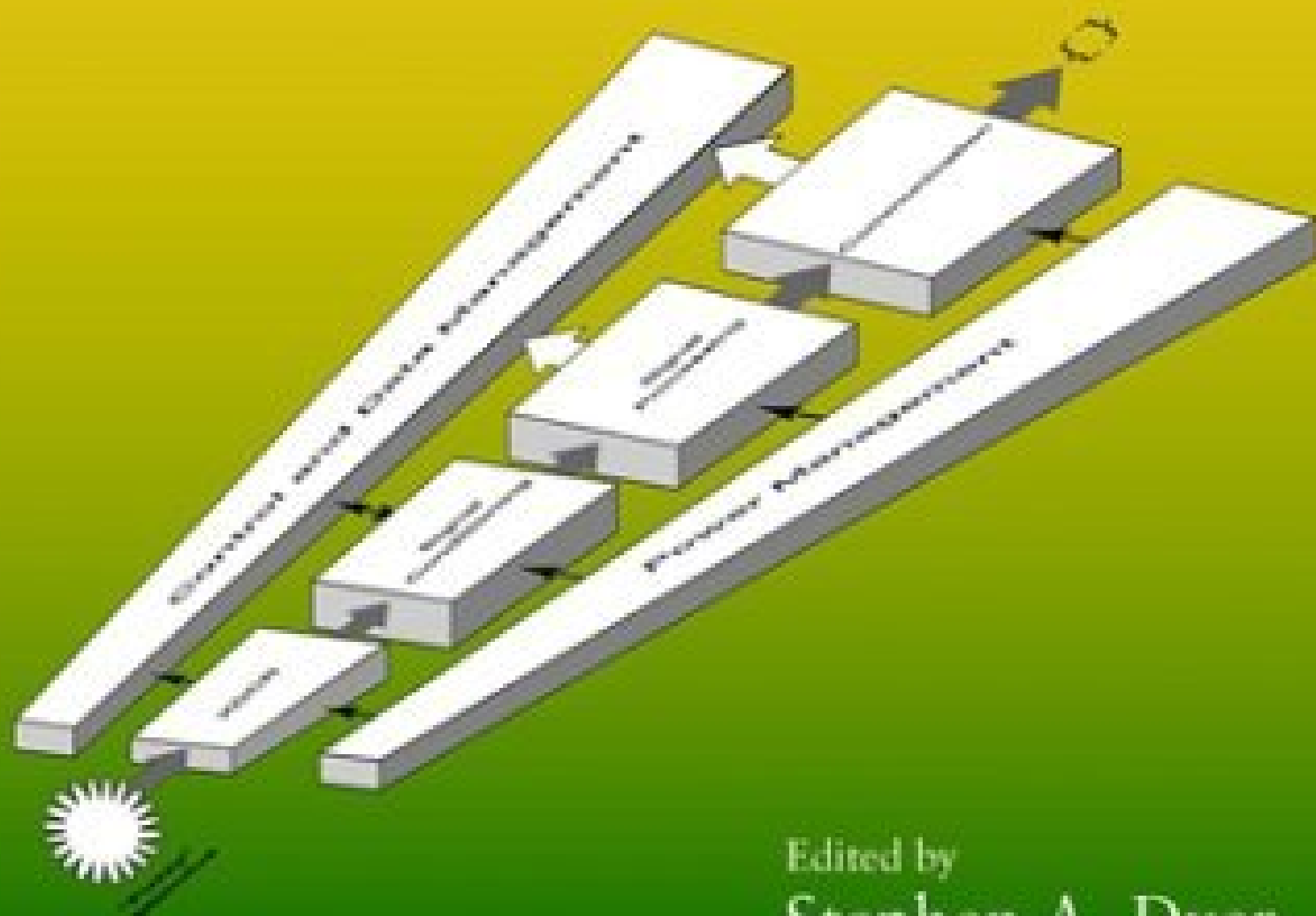


*Wiley
Survey
of*

Instrumentation and Measurement



Edited by
Stephen A. Dyer

SURVEY OF INSTRUMENTATION AND MEASUREMENT

SURVEY OF INSTRUMENTATION AND MEASUREMENT

Stephen A. Dyer, *Editor*

Department of Electrical and Computer Engineering
Kansas State University



WILEY-INTERSCIENCE
John Wiley & Sons, Inc.

New York/Chichester/Weinheim/Brisbane/Singapore/Toronto

This book is printed on acid-free paper. (∞)

Copyright © 2001 by John Wiley & Sons. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:

Survey of instrumentation and measurement / Stephen A. Dyer, editor.

p. cm.

Articles selected from the Wiley encyclopedia of electrical and electronics engineering.

Includes index.

ISBN 0-471-39484-X (cloth : alk. paper)

1. Mensuration. 2. Measuring instruments. 3. Electric meters. I. Dyer, Stephen A.

T50 .S89 2001

681'.2—dc21

00-068602

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE vii

INTRODUCTION ix

INSTRUMENTATION AND MEASUREMENT— INTRODUCTORY 1

Instruments	1
Measurement Errors	14
Automatic Test Equipment	25

SENSORS AND TRANSDUCERS 37

Electric Sensing Devices	37
Magnetic Sensors	46
Hall Effect Transducers	60
Pressure Sensors	71
Piezoresistive Devices	84
Strain Sensors	87
Temperature Sensors	101
Thermistors	122
Thermocouples	129
Thermopiles	135
Accelerometers	147
Tachometers	163
Fiber-Optic Sensors	178
Ultraviolet Detectors	199
Chemical Sensors	207
Intelligent Biosensors	222

SIGNAL CONDITIONING 241

Instrumentation Amplifiers	241
Current Shunts	252
Current Transformers	259
Instrument Transformers	266

GENERAL-PURPOSE INSTRUMENTATION AND MEASUREMENT 277

Multimeters	277
Digital Multimeters	286
Ohmmeters	291
Electrometers	293
Balances	301
Bridge Instruments	309
Oscilloscopes	326

ELECTRICAL VARIABLES 339

Charge Measurement	339
Electric Current Measurement	343
Inductance Measurement	347
Capacitance Measurement	358
Q-Factor Measurement	372
Q-Meters	387

ELECTROMAGNETIC VARIABLES 397

Electric Strength	397
Electromagnetic Field Measurement	401
Magnetic Field Measurement	410
Attenuation Measurement	422
Loss-Angle Measurement	439
Wavelength Meter	452
Reflectometers, Time-Domain	467
Standing Wave Meters and Network Analyzers	476
Millimeter-Wave Measurement	496

MECHANICAL VARIABLES 509

Displacement Measurement	509
Velocimeters	521
Acceleration Measurement	535
Density Measurement	546
Level Meters	557
Flowmeters	568
Dynamometers	580
Torquemeters	587
Weighing	596
Gravimeters	607
Viscometers	611
Compasses	624
Goniometers	635
Gyroscopes	644
Hygrometers	666
Pyrometers	680
Acoustic Variables Measurement	688

TIME, FREQUENCY, AND PHASE 697

Frequency and Time Interval Meters	697
Frequency Standards, Characterization	707
Frequency Stability	715
Phase Meters	721

Time Measurement	733
Clocks In Telecommunications	743

NOISE AND DISTORTION 759

Electric Noise Measurement	759
Electric Distortion Measurement	771
Intermodulation Measurement	778
Measurement of Frequency, Phase Noise and Amplitude Noise	793

POWER AND ENERGY 809

Power Measurement	809
Power Meters	826
Wattmeters	840
Volt-Ampere Meters	848
Power Factor Measurement	855
Electrical Energy Measurement	867
Watthour Meters	874

INSTRUMENTATION FOR CHEMISTRY AND PHYSICS 885

Photometers	885
Photometric Light Sources	892
pH Measurement and Control	900
Pulse Height Analyzers	915
Scintillation Counters	919

Thin Film Analyzers	931
Radiometry	937

INTERFEROMETERS AND SPECTROMETERS 947

Interferometers	947
Acoustic Wave Interferometers	957
Mass Spectrometers	964

MICROSCOPY 971

Electron Microscopes	971
X-Ray Microscopy	995

DATA ACQUISITION AND RECORDING 1013

Radiotelemetry	1013
Data Recording	1026
Recorders	1035

TESTING METHODS 1053

Eddy Current Testing	1053
Eddy Current Nondestructive Evaluation	1062
Impulse Testing	1075

INDEX 1083

PREFACE

“Measurement is the heart of all science, engineering, and technology.” Thus wrote Peter K. Stein, a measurement engineer and educator, several decades ago.

And it is true. Beginning some 5000 years ago with the establishment of measurement standards—important to trade, government, work, and religion—and continuing past the last few centuries B.C.E., when primitive instruments were developed for measuring time, this now vast and all-important field of instrumentation and measurement (I&M) got its kickstart during the sixteenth century with what can be considered the birth of experimental science. The eighteenth and following centuries have seen great progress in science and engineering in general and I&M in particular. The development of electronics and computers has allowed very sophisticated—and relatively easy-to-use—instruments to be designed. These instruments, when coupled with the ingenuity of careful and clever investigators, have in turn led to the ability to make very accurate, precise, important measurements of sometimes-very-hard-to-access quantities, both macroscopic and microscopic.

What do I&M systems do? They measure physical quantities, such as temperature, pressure, voltage, and current. They locate events and objects. They count things. They discriminate based on one or more parameters. The need for I&M is everywhere. Everyone uses it.

Indeed, the number and sheer breadth of applications make it challenging to give coverage to such a field within a single

volume. The 97 articles appearing herein are drawn from the *Wiley Encyclopedia of Electrical and Electronics Engineering*. They are chosen to give a broad representation of I&M, but with the realization that most of the other topics in the full encyclopedia remain in some relationship with I&M, either affected by it or having some important contribution to this fascinating field.

The articles contained within are grouped into topical categories selected to help you navigate this volume. Of course, there are some articles that might comfortably fit within more than one category, and, likewise, there are occasions in which articles from two or more categories are closely related. We hope that, given the table of contents, the index, and your own curiosity, it will be reasonably easy to make the appropriate connections.

Physicists, chemists, mathematicians, statisticians, computer scientists, a variety of scientists from other disciplines, engineers of various types, talented technicians, and others—many have, either directly or indirectly, made contributions to I&M. It is our hope that this survey volume will indeed prove a significant aid as you make use of—and perhaps your own contributions to—this diverse, exciting, and continually growing field.

STEPHEN A. DYER

INTRODUCTION

Instrumentation and Measurement—Introductory

Kelvin's first rule of instrumentation states, in essence, that the measuring instrument should not alter the event being measured. That rule impacts the design of any instrument.

In broad terms, we can view a measuring instrument or instrument subsystem as comprising an input transducer followed by a signal-conditioning section, which in turn drives a data-processing and display section.

The instrument or piece of test equipment may be completely manual, requiring (sometimes significant) tending by its operator, or it may be automated to one extent or another. The increasing availability of inexpensive computing power, beginning in the early 1970s, has encouraged the design of instrumentation capable of being controlled by computer, with all but the simplest of instruments today having interfaces built in.

Sensors and Transducers

Transducers are energy-conversion devices and are the “molecule” of a measurement system (i.e., the smallest unit from which measurement systems are constructed). Some transducers, such as thermocouples, are active (or self-generating); i.e., they are such that the input energy in one form directly produces another form of energy at its output. Other transducers, such as resistance thermometers, require two forms of input energy to produce some form of output energy; they are termed passive, or non-self-generating.

Traditionally, sensors have been taken to be the passive portions of non-self-generating transducers. They are usually elements which have one or more physical properties affected by some physical quantity.

In today's literature there is often a lack of distinction between “sensor” and “transducer.” The terms are often used interchangeably. Furthermore, integrated-circuit technology has led to the inclusion of signal conditioning circuitry and data-acquisition-and-communication interfaces on the same chip as the sensor or transducer. While the entire package is typically referred to as a sensor, in reality it is an instrumentation subsystem, or perhaps a complete instrument, on a chip.

While the need for traditional, discrete sensors and transducers will certainly continue, present research and development of integrated sensors, intelligent sensors, and integrated systems on a chip will result in an expanding offering

of compact, rugged, relatively inexpensive, and easy-to-apply measurement front ends and systems based on one or more sensors or transducers.

Signal Conditioning

It is the task of the signal conditioner to accept the signal output from the transducer and produce from it a signal acceptable for introduction to the remainder of the instrument.

Signal conditioning is typically analog in nature, and it can involve strictly linear operations, strictly nonlinear operations, or some combination of the two. Examples of linear operations include amplitude scaling (amplification or attenuation), impedance transformation, linear filtering, and modulation. Some nonlinear operations include obtaining the root-mean-square (rms) value, absolute value, square root or logarithm of the input signal.

The signal conditioning may also provide auxiliary services, such as producing an excitation signal for the transducer, providing a reference signal for the transducer, or introducing electrical isolation.

Building blocks for providing analog signal processing are available in many forms—as modules, as integrated circuits (ICs), and as complete subsystems having universal backplanes and plug-in input and output modules. Operational amplifiers, instrumentation amplifiers, isolation amplifiers, nonlinear processing circuits such as comparators, multiplier/dividers, log/antilog amplifiers, trigonometric-function generators, and rms-to-dc converters are but a few of the building blocks available.

General-Purpose Instrumentation

The electrical quantities most commonly measured include voltage, current, charge, resistance, and impedance. This section surveys general-purpose instrumentation typically used for measurement of such quantities. Many of these instruments can be found on even the most modest of laboratory or test benches.

Electrical Variables

There are a variety of approaches to measuring the electrical variables of charge, current, inductance, capacitance, and quality (Q). This section provides an overview and, in several cases,

acts as supplemental material to the discussions on various general-purpose instruments.

Electromagnetic Variables

The measurement of electromagnetic variables is useful in a wide variety of applications, from determination of the moisture content in grain to satisfying governmental regulations on radiated and conducted emissions associated with electronic equipment. This section provides coverage of a variety of instruments and measurement techniques applicable at frequencies up through the millimeter range of the electromagnetic spectrum.

Mechanical Variables

Linear displacement, velocity, acceleration, and jerk; angular displacement, velocity, and acceleration; length; force; torque; shaft power; pressure; sound; fluid flow; temperature; heat flux; humidity; and liquid level all fall broadly under “mechanical variables.” The fundamental quantities are length, time, mass, and temperature, and there exist independent standards for each. Other quantities such as force, pressure, etc., are derived quantities. This section provides coverage of a breadth of measurement techniques applicable to mechanical variables. The reader should also refer to the “Sensors and Transducers” section, as well as “Time, Frequency and Phase” section for additional and related information.

Time, Frequency, and Phase

Time is one of the fundamental quantities in nature, and it plays a very important role in the vast majority of measurements made, either directly or through its relationship to two other often-measured quantities—frequency and phase. This section presents the important aspects of the measurement of time, frequency and phase. It provides general information on the measurement of time, addresses frequency standards, and discusses time-interval, frequency and phase meters, along with additional topics.

Instrumentation for Chemistry and Physics

There is a vast collection of instruments and instrumental methods used in chemical and physical analysis, and the progress in instrumental methods of analysis over the past few decades can be aptly described as “explosive.” As with many other areas of instrumentation and measurement, the computer has played a significant role.

Even a brief list of instrumental methods would probably include colorimetry; fluorescence; spectrophotometry; flame photometry; spectrometry; turbidimetry; nephelometry; thermal conductivity; X-ray diffraction; radiochemical methods; refractometry; interferometry; pH determination; potentiometric, conductometric and amperometric titrations; electrolytic separations; polarography; chromatography; and nuclear magnetic resonance.

This section gives an introduction to several common instruments and techniques. Coverage of interferometers and

spectrometers is given under a separate section. Likewise, two important types of microscopy—electron and X-ray—have coverage under their own sections.

Interferometers and Spectrometers

Interferometers are nondispersive devices in which patterns of constructive and destructive interference of electromagnetic or mechanical waves can be set up. A. A. Michelson pioneered the use of interferometry in 1880 in making accurate measurements of distance. The invention of the laser and its subsequent application to interferometry have resulted in the commercial practicality of interferometric measurement of distances. Another very important use of interferometers is in obtaining spectral information. Indeed, interferometers form the basis for many types of spectrometers.

Spectrometers are instruments used to provide information about a spectrum associated with a sample. The spectrum, for example, might be of radiation absorbed, emitted, or scattered by the sample over a particular range of wavelengths; it might be a mass spectrum, a record of the relative numbers of the various kinds of ions associated with a molecule, sorted by their charge-to-mass ratio; or it might indicate the characteristic absorption of energy by spinning nuclei in a strong magnetic field.

This section provides an introduction to interferometers and describes the mass spectrometer, one of the most broadly useful of all the tools employed in instrumental analysis.

Data Acquisition and Recording

The “storage and readout” portions of instruments have made tremendous strides over the past decades. The large, analog industrial recorder has been supplanted to a large extent by systems that acquire the analog data, convert it to digital form, and store it on media such as magnetic tape or disk, magneto-optical disk, solid-state memory cards, or compact disc (either CD-R or CD-RW). The “readout”—or display—might be a cathode-ray tube (CRT), a liquid-crystal display (LCD), or something similar. There remain, of course, recorders that produce hard-copy on paper, but most of these are hybrid recorders which are microprocessor-based and which typically expect the data in digital form.

The sophisticated, expensive, analog recorders which relied on frequency-modulation (FM) techniques likewise are hard to find nowadays. Analog-to-digital converters with 16- to 24-bit resolution are inexpensive, and the cost-per-gigabyte of hard-disk storage is measured in tens of dollars (US); CD-RW media are one to two orders of magnitude less expensive yet. Even when considered only from the standpoint of cost, digital has proven itself the route to take for most data-storage applications associated with instrumentation and measurement.

Recording instruments, as more and more computing power is provided them, will continue to be placed in increasingly sophisticated applications. Their use as smart network devices, as data nodes, and even as control elements in fieldbus-based systems are just a few of the uses for some of the newer recording instruments.

CONTRIBUTORS

- Muhammad Taher Abuelma'atti**, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
- Farid Ahmed**, Penn State University—Erie, The Behrend College
- Mohammad S. Alam**, Purdue University, Fort Wayne, IN
- S. Amelinckx**, University of Antwerp, Belgium
- G. Andreoli**, Space Applications Institute, Ispra, Italy
- David R. Armitage**, National Physical Laboratory, Middlesex, England
- Aydin Aykan**, Rohde & Schwarz GmbH & Co.KG, Munich, Germany
- Daniël H. J. Baert**, University of Ghent, Belgium
- Billur Barshan**, Bilkent University, Ankara, Turkey
- V. R. Basker**, University of Florida, Gainesville
- N. K. Batra**, Naval Research Laboratory, Washington, DC
- Gerard Beaudin**, Observatoire de Paris—Meudon, France
- Necip Berme**, The Ohio State University, Columbus
- Giovanni Betta**, University of Cassino, Italy
- Peter J. Betts**, Commonwealth Scientific and Industrial Research Organisation, Division of Telecommunications and Industrial Physics, Lindfield, Australia
- H. R. Bilger**, Oklahoma State University, Stillwater
- Jerome J. Blair**, Bechtel Nevada, Las Vegas
- Wojtek J. Bock**, Université du Québec à Hull
- Derek F. Bowers**, Analog Devices Incorporated, Santa Clara, CA
- George Braoudakis**, The University of Sydney, Australia
- Stefano Bregni**, Politecnico di Milano, Milan, Italy
- Giovanni Bucci**, University of Aquila, Italy
- Theodore W. Cannon**, National Renewable Energy Laboratory, Golden, CO
- Bruce W. Carsten**, Bruce Carsten Associates, Inc., Corvallis, OR
- Kar-Ming Cheung**, California Institute of Technology, Pasadena
- D. D. L. Chung**, State University of New York at Buffalo
- Oscar D. Crisalle**, University of Florida, Gainesville
- Dominique Cros**, Université de Limoges, France
- Paul Crozat**, Université Paris-Sud, Orsay, France
- António M. Cruz Serra**, Ecole Supérieure d'Electricité (Supelec), Lisbon, Portugal
- Leszek S. Czarniecki**, Louisiana State University, Baton Rouge
- Gilles Dambrine**, Institut d'Electronique et Microélectronique du Nord (IEMN), Cedex, France
- Massimo D'apuzzo**, University of Napoli, Naples, Italy
- Claudio De Capua**, Federico II University of Naples, Italy
- Ernesto DeLa Cruz**, Rutgers, The State University of New Jersey, New Brunswick
- P. P. Delsanto**, Politecnico di Torino, Turin, Italy
- Alexander Domijan, Jr.**, University of Florida, Gainesville
- Halit Eren**, Curtin University of Technology, Perth, Australia
- Anthony H. Falkner**, Coventry University, United Kingdom
- Eva S. Ferre-Pikal**, National Institute of Standards and Technology, Boulder, CO
- Fabio Filicori**, University of Bologna, Italy
- Antonino S. Fiorillo**, University of Salerno, Pisa, Italy
- Arthur L. Fricke**, University of Florida, Gainesville
- Robert E. Gander**, University of Saskatchewan, Saskatoon
- Michael F. Gard**, Subsite Electronics, Perry, OK
- Helena M. Geirinhas Ramos**, Ecole Supérieure d'Electricité (Supelec), Lisbon, Portugal
- Richard G. Geyer**, National Institute of Standards and Technology, Boulder, CO
- David M. Gleason**, United States Air Force Research Laboratory, Hanscom AFB, MA
- Ian A. Glover**, University of Bradford, Yorkshire, England
- Andreas Gronefeld**, Ruhr-Universität Bochum, Germany
- Li-Qun Gu**, Texas A&M University, College Station
- Martin Guy**, Institut National d'Optique, Québec
- Albert Helfrick**, Embry-Riddle Aeronautical University, Daytona Beach, FL
- Kurt Hilty**, Swiss Federal Office of Metrology, Bern, Switzerland
- B. Hosgood**, Space Applications Institute, Ispra, Italy
- Alan L. Huston**, Naval Research Laboratory, Washington, DC
- Bernard Huyart**, Ecole Nationale Supérieure des Télécommunications (ENST), Orsay, France
- Gaetano Iuculano**, University of Florence, Italy
- Brian L. Justus**, Naval Research Laboratory, Washington, DC
- Pentti Jutila**, Helsinki University of Technology, Finland
- S. Kaliyugavaradan**, Madras Institute of Technology, India
- Motohisa Kanda**, National Institute of Standards and Technology, Boulder, CO
- Mirosław A. Karpierz**, Warsaw University of Technology, Poland
- Irving Kaufman**, Arizona State University, Tempe
- M. Kaykobad**, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
- Andy J. Knott**, National Physical Laboratory, Teddington, England
- C. Koehler**, Space Applications Institute, Ispra, Italy
- Juha Kostamovaara**, University of Oulu, Finland
- Jerzy Krupka**, Instytut Mikroelektroniki i Optoelektroniki Politechniki Warszawskiej, Warszawa, Poland
- V. Jagadeesh Kumar**, Indian Institute of Technology, Madras, India
- Carmine Landi**, University of Aquila, Italy
- Christine Latrasse**, Laval University, Quebec, Canada
- Anthony D. Lavietes**, Lawrence Livermore National Laboratory, Livermore, CA
- Stanisław F. Legowski**, University of Wyoming, Laramie, WY
- Angelica Leitmannova-Ottova**, Slovak Technical University, Bratislava, Slovakia
- Manfred Lenzen**, The University of Sydney, Australia
- Consolatina Liguori**, University of Cassino, Italy
- Fred Liguori**, ATE Consulting Services, Browns Mills, New Jersey
- Baoming Ling**, Zhejiang University, Hangzhou, China
- Andre Luiten**, University of Western Australia, Perth, Australia
- Kari Määttä**, University of Oulu, Finland
- Takashi Maeno**, Communications Research Laboratory, Tokyo, Japan
- David W. Martyny**, University of Florida, Gainesville, FL
- Vojko Matko**, University of Maribor, Slovenia

Mark McDowell, Microgravity Sciences Division, Cleveland, OH
James H. McQuaid, Lawrence Livermore National Laboratory, CA
Christopher J. Melhorn, Electrotek Concepts, Inc., Knoxville, TN
Domenico Mirri, University of Bologna, Italy
Devendra K. Misra, University of Wisconsin, Milwaukee
Douglas R. Morr, Bertec Corporation, Columbus, OH
Hitoshi Muguruma, Rutgers, The State University of New Jersey, Piscataway, NJ

V. G. K. Murti, Indian Institute of Technology, Madras, India
M. M. Nayak, Indian Space Research Organisation (ISRO), Bangalore, India
Jean-Michel Nebus, Universite de Limoges, France
Mário Ventim Neves, Technical University of Lisbon, Portugal
Charles B. Newcombe, Fluke Corporation, Everett, WA
Jacques Nicolas, Belgian Ministry of Economic Affairs, Brussels
Seppo J. Ovaska, Helsinki University of Technology, Finland
Gaetano Pasini, University of Bologna, Italy
Thomas J. Paulus, EG&G ORTEC, Oak Ridge, TN
Henrik Pedersen, Rutgers, The State University of New Jersey, Piscataway
Antonio Pietrosanto, University of Salerno, Italy
J. Piironen, Space Applications Institute, Ispra, Italy
Roushdy M. Radwan, Cairo University, Egypt
Elvi Räisänen-Ruotsalainen, University of Oulu, Finland
K. Rajanna, Indian Institute of Science, Bangalore, India
A. Romano, Politecnico di Torino, Italy
Wayne D. Ruhter, Lawrence Livermore National Laboratory, CA
Andrew Rusek, Oakland University, Rochester, MI
St. Sandmeier, University of Zurich, Switzerland
P. Sankaran, Indian Institute of Technology, Madras, India
Renato Sasdelli, University of Bologna, Italy
M. Scalerandi, Politecnico di Torino, Italy
Burkhard Schiek, Ruhr-Universität Bochum, Germany

M. Roman Serbyn, Morgan State University, Baltimore, MD
Pedro M. B. Silva Girão, Ecole Supérieure d'Electricité (Supelec), Lisbon Portugal
Sangeet Singh-Gasson, Naval Research Laboratory, Washington, DC
K. Srinivasan, St. Bruno, Quebec, Canada
Manfred Stecher, Rohde & Schwarz GmbH & Co.KG, Tutzing, Germany
Alex Stein, Quantum Logic Corporation, Secaucus, NJ
Srinivas Tadigadapa, Integrated Sensing Systems Incorporated
Michel Têtu, Laval University, Québec
H. Ti Tien, Michigan State University, E. Lansing
Michael E. Tobar, The University of Western Australia, Perth
Pierre Tremblay, Laval University, Québec
Haiping Tsou, California Institute of Technology, Pasadena
Hidemi Tsuchida, Electrotechnical Laboratory, Tsukuba, Japan
Lalita Udpa, Iowa State University, Ames
Satish S. Udpa, Iowa State University, Ames
Adriaan van den Bos, Delft University of Technology, Netherlands
D. Van Dyck, University of Antwerp, Belgium
Guy Vernet, Université Paris—Sud, France
Bernard Villeneuve, Nortel, Ottawa, Ontario
Fred L. Walls, National Institute of Standards and Technology, Boulder, CO
L. I. Winkler, Appalachian State University, Boone, NC
Roland Wiśniewski, Warsaw University of Technology, Poland
Darold C. Wobschall, State University of New York at Buffalo
Lawrence B. Wolff, The Johns Hopkins University, Baltimore
Hermann Wollnik, Universität Giessen, Germany
Ken Xu, Huazhong University of Science and Technology, Wuhan, China
Jean-Peter Ylén, Helsinki University of Technology, Finland
Jan Zakrzewski, Silesian Technical University, Poland
Xiangbin Zhuge, Zhejiang University, Hangzhou, China

INSTRUMENTATION AND MEASUREMENT—INTRODUCTORY

INSTRUMENTS

Measurement is essential for observing and testing scientific and technological investigations. It is so fundamental and important to science and engineering that the whole science can be said to be dependent on it. Instruments are developed for monitoring the conditions of physical variables and converting them into symbolic output forms. They are designed to maintain prescribed relationships between the parameters being measured and the physical variables under investigation. The physical parameter being measured is known as the *measurand*. The sensors and transducers are the primary sensing elements in the measuring systems that sense the physical parameters to produce an output. The output from the sensor is supplied to a transducer, which converts energy from one form to another. Therefore, a transducer is a device capable of transferring energy between two physical systems.

Measurement is a process of gathering information from a physical world and comparing this information with agreed standards. Measurement is carried out with instruments that are designed and manufactured to fulfill given specifications. After the sensor generates the signals, the type of signal processing depends on the information required from it. A diverse range of sensors and transducers may be available to meet the measurement requirements of a physical system. The sensors and transducers can be categorized in a number of ways depending on the energy input and output, input variables, sensing elements, and electrical or physical principles. For example, from an energy input and output point of view, there are three fundamental types of transducers: modifiers, self-generators, and modulators.

In modifiers, a particular form of energy is modified rather than converted; therefore, the same form of energy exists at the input and at the output. In self-generators, electrical signals are produced from nonelectric inputs without the application of external energy. These transducers produce very small signals, which may need additional conditioning. Typical examples are piezoelectric transducers and photovoltaic cells. Modulators, on the other hand, produce electric outputs from non-electric inputs, but they require an external source of energy. Strain gauges are typical examples of such devices.

The functionality of an instrument can be broken into smaller elements, as illustrated in Fig. 1. Most measurement systems have a sensor or transducer stage, a signal-conditioning stage, and an output or termination stage. All

instruments have some or all of these functional blocks. Generally, if the behavior of the physical system under investigation is known, its performance can be assessed by means of a suitable method of sensing, signal conditioning, and termination.

In the applications of instruments, the information about a physical variable is collected, organized, interpreted, and generalized. Experiments are conceived, performed, and repeated; as we acquire confidence in the results, they are expressed as scientific laws. The application of instruments ranges from laboratory conditions to arduous environments such as inside nuclear reactors or remote locations such as satellite systems and spaceships. In order to meet diverse application requirements of high complexity and capability, many manufacturers have developed a large arsenal of instruments. Some of these manufacturers are listed in Table 1.

In recent years, rapid growth of IC electronics and the availability of cheap analog-to-digital and microprocessors have led to rapid progress in the instrumentation field, with the development of instruments, measuring techniques, distributed architectures, and standards aimed to improve performances.

Instruments are applied for static or dynamic measurements. The static measurements are relatively easy since the physical quantity (e.g., fixed dimensions and weights) does not change in time. If the physical quantity is changing in time, which often is the case, the measurement is said to be dynamic. In this case, steady-state and transient behavior of the physical variable must be analyzed so that it can be matched with the dynamic behavior of the instrument.

DESIGN, TESTING, AND THE USE OF INSTRUMENTS

Instruments are designed on the basis of existing knowledge, which is gained either from the experiences of people about the physical process or from the structured understanding of the process. In any case, ideas conceived about an instrument must be translated into hardware and/or software that can perform well within the expected standards and easily be accepted by the end users.

Usually, the design of instruments requires many multidisciplinary activities. In the wake of rapidly changing technology, instruments are upgraded often to meet the demands of the marketplace. Depending on the complexity of the proposed instrument, it may take many years to produce an instrument for a relatively short commercial lifetime. In the design and

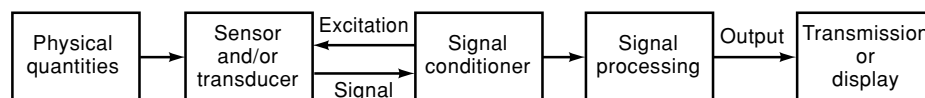


Figure 1. An instrument has a number of relatively independent components that can be described as functional elements. These functional elements are the sensors and transducers, signal conditioners, and output or terminations. In general, if the behavior of the physical system is known, its performance is measured by a suitable arrangement and design of these components.

Table 1. List of Manufacturers

ABB, Inc. 501 Merritt 7, P.O. Box 5308 Norwalk, CT 06856-5308 Tel: 800-626-4999 Fax: 203-750-2263	Keithley Instrument, Inc. 28775-T Aurora Road Cleveland, OH 44139-1891 Tel: 800-552-1115 Fax: 440-248-6168
Allied Signal, Inc. 101 Columbia Road Morristown, NY 07962 Tel: 800-707-4555 Fax: 608-497-1001	MCS Calibration, Inc. Engineering Division 1533 Lincoln Avenue Halbrook, NY 11741 Tel: 800-790-0512 Fax: 512-471-6902
Bailey-Fisher and Porter Company 125 E County Line Road Wanminster, PA 18974 Tel: 800-268-8520 Fax: 215-674-7183	MSC Industrial Supply Company 151-T Sunnyside Boulevard Plainview, NY 11803 Tel: 800-753-7937 Fax: 516-349-0265
Consolidated Instrument, Inc. 510 Industrial Avenue Teterboro, NC 07608 Tel: 800-240-3633 Fax: 201-288-8006	National Instruments 6504 Bridge Point Parkway Austin, TX 78730-7186 Tel: 512-794-0100; 888-217-7186 Fax: 512-794-8411
Davies Instrument Manufac- turing Company, Inc. 4701 Mt. Hope Drive Baltimore, MD 21215 Tel: 800-548-0409 Fax: 410-358-0252	Omega Engineering, Inc. P.O. Box 4047 Stamford, CT 06907 Tel: 800-826-6342 Fax: 203-359-7700
Dwyer Instrument, Inc. P.O. Box 373-T Michigan City, IN 46361-0373 Tel: 219-879-8000 Fax: 219-872-9057	Rosemount Analytical 600 S. Harbor Boulevard, Dept TR La Habra, CA 90631-6166 Tel: 800-338-8099 Fax: 562-690-7127
Fuji Corporation of America Park 80 West, Plaza Two Saddlebrook, NJ 07663 Tel: 201-712-0555 Fax: 201-368-8258	Scientific Instruments, Inc. 518 W Cherry Street Milwaukee, WI 53212 Tel: 414-263-1600 Fax: 415-263-5506
Hanna Instrument, Inc. Highland Industrial Park 584 Park East Drive Woonsocket, RI 02895-0849 Tel: 800-999-4144 Fax: 401-765-7575	Space Age Control, Inc. 38850 20th Street East Palmdale, CA 93550 Tel: 800-366-3408 Fax: 805-273-4240
Hewlett-Packard Company 5301 Stevens Creek Bou- levard Santa Clara, CA 95052-8059 Tel: 800-452-6866 Fax: 303-756-6800	Tektronix, Inc. P.O. Box 500 Beaverton, OR 97077 Tel: 503-627-7111
Industrial Instruments and Supply, Inc. P.O. Box 416 12 County Line Industrial Park Southampton, PA 18966 Tel: 800-523-6079 Fax: 215-396-0833	Texas Instrument, Inc. 34 Forest Street, MS 23-01 P.O. Box 2964 Attleboro, MA 02703 Tel: 508-236-3287 Fax: 508-236-1598
Instrument and Control Ser- vices Company 1351-T Cedar Lake Road Lake Villa, IL 60046 Tel: 800-747-8367 Fax: 847-356-9007	Warren-Knight Instrument Company 2045 Bennett Drive Philadelphia, PA 19116 Tel: 215-464-9300 Fax: 215-464-9303
	Yokogawa Corporation of America 2 Dart Road Newnon, GA 30265-1040 Tel: 800-258-2552 Fax: 770-251-2088

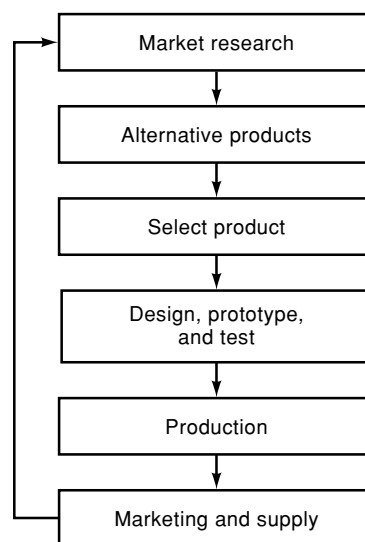


Figure 2. The design process from the conception of ideas to marketing follows carefully considered stages. The proper identification and effective implementation of these stages is important in the success and acceptance of a specific instrument in the marketplace.

production of instruments, one must consider such factors as simplicity, appearance, ease and flexibility of use, maintenance requirements, lower production costs, lead time to product, and positioning strategy in the marketplace.

In order to design and produce instruments, a firm must consider many factors. These include sound business plans, suitable infrastructure, plant, equipment, understanding of technological changes, skilled and trained personnel, adequate finance, marketing and distribution channels, and a clear understanding about worldwide instrument and instrumentation system trends. It is important to choose the right product that is very likely to be in demand in the years to come. Here entrepreneurial management skills may be an important factor.

The design process itself may follow well-ordered procedures from idea to marketing stages. The process may be broken down into smaller tasks such as identifying specifications, developing possible solutions for these specifications, modeling, prototyping, installing and testing, making modifications, manufacturing, planning marketing and distribution, evaluating customer feedback, and making design and technological improvements. Figure 2 illustrates the stages for the design and marketing of an instrument. Each one of these stages can be viewed in detail in the form of subtasks. For example, many different specifications may be considered for particular product. These specifications include but are not limited to operational requirements, functional and technological requirements, quality, installation and maintenance, documentation and servicing, and acceptance level determination by the customers.

In recent years, computers have been used extensively in the instrument manufacturing industry in the form of computer-aided-design (CAD), automated testing, and in other applications. The computer enables rapid access to knowledge-based information and makes design time considerably shorter, thus enabling manufacturers to meet the customer demand. In CAD systems, mechanical drafting software, electronic circuit design tools, control analysis tools, and mathematical and word processing tools are integrated to assist the design procedure.

Design software is available from various manufacturers listed in Table 1.

Testing and Use of Instruments

After the instrument is designed and prototyped, various evaluation tests may be conducted. These tests may be made under reference conditions or under simulated environmental conditions. Some examples of reference condition tests are accuracy, response time, drift, and warm up time. Simulated environmental tests may be compulsory, being regulated by governments and other authorities. Some simulated environment tests include climatic test, drop test, dust test, insulation resistance test, vibration test, electromagnetic compatibility (EMC) tests, and safety and health hazard tests. Many of these tests are strictly regulated by national and international standards.

Adequate testing and proper use of instruments is important to achieve the best results out of them. When the instruments are installed, a regular calibration is necessary to ensure the consistency of the performance over the time period of operation. Incorrect measurements can cost a considerable amount of money or even lead to loss of lives.

For maximum efficiency, appropriate instrument for the measurement must be selected. Users should be fully aware of their application requirements, since instruments that do not fit their purpose will deliver false data thus resulting in wasted time and effort. When selecting the instrument, users must evaluate many factors such as accuracy, frequency response, electrical and physical loading effects, sensitivity, response time, calibration intervals, power supply needs, spare parts, technology, and maintenance requirements. They must ensure compatibility with their existing equipment.

Also, when selecting and implementing instruments, quality becomes an important issue from both quantitative and qualitative perspectives. The quality of an instrument may be viewed differently depending on the people involved. For example, quality as viewed by the designer may be an instrument designed on sound physical principles, whereas from the user's point of view quality may be reliability, maintainability, cost, and availability.

For the accuracy and validity of information collected from the instruments, correct installation and proper use become very important. The instruments must be fully integrated with the overall system. Sufficient background work must be conducted prior to installation to avoid a possible shutdown of the process that is longer than necessary.

Once the system is installed, the reliability of the instrument must be assessed, and its performance must be checked regularly. The *reliability* of the system may be defined as the probability that it will operate at an agreed level of performance for a specified period of time. The reliability of instruments follows a bath tub shape curve against time. Instruments tend to be unreliable in the early and later stages of their lives. During normal operations, if the process conditions change (e.g., installation of large machinery nearby), calibrations must be conducted to avoid possible performance deterioration of the instrument. Therefore, the correct operations of instruments in service must be assured at all times throughout the lifetime of the device.

Once the instruments are installed they may be left alone and expected to operate reliably. In some applications, the instruments may be part of a large instrumentation system, taking a critical role in monitoring and/or controlling the process

and operations. They may be communicating with other devices and their performance may affect the performance of the rest of the system, as in the case of the process industry. However, in many applications, instruments are used on a stand-alone basis for laboratory and experimental work, and the success of the experiments may entirely depend on their correct usage and performance. In these cases, the experiments must be designed and conducted carefully by identifying the primary variables, controlling, selecting the correct instruments, assessing the relative performances, validating the results, and using the data effectively by employing comprehensive data analysis techniques. Set procedures for experimental designs can be found in various sources given in the Bibliography (e.g., Sydenham et al., 1989).

After having performed the experiments, the data must be analyzed appropriately. This can be done at various stages by examining the consistency of the data, performing appropriate statistical analyses, estimating the uncertainties of the results, relating the results to the theory, and correlating the data. Details of statistical data analysis can be found in many books; also many computer software programs are available for the purpose analysis including common packages such as Microsoft Excel.

INSTRUMENT RESPONSE AND DRIFT

Instruments respond to physical phenomena by sensing and generating signals. Depending on the type of instrument used and the physical phenomenon observed, the signals may be either slow or fast to change, and may also contain transients. The response of the instruments to the signals can be analyzed in a number of ways by establishing static and dynamic performance characteristics. Although, the static performances are relatively simple, the dynamic performances may be complex.

Static Response

Instruments are often described by their dynamic ranges and full-scale deflections (span). The *dynamic range* indicates the largest and smallest quantities that can be measured. The *full-scale deflection* of an instrument refers to the maximum permissible value of the input quoted in the units of the particular quantity to be measured.

In instruments, the change in output amplitude resulting from a change in input amplitude is called the *sensitivity*. System sensitivity often is a function of external physical variables such as temperature and humidity. The relative ratio of the output signal to the input signal is the *gain*. Both, the gain and sensitivity are dependent on the amplitude of the signals and the frequency, which will be discussed in the section on dynamic response.

In the design stages or during manufacturing, there might be small differences between the input and output, which is called the *offset*. In other words, when the input is zero the output is not zero or vice versa. The signal output also may change in time, which is known as the *drift*. The drift can occur for many reasons including temperature and aging. Fortunately, drift usually occurs in a predictable manner. A typical drift curve of an instrument against time is illustrated in Fig. 3.

During practical applications, readings taken from an instrument under the same conditions may not be repeatable. In this case, a repeatability test may be conducted, and statistical

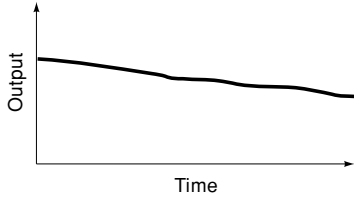


Figure 3. Drift in the output of an instrument. The main causes of the drift are aging, temperature, ambient conditions, and component deterioration. The drift in an instrument may be predicted by performance analysis of components, past experience, environmental tests, and so on.

techniques must be employed to evaluate the repeatability characteristics of the instrument.

Dynamic Response

The dynamic responses of an instrument is characterized by its natural frequency, amplitude, frequency response, phase shift, linearity and distortions, rise and settling times, slew rate, and the like. These characteristics are a common theme in many instrumentation, control, and electronics books. Although sufficient analysis will be given here, the detailed treatment of the topic can be very lengthy and complex; hence, the full treatment of this topic is not within the scope of this article. Interested readers should refer to the Bibliography (e.g., Doebelin, 1990).

The dynamic response of an instrument can be linear or nonlinear. Fortunately, most instruments exhibit linear characteristics, leading to simple mathematical models by using differential equations such as

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_0 y = x(t) \quad (1)$$

where x is the input variable or the forcing function, y is the output variable, and $a_n, a_{n-1}, \dots, a_0$ are the coefficients or the constants of the system.

The dynamic response of instruments can be categorized as zero-order, first-order, or second-order responses. Although higher-order instruments may exist, their behaviors can be understood adequately in the form of a second-order system. From Eq. (1)

$$a_0 y = x(t) \quad \text{zero order} \quad (2)$$

$$a_1 \frac{dy}{dt} + a_0 y = x(t) \quad \text{first order} \quad (3)$$

$$a_2 \frac{d^2 y}{dt^2} + a_1 \frac{dy}{dt} + a_0 y = x(t) \quad \text{second order} \quad (4)$$

Equations (2)–(4) can be written as Laplace transforms, thus enabling analysis in the frequency domain,

$$\frac{Y(s)}{X(s)} = 1 \quad (5)$$

$$\frac{Y(s)}{X(s)} = \frac{1}{(\tau_1 s + 1)} \quad (6)$$

$$\frac{Y(s)}{X(s)} = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)} \quad (7)$$

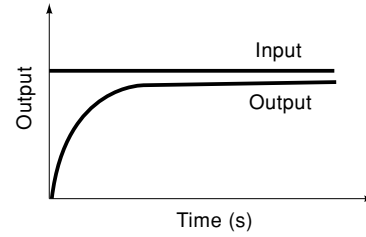


Figure 4. A first-order-hold instrument responds to a step input in an exponential form. For a good response the time delay must be small. Drift is usually expressed in percentage of output.

where s is the Laplace operator and τ is the coefficient, also called time constant.

In zero-order instruments, there is no frequency dependence between the input and output. The amplitude change is uniform across the spectrum of all possible frequencies. In practice, such instruments are difficult to obtain, except in a limited range of operations.

In first-order instruments, the relation between the input and the output is frequency dependent. Figure 4 illustrates the response of a first-order instrument for a unit step input in the time domain. Mathematically, the output may be written as

$$y(t) = K e^{-t/\tau} \quad (8)$$

where K and τ are constants determined by the system parameters. In many cases, the input signals may be a complex rather than a simple step input. In the analysis, we need to multiply the transfer function, the second member of Eq. (6), by the Laplace transform of the input signal and then transform it back to the time domain if we are to understand the nature of transient and steady-state responses. Also, if the first-order systems are cascaded, the relative magnitudes of the time constants become important; some may be dominant, and others may be neglected.

Second-order systems exhibit the laws of simple harmonic motion, which can be described by linear wave equations. Equation (7) may be rearranged as

$$\frac{X(s)}{Y(s)} = \frac{1/a_0}{s^2/\omega_n^2 + 2\zeta s/\omega_n + 1} \quad (9)$$

where ω_n is the natural or undamped frequency (rad/s) and ζ is the damping ratio.

As can be seen, the performance of instruments become a function of natural frequency and the damping ratio of the system. The natural frequency and damping ratios are related to the physical parameters of the devices, such as mass and dimensions. In the design stages, these physical parameters may be selected, tested, and modified to obtain a desired response from the system.

Typical time response of a second-order system to unit step inputs is illustrated in Fig. 5. The response here indicates that a second-order system oscillates or it can even be unstable. Furthermore, we can deduce that, since the second-order system is dependent on time, wrong readings can be made depending on the time that the results are taken. Clearly, recording the output when the instrument is still under transient conditions will give an inadequate representation of the physical variable.

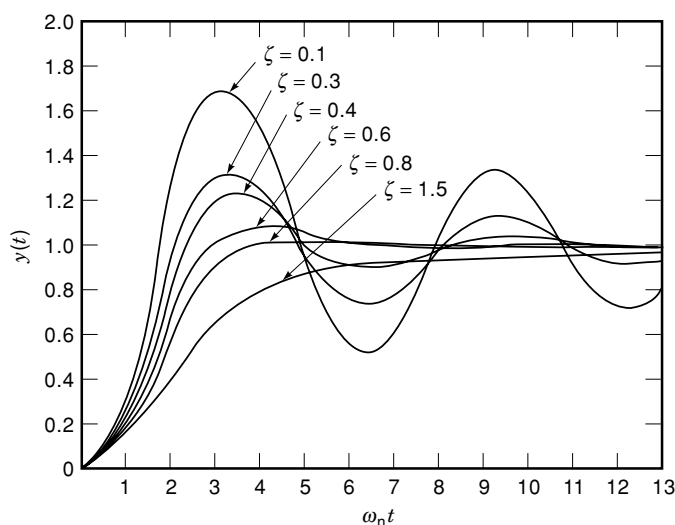


Figure 5. Unit step time responses of a second-order system with various damping ratios. The maximum overshoot, delay, rise, settling times and frequency of oscillation depend on the damping ratio. A smaller damping ratio gives a faster response but larger overshoot. In many applications, a damping ratio of 0.707 is preferred.

The frequency compensation, selection of appropriate damping, acceptable time responses, and rise time and settling times of instruments may need careful attention in both the design and application stages. In these systems, system analysis is essential to ensure that they can measure the input measurand adequately.

A typical frequency dependence of gain and phase angle between input and output is illustrated in Fig. 6 in the form of Bode diagrams. Here, the *bandwidth*, which is the frequencies over which the gain is reasonably constant, is also shown. Usually, half power point (3 dB), which symbolizes 70.7% of the maximum value, is taken as the bandwidth.

An important concept in instruments is *response time*, which can be described as the time required for the instrument to respond to an input signal change. For automatic measurements, the response time is an indication of how many readings can be done per second. Response time is affected by many factors such as analog-to-digital (A/D) conversion times, settling times, delays in electronic components, and delays in sensors.

MEASUREMENT ERRORS AND ERROR CONTROL SYSTEMS

The performance of an instrument depends on its static and dynamic characteristics. The performance may be indicated by its *accuracy*, which may be described as the closeness of measured values to the real values of the variable. The total response is a combination of dynamic and static responses. If the signals generated by the physical variable are changing rapidly, then the dynamic properties of the instrument become important. For slowly-varying systems, the dynamic errors may be neglected. In order to describe the full relationships between the inputs and outputs, differential equations can be used, as discussed previously.

The performance of an instrument may also be decided by other factors, such as the magnitudes of errors; the *repeatability*, which indicates the closeness of sets of measurements

made in the short term; and the *reproducibility* of the instrument. The reproducibility is the closeness of sets of measurements when repeated in similar conditions over a longer period of time.

The ideal or perfect instrument would have perfect sensitivity, reliability, and repeatability without any spread of values and would be within the applicable standards. However, in many measurements, there will be imprecise and inaccurate results due to many internal and external factors. The departure from the expected perfection is called the *error*. Often, sensitivity analyses are conducted to evaluate the effect of individual components that are causing these errors. Sensitivity to the affecting parameter can be obtained by varying that one parameter and keeping the others constant. This can be done practically by using the developed instruments or mathematically by means of appropriate models.

When determining the performance of an instrument, it is essential to appreciate how errors arise. There may be many sources of errors; therefore, it is important to identify these sources and draw up an error budget. In the error budget, there may be many factors, such as (1) imperfections in electrical and mechanical components (e.g., high tolerances and noise or offset voltages), (2) changes in component performances (e.g., shift in gains, changes in chemistry, aging, and drifts in offsets), (3) external and ambient influences (e.g., temperature, pressure, and humidity), and (4) inherent physical fundamental laws (e.g., thermal and other electrical noises, Brownian motion in materials, and radiation).

In instrumentation systems, errors can be broadly classified as systematic, random, or gross.

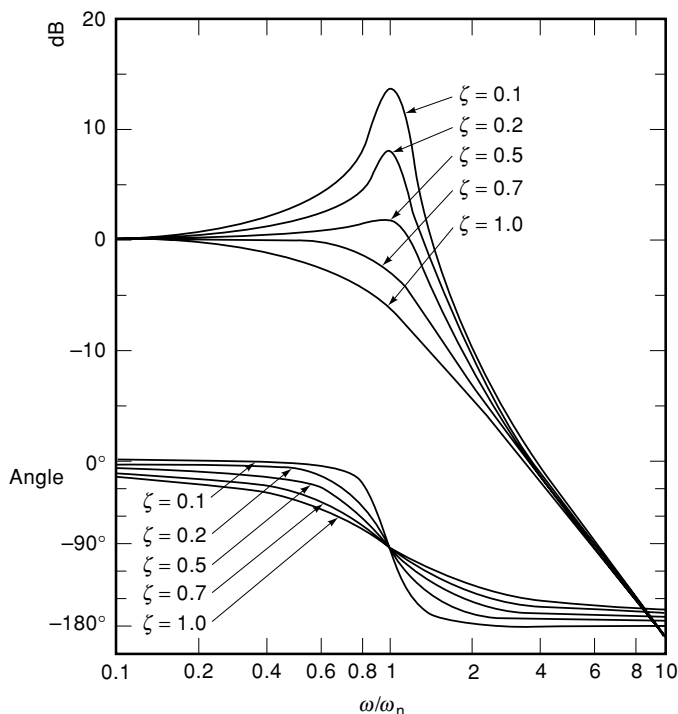


Figure 6. Bode plots of gains and phase angles against frequency of a second-order system. Curves are functions of frequencies as well as damping ratios. These plots can be obtained theoretically or by practical tests conducted in the frequency range of interest.

Systematic Errors

Systematic errors remain constant with repeated measurements. They can be divided into two basic groups: the instrumental errors and the environmental errors. Instrumental errors are inherent within the instrument, arising because of the mechanical structure, electronic design, improper adjustments, wrong applications, and so on. They can also be subclassified as loading error, scale error, zero error, and response time error. The environmental errors are caused by environmental factors such as temperature and humidity. Systematic errors can also be viewed as static or dynamic errors.

Systematic errors can be quantified by mathematical and graphical means. They can be caused by the nonlinear response of the instrument to different inputs as a result of hysteresis. They also emerge from wrong biasing, wear and aging, and other factors such as modifying the effects environment (e.g., interference). Typical systematic error curves are illustrated in Fig. 7.

Because of the predicability of systematic errors, deterministic mathematics can be employed. In the simplest form, the error of a measurement may be expressed as

$$\Delta x(t) = x_m(t) - x_r(t) \quad (10)$$

where $\Delta x(t)$ is the absolute error, $x_r(t)$ is the correct reference value, and $x_m(t)$ is the measured value.

From Eq. (10), the relative error $r_e(t)$ may be calculated as

$$r_e(t) = \frac{\Delta x(t)}{x_r(t)} \quad (11)$$

However, in complex situations, correction curves obtained either empirically or theoretically may be used. Manufacturers usually supply correction curves, especially if their products embrace wide ranging and different applications (e.g., slurries with changing characteristics in time).

In many applications, the measurement system is made up of many components that have errors in their own rights. The

deterministic approach may be adapted to calculate the overall propagated error of the system, as

$$y = f(x_1, x_2, x_3, \dots, x_n) \quad (12)$$

where y is the overall output and $x_1, x_2, \dots$ are the components affecting the output.

Each variable affecting the output will have its own absolute error of Δx_i . The term Δx_i indicates the mathematically or experimentally determined error of each component under specified operating conditions. The performance of the overall system with the errors may be expressed as

$$y \pm \Delta y = f(x_1 \pm \Delta x_1, x_2 \pm \Delta x_2, \dots, x_n \pm \Delta x_n) \quad (13)$$

For an approximate solution, the Taylor series may be applied to Eq. (13). By neglecting the higher-order terms of the series, the total absolute error Δy of the system may be written as

$$\Delta y = |\Delta x_1 \delta y / \delta x_1| + |\Delta x_2 \delta y / \delta x_2| + \dots + |\Delta x_n \delta y / \delta x_n| \quad (14)$$

The absolute error is predicted by measuring or calculating the values of the errors of each contributing component.

Slight modification of Eq. (13) leads to uncertainty analysis, where

$$w_y = [(w_1 \delta y / \delta x_1)^2 + (w_2 \delta y / \delta x_2)^2 + \dots + (w_n \delta y / \delta x_n)^2]^{1/2} \quad (15)$$

where w_y is the uncertainty of the overall system and $w_1, w_2, \dots, w_n$ are the uncertainties of affecting the component.

Uncertainty differs from error in that it involves such human judgemental factors as estimating the possible values of errors. In measurement systems, apart from the uncertainties imposed by the instruments, experimental uncertainties also exist. In evaluating the total uncertainty, several alternative measuring techniques should be considered and assessed, and the estimated accuracies must be worked out with care.

Random and Gross Errors

Random errors appear as a result of rounding of reading, noise and interference, backlash and ambient influences, and so on. In experiments, the random errors vary by small amounts around a mean value. Therefore, the future value of any individual measurement cannot be predicted in a deterministic manner. Random errors may not easily be offset electronically; therefore, in the analysis and compensation, stochastic approaches are adapted by using the laws of probability.

Depending on the system, the random error analysis may be made by applying different probability distribution models. But, most instrumentation systems obey normal distribution laws; therefore, the Gaussian model can broadly be applied enabling the determination of the mean values, standard deviations, confidence intervals, and the like, depending on the number of samples being taken. A typical example of a Gaussian curve is given in Fig. 8. The mean value $\bar{x}$ and the standard deviation σ may be found by

$$\bar{x} = \frac{\sum x_i}{n} \quad (16)$$

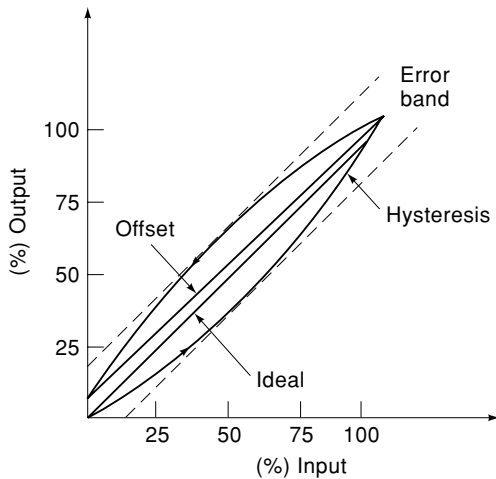


Figure 7. Systematic errors are static errors and they can be quantified theoretically or experimentally. There are many different types, including hysteresis, linearity, and offset. They are contained within an error band typical to the particular instrument.

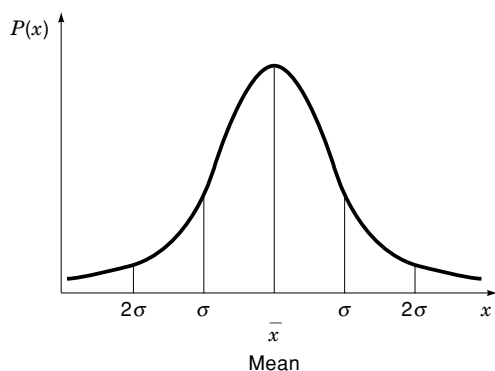


Figure 8. Random errors of instruments can be analyzed by using probability methods. In many instruments, the errors can be described by a Gaussian distribution curve.

and

$$\sigma = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad (17)$$

Discussions relating to the application of stochastic theory in error analysis are very lengthy and will not be repeated here. Interested readers should refer to the Bibliography (e.g., Holman, 1989).

Gross errors are the result of human mistakes, equipment fault, and the like. Human errors may occur in the process of observations or during the recording and interpretation of experimental results. A large number of errors can be attributed to carelessness, the improper adjustment of instruments, the lack of knowledge about the instrument and the process, and so on. These errors cannot be treated mathematically and cannot be eliminated completely, but they can be minimized by having different observers repeat the experiments.

Error Reduction Techniques

Controlling errors is an essential part of instruments and instrumentation systems. Various techniques are available to achieve this objective. The error control begins in the design stages by choosing the appropriate components, filtering, and bandwidth selection, by reducing the noise, and by eliminating the errors generated by the individual subunits of the complete system. In a good design, the errors of the previous group may be compensated adequately by the following groups.

The accuracy of instruments can be increased by postmeasurement corrections. Various calibration methods may be employed to alter parameters slightly to give correct results. In many cases, calibration graphs, mathematical equations, tables, the experiences of the operators, and the like are used to reduce measurement errors. In recent years, with the application of digital techniques and intelligent instruments, error corrections are made automatically by computers or the devices themselves.

In many instrumentation systems, the application of compensation strategy is used to increase static and dynamic performances. In the case of static characteristics, compensations can be made by many methods including the introducing opposing nonlinear elements in the system, using isolation and

zero environmental sensitivity, opposing compensating environmental inputs, using differential systems, and employing feedback systems. On the other hand, the dynamic compensation can be achieved by applying these techniques as well as by reducing harmonics, using filters, adjusting bandwidths, using feedback compensation techniques, and the like.

Open loop and close loop dynamic compensations are popular methods employed in both static and dynamic error corrections. For example, using high-gain negative feedback can reduce the nonlinearity generated by the system. A recent and fast-developing trend is the use of computers for estimating measured values and providing compensation during the operations if any deviations occur from the estimated values.

STANDARDS AND REFERENCE MATERIALS

Standards of fundamental units of length, time, weight, temperature, and electrical quantities have been developed for measurements to be consistent all over the world. The length and weight standards—the meter and the kilogram—are kept in the International Bureau of Weights and Measures in Sèvres, France. Nevertheless, in 1983 the meter was defined as the length of the path traveled by light in vacuum in the fraction $1/299,792,458$ of a second, which was adopted as the standard meter. The standard unit of time—second—is established in terms of known oscillation frequencies of certain devices, such as the radiation of the cesium-133 atom. The standards of electrical quantities are derived from mechanical units of force, mass, length, and time. Temperature standards are established as international scale by taking eleven primary fixed points. If different units are involved, the relationship between different units are defined in fixed terms. For example, $1 \text{ lb} = 453.59237 \text{ g}$.

Based on these standards, primary international units, SI (Système International d'Unités), are established for mass, length, time, electric current, luminous intensity, and temperature, as illustrated in Table 2. From these units, SI units of all physical quantities can be derived as exemplified in Table 3. The standard multiplier prefixes are illustrated in Table 4.

In addition to primary international standards, standard instruments are available having stable and precisely defined characteristics that are used as references for other instruments that are performing the same function. Hence, the performance of an instrument can be cross-checked against a known device. At a global level, checking is done by using an international network of national and international laboratories,

Table 2. Basic SI Units

Quantity	Unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd
Plane angle	radian	rad
Solid angle	steradian	sr

Table 3. Fundamental, Supplementary, and Derived Units

Quantity	Symbol	Unit Name	Unit Symbol
<i>Mechanical Units</i>			
Acceleration	<i>a</i>	Meter/second ²	m/s ²
Angular acceleration	<i>a</i>	Radian/second ²	rad/s ²
Angular frequency	<i>g</i>	Radian/second	rad/s
Angular velocity	<i>g</i>	Radian/second	rad/s
Area	<i>A</i>	Square meter	m ²
Energy	<i>E</i>	Joule	J(kg · m ² /s ²)
Force	<i>F</i>	Newton	N(kg · m/s ²)
Frequency	<i>f</i>	Hertz	Hz
Gravitational field strength	<i>g</i>	Newton/kilogram	N/kg
Moment of force	<i>M</i>	Newton · meter	N · m
Plane angle	<i>a, b, θ, φ</i>	Radian	Rad
Power	<i>P</i>	Watt	W(J/s)
Pressure	<i>p</i>	Newton/meter ²	N/m ²
Solid angle	<i>g</i>	Steradian	Sr
Torque	<i>T</i>	Newton meter	N · m
Velocity	<i>v</i>	Meter/second	m/s
Volume	<i>V</i>	Cubic meter	m ³
Volume density	<i>ρ</i>	Kilogram/meter ³	kg/m ³
Wavelength	<i>λ</i>	Meter	m
Weight	<i>W</i>	Newton	N
Weight density	<i>c</i>	Newton/cubic meter	N/m ³
Work	<i>w</i>	Joule	J
<i>Electrical Units</i>			
Admittance	<i>Y</i>	Mho (siemen)	mho (S)
Capacitance	<i>C</i>	Farad	F(A · s/V)
Conductance	<i>G</i>	Mho(siemen)	mho(S)
Conductivity	<i>c</i>	Mho/meter	mho/m(S/m)
Current density	<i>J</i>	Ampere/meter ²	A/m ²
Electric potential	<i>V</i>	Volt	V
Electric field intensity	<i>E</i>	Volt/meter	V/m
Electrical energy	<i>W</i>	Joule	J
Electrical power	<i>P</i>	Watt	W
Impedance	<i>Z</i>	Ohm	Ω
Permittivity of free space	<i>ε</i>	Farad/meter	F/m
Quantity of electricity	<i>Q</i>	Coulomb	C(A · s)
Reactance	<i>X</i>	Ohm	Ω
Resistance	<i>R</i>	Ohm	Ω
Resistivity	<i>ρ</i>	Ohm · meter	Ω · m
<i>Magnetic Units</i>			
Magnetic field intensity	<i>H</i>	Ampere/meter	A/m
Magnetic flux	<i>Φ</i>	Weber	Wb
Magnetic flux density	<i>B</i>	Tesla (weber/meter ²)	T (Wb/m ²)
Magnetic permeability	<i>e</i>	Henry/meter	H/m
Mutual inductance	<i>M</i>	Henry	H
Permeability of free space	<i>e₀</i>	Henry/meter	H/m
Permeance	<i>P</i>	Henry	H
Relative permeability	<i>e_r</i>	—	—
Reluctance	<i>R</i>	Henry ⁻¹	H ⁻¹
Self inductance	<i>L</i>	Henry	H
<i>Optical Units</i>			
Illumination	<i>lx</i>	Lux	cd · sr/m ²
Luminous flux	<i>lm</i>	Lumen	cd · sr
Luminance	<i>cd</i>	Candela/meter ²	cd/m ²
Radiance	<i>L_e</i>	Watt/steradian · meter ³	W/sr · m ³
Radiant energy	<i>W</i>	Joule	J
Radiant flux	<i>P</i>	Watt	W
Radiant intensity	<i>I_e</i>	Watt/steradian	W/sr

such as the National Bureau of Standards (NBS), the National Physical Laboratory (NPL), and the Physikalisch-Technische Bundesanstalt of Germany. A treaty among the world's national laboratories regulates the international activity and coordinates development, acceptance, and intercomparisons. Standards are kept in four stages:

Table 4. Decimal Multiples

Name	Symbol	Equivalent
Exa	E	10 ¹⁸
Peta	P	10 ¹⁵
Tera	T	10 ¹²
Giga	g	10 ⁹
Mega	M	10 ⁶
Kilo	k	10 ³
Hecto	h	10 ²
Deca	da	10
Deci	d	10 ⁻¹
Centi	c	10 ⁻²
Milli	m	10 ⁻³
Micro	e	10 ⁻⁶
Nano	n	10 ⁻⁹
Pico	p	10 ⁻¹²
Femto	f	10 ⁻¹⁵
Atto	a	10 ⁻¹⁸

1. *International standards* represent certain units of measurement with maximum accuracy possible within today's available technology. These standards are under the responsibility of an international advisory committee and are not available to ordinary users for comparison or calibration purposes.
2. *Primary standards* are the national standards maintained by national laboratories in different parts of the world for verification of secondary standards. These standards are independently calibrated by absolute measurements that are periodically made against the international standards. The primary standards are compared against each other.
3. *Secondary standards* are maintained in the laboratories of industry and other organizations. They are periodically checked against primary standards and certified.
4. *Working standards* are used to calibrate general laboratory and field instruments.

Another type of standard is published and maintained by the IEEE in New York. These standards are for test procedures, safety rules, definitions, nomenclature, and so on. The IEEE standards are adopted by many organizations around the world. Many nations also have their own standards for test procedures, instrument usage procedures, safety, and the like.

CALIBRATION, CALIBRATION CONDITIONS, AND THE LINEAR CALIBRATION MODEL

The calibration of all instruments is essential for checking their performances against known standards. This provides consistency in readings and reduces errors, thus validating the measurements universally. After an instrument is calibrated, future operation is deemed to be error-bounded for a given period of time for similar operational conditions. The calibration procedure involves comparison of the instrument against primary or secondary standards. In some cases, it may be sufficient to calibrate a device against another one with a known accuracy.

Many nations and organizations maintain laboratories with the primary functions of calibrating instruments and field measuring systems that are used in everyday operations. Examples of these laboratories are National Association of Testing

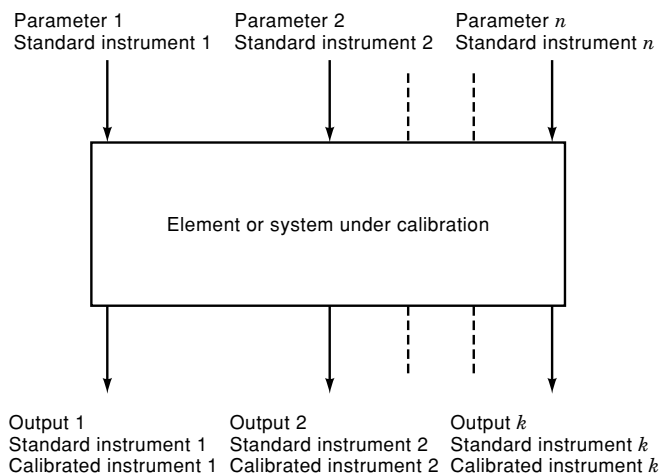


Figure 9. Instruments are frequently calibrated sequentially for all inputs. Calibrations are made under static or dynamic conditions, usually keeping all inputs constant and varying only one and observing the output. The calibration continues until all other inputs are covered.

Authorities (NATA) of Australia and the British Calibration Services (BCS).

Calibrations may be made under static or dynamic conditions. A typical calibration procedure of a complex process involving many instruments is illustrated in Fig. 9. In an ideal situation, for instruments that respond to a multitude of physical variables, a commonly employed method is to keep all inputs constant except one. The input is varied in increments in increasing and decreasing directions over a specified range. The observed output then becomes a function of that single input. The calibration is continued in a similar manner until all other inputs are covered. For better results, this procedure may be repeated by varying the sequences of inputs thus developing a family of relationships between the inputs and outputs. As a result of these calibration readings, the input and output relation usually demonstrates statistical characteristics. From these characteristics, appropriate calibration curves can be obtained, and other statistical techniques can be applied.

In many instruments, the effect of a single input may not represent the true output values when one input is varied and all others are kept constant. In these cases, calibration is conducted by varying several inputs simultaneously. Throughout the calibration procedure, the n number of variables of the system are monitored by appropriate standard instruments. The rule of thumb is that each calibrated variable must have a traceable ladder starting from laboratory standards and secondary standards leading to primary standards. This is known as the linear calibration model or traceability.

Most instrument manufacturers supply calibrated instruments and reliable information about their products. But their claims of accuracy and reliability must be taken at face value. Therefore, in many cases, application-specific calibrations must be made periodically within the recommended calibration intervals. Usually, manufacturers supply calibration programs. In the absence of such programs, it is advisable to conduct frequent calibrations in the early stages of installation and lengthen the period between calibrations as the confidence builds based on satisfactory performance. Recently, with the wide applications of digital systems, computers can make au-

tomatic and self calibrations as in the case of many intelligent instruments. In these cases, post measurement corrections are made, and the magnitudes of various errors are stored in the memory to be recalled and used in laboratory and field applications.

ANALOG AND DIGITAL INSTRUMENTS

Instruments can be analog or digital or a combination of the two. Nowadays, most instruments are produced to be digital because of the advantages that they offer. However, the front end of many of instruments are still analog; that is, the majority of sensors and transducers generate analog signals. Initially, the signals are conditioned by analog circuits before they are put into digital form for signal processing. It is important to mention that digital instruments operating purely on digital principles are developing fast. For instance, today's smart sensors contain the complete signal condition circuits in a single chip integrated with the sensor itself. The output of smart sensors can be interfaced directly with other digital devices.

In analog instruments, the useful information is conveyed by changes in amplitudes, phases, or frequencies or a combination of the three. These signals can be deterministic or non-deterministic. In all analog or digital instruments, as in the case with all signal-bearing systems, there are useful signals that respond to the physical phenomena and unwanted signal resulting from various forms of noise. In the case of digital instruments, additional noise is generated in the process of A/D conversion.

Analog signals can also be nondeterministic; that is, the future state of the signal cannot be determined. If the signal varies in a probabilistic manner, its future can be foreseen only by statistical methods. The mathematical and practical treatment of analog and digital signals having deterministic, stochastic, and nondeterministic properties is a very lengthy subject and a vast body of information can be found in literature; therefore, they will not be treated here.

As is true of all instruments, when connecting electronic building blocks, it is necessary to minimize the loading effects of each block by ensuring that the signal is passed without attenuation, loss, or magnitude and phase alterations. It is also important to ensure maximum power transfer between blocks by appropriate impedance-matching techniques. Impedance matching is very important in all instruments but particularly at a frequency of 1 MHz and above. As a rule of thumb, output impedances of the blocks are usually kept low and input impedances are kept high so that the loading effects can be minimized.

Analog Instruments

Analog instruments are characterized by continuous signals. A purely analog system measures, transmits, displays, and stores data in analog form. The signal conditioning is usually made by integrating many functional blocks such as bridges, amplifiers, filters, oscillators, modulators, offsets and level converters, buffers, and the like, as illustrated Fig. 10. Generally, in the initial stages, the signals produced by the sensors and transducers are conditioned mainly by analog electronics, even if they are configured as digital instruments later. Therefore, we pay more attention to analog instruments, keeping in mind that much of the information given here also may be used in various stages of the digital instruments.

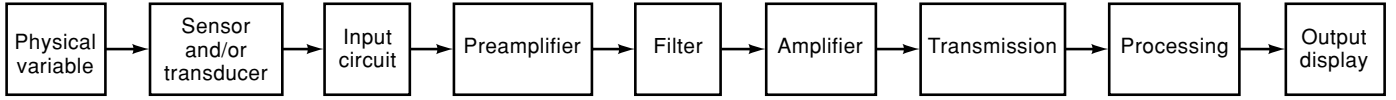


Figure 10. Analog instruments measure, transmit, display, and store data in analog form. The signal conditioning usually involves such components as bridges, amplifiers, filters, oscillators, modulators, offsets and level converters, buffers, and so on. These components are designed and tested carefully to suit the characteristics of a particular instrument.

Instrument bridges are commonly used to measure such basic electrical quantities as resistance, capacitance, inductance, impedance, and admittance. Basically, they are two-port networks in which the component to be measured is connected to one of the branches of the network. There are two basic groups, ac and dc bridges. Also, there are many different types in each group, such as Wheatstone and Kelvin dc bridges and Schering, Maxwell, Hay, and Owen ac bridges. In a particular instrument, the selection of the bridge to be employed and the determination of values and tolerances of its components is very important. It is not our intent to cover all bridges here; however, as typical example of an ac bridge, a series RC bridge is given in Fig. 11. We also offer some analysis to illustrate briefly their typical operational principles. At balance,

$$Z_1 Z_3 = Z_x Z_2 \quad (18)$$

Substitution of impedance values gives

$$R_3(R_1 - j/\omega C_1) = (R_x - j/\omega C_x)R_2 \quad (19)$$

Equating the real and imaginary terms gives the values of unknown components as

$$R_x = \frac{R_1 R_3}{R_2} \quad (20)$$

and

$$C_x = \frac{C_1 R_2}{R_3} \quad (21)$$

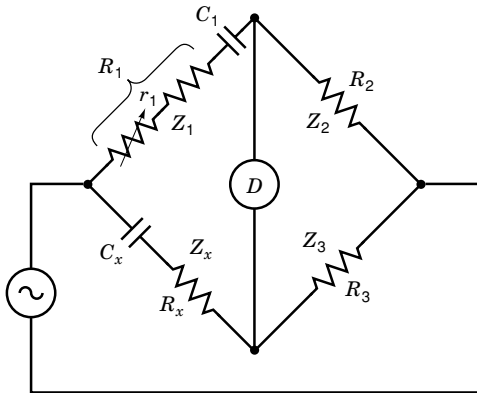


Figure 11. A series RC bridge wherein the unknown capacitance is compared with a known capacitance. The voltage drop across R_1 balances the resistive voltage drop in branch Z_2 . The bridge balance is achieved relatively easily when capacitive branches have substantial resistive components. The resistors R_1 and either R_2 or R_3 are adjusted alternately to obtain the balance.

In instruments, the selection and use of amplifiers and filters are also very important since many transducers generate extremely weak signals in comparison to the noise existing in the device. Today, operational amplifiers and high-precision instrumentation amplifiers form the building blocks of modern instruments.

The operation amplifiers may be used as inverting and noninverting amplifiers, and by connecting suitable external components, they can be configured to perform many other functions, such as multipliers, adders, limiters, and filters. Instrumentation amplifiers are used in situations where operational amplifiers do not meet the requirements. They are essentially high-performance differential amplifiers consisting of several closed-loop operational amplifiers. The instrumentation amplifiers have improved common mode rejection ratios (CMRR) (up to 160 dB), high input impedances (up to 500 M Ω), low output impedance, low offset currents and voltages and better temperature characteristics. To illustrate amplifiers in instrumentation systems, a typical current amplifier used in charge amplification is illustrated in Fig. 12. In this circuit, if the input impedance of the operational amplifier is high, output is not saturated, and the differential input voltage is small, it is possible to write

$$\frac{1}{C_f} \int i_f dt = e_{ex} - e_{ai} = e_{ex} \quad (22)$$

$$\frac{1}{C_x} \int i_x dt = e_0 - e_{ai} = e_0 \quad (23)$$

$$i_f + i_x - i_{ai} = 0 = i_f + i_x \quad (24)$$

Manipulation of these equations gives

$$e_0 = \frac{-C_f e_{ex}}{C_x} \quad (25)$$

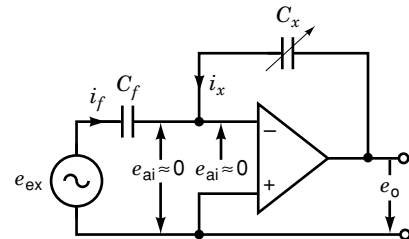


Figure 12. Using an operational amplifier signal processor is useful to eliminate the nonlinearity in the signals generated by capacitive sensors. With this type of arrangement, the output voltage can be made to be directly proportional to variations in the signal representing the nonlinear operation of the device.

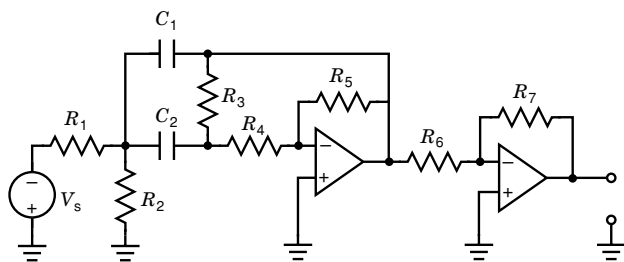


Figure 13. Filtering is used in various stages of signal processing to eliminate unwanted components of signals. They can be designed and constructed to eliminate or pass signals at certain frequency ranges. Suitable arrangements of components yield to lowpass, highpass, band-pass, bandstop and notch filters. Filters can be classified as active and passive.

However, a practical circuit requires a resistance across C_f to limit output drift. The value of this resistance must be greater than the impedance of C_f at the lowest frequency of interest.

Filtering is used to reject unwanted components of signals. For example, by using a filter that narrows the bandwidth, the broadband noise energy is reduced, and unwanted signals outside the passband are rejected. Analog filters can be designed by using various techniques, such as Butterworth, Chebyshev, and Bessel-Thomson filters etc. They can be lowpass, highpass, bandpass, bandstop, and notch filters. Filters can be classified as active filters or passive filters. Active filters involve active components such as operational or instrumentation amplifiers, whereas passive filters are configured completely by inductive, capacitive, and resistive components. The choice of active or passive filters depends on the available components, the precision required, and the frequency of operations. A typical filter used in instrument is given in Fig. 13.

Digital Instruments

In modern instruments, the original data acquired from the physical variables are usually in analog form. This analog signal is converted to digital before being passed on to the other parts of the system. For conversion purposes, analog-to-digital converters are used together with appropriate sample-and-hold devices. In addition, analog multiplexers enable the connection of a number of transducers to the same signal-processing media. The typical components of a digital instrument are illustrated in Fig. 14. The digital systems are particularly useful in performing mathematical operations and storing and transmitting data.

Analog-to-digital conversion involves three stages: sampling, quantization, and encoding. The Nyquist sampling theorem must be observed during sampling; that is, “the

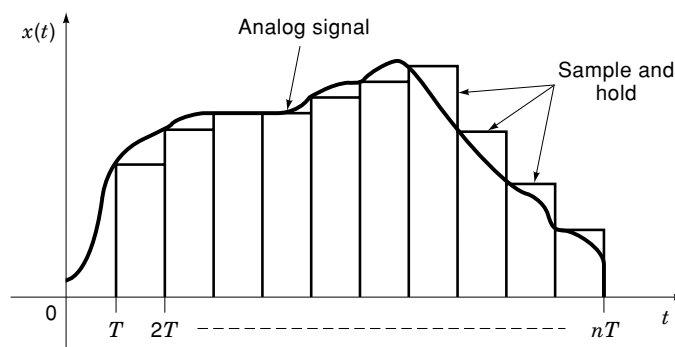


Figure 15. Analog-to-digital converters involve three stages: sampling, quantization, and encoding. However, the digitization introduces a number of predictable errors. After the conversion, the data can be processed by techniques such as FFT analysis, DFT analysis, digital filtering, sequential or logical decision making, correlation methods, spectrum analysis, and so on.

number of samples per second must be at least twice the highest frequency present in the continuous signal.” As a rule of thumb, depending on the significance of the high frequencies, the sampling must be about five to ten times of the highest frequency of the signal. The next stage is the quantization, which determines the resolution of the sampled signals. The quantization error decreases as the number of bits increases. In the encoding stage, the quantized values are converted to binary numbers to be processed digitally. Figure 15 illustrates a typical A/D sampling process of an analog signal.

After the signals are converted to digital forms, the data can be further processed by employing various techniques as FFT analysis, digital filtering, sequential or logical decision making, correlation methods, spectrum analysis, and so on.

Virtual Instruments (VIs)

Traditional instruments have three basic components—acquisition and control, data analysis, and data presentation. In VIs, the use of digital techniques, software, and computers replace the display and processing capabilities of most traditional instruments. In this technology, plug-in data acquisition (DAQ) boards, PC cards (PCMCIA), and parallel port I/O devices are used to interface sensors and transducers of the system under investigation to computers. There are standard interface buses such as VXIbus, which stands for VMEbus Extensions for Instrumentation (also known as the IEEE Standard 1155-1992).

Once the system is interfaced, the computer can be programmed to act just like a stand-alone instrument, but offering additional benefits of flexibility of the processing, display, and storage. In VIs, the data can be saved or loaded in memory to

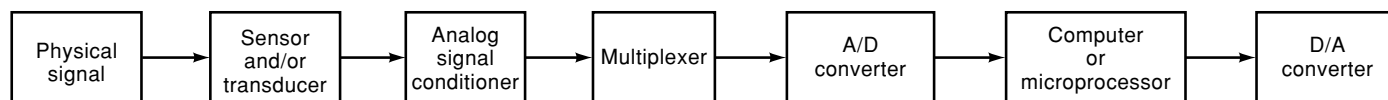


Figure 14. Digital instruments have more signal-processing components than analog instruments. Usually, analog signals are converted to digital form by analog-to-digital (A/D) converters. The digital instruments have the advantage of processing, storing, and transmitting signals more easily than their analog counterparts.

be processed in popular spreadsheet programs and word processors, and a report generation capability complements the raw data storage by adding timestamps, measurements, user names, and comments.

VI technology allows the user to build test systems that fit specific applications. The VI software can be programmed to resemble familiar instrument panels, including buttons and dials. The user interface tools include knobs, meters, gauges, dials, tank displays, thermometers, graphs, strip charts, and the like to simulate the appearance of traditional instruments. Computer displays can show more colors, and allow users to quickly change the way they display test data and controls as required. The software also contains analysis libraries with high-powered statistics, curve fitting, signal processing, and filtering to standard dynamic link libraries (DLLs).

Designing a VI system is similar to designing a test system with stand-alone instruments. The first step is to determine what types of signals are needed to measure, including their frequencies, amplitudes, and other signal characteristics together with the level of accuracy expected from these signals. To develop the software for the test application, a programming language or test-development software package needs to be selected such as C or Microsoft Visual Basic. Since the display is not fixed, as on a stand-alone instrument, it can be as complex or as simple as the application requires.

Nowadays, users can configure their VIs to update front panels and display real-time, animated VIs over the Internet. The toolkits let applications be published over the Web and viewed with a standard Web browser with little additional programming. With these tools, developers can monitor VIs running in remote locations, publish experiment results on the Web, and automatically notify operators of alarm conditions or status information.

CONTROL OF INSTRUMENTS

Instruments can be manual, semiautomatic, or fully automatic. Manual instruments need human intervention for adjustment, parameter setting, and interpreting the readings. Semiautomatic instruments need limited intervention such as the selection of operating conditions and so on. In the fully automatic instruments, however, the variables are measured either periodically or continuously without human intervention. The information is either stored or transmitted to other devices automatically. Some of these instruments can also measure the values of process variables and regulate their deviations from preset points.

It is often necessary to measure many parameters of a process by using two or more instruments. The resulting arrangement for performing the overall measurement function is called the measurement system. In measurement systems, instruments operate in an autonomously but coordinated manner. The information generated by each device is communicated between instruments themselves, or between the instrument and other devices such as recorders, display units, and computers. The coordination of instruments can be done in three ways: analog-to-analog, analog-to-digital, and digital-to-digital.

Analog systems consist of instruments that generate continuous current and voltage waveforms in response to the physical variations. The signals are processed by using analog

electronics; therefore, signal transmission between the instruments and other devices is also done in the analog form. In assembling these devices, the following characteristics must be considered:

- Signal transmission and conditioning
- Loading effects and buffering
- Proper grounding and shielding
- Inherent and imposed noises
- Ambient conditions
- Signal level compatibility
- Impedance matching
- Proper display units
- Proper data storage media

Offset and level conversion is used to convert the output signal of an instrument from one level to another, compatible with the transmission medium in use. In analog systems, signals are usually transmitted at suitable current levels (4–20 mA). In this way, change in impedance does not affect the signal levels, and standard current signal levels can easily be exchanged.

In digital instrumentation systems, analog data are converted and transmitted in digital form. The transmission of data between digital devices can be done relatively easily, by using serial or parallel transmission techniques. However, as the measurement system becomes large by the inclusion of many instruments, the communication becomes complex. To avoid this complexity, message interchange standards are used for digital signal transmission such as RS-232 and IEEE-488, VXIbus.

Many instruments are manufactured with output ports to pass measurement data and various control signals. The IEEE-488 (also known as the GPIB) bus is one of the established industry standard instrumentation interfacing. It enables simultaneous measurements by interfacing up to 15 instruments together at the same time. It has 16 signal lines distributed as 8 data lines, 3 control lines, and 5 general interface management lines. The line configuration of an IEEE-488 bus is given in Fig. 16. Once connected, any one device can transfer data to one or more other devices on the bus. All devices must be able to perform at least one of the following roles: talker, listener, controller. The minimum system consists of one talker and one listener without a controller. The length of cables connected to the bus cannot exceed 20 m, and the maximum data rate is restricted to 250 kilobytes per second.

RS-232 is issued by the Electronic Industries Association (EIA). It uses serial binary data interchange and applies specifically to the interconnection of data communication equipment (DCE) and data terminal equipment (DTM). Data communications equipment may include modems, which are the devices that convert digital signals suitable for transmission through telephone lines. The RS-232 uses standard DB-25 connectors, the pin connection is given in Table 5. Although 25 pins are assigned, a complete data transmission is possible by using only three pins—2, 3 and 7. The transmission speed can be set to certain baud rates such as 19200 bits per second and can be used for synchronous or nonsynchronous communication purposes. The signal voltage levels are very flexible, with any voltage between –3 V to –25 V representing logic 1 and any voltage between +3 V to +25 V representing logic 0.

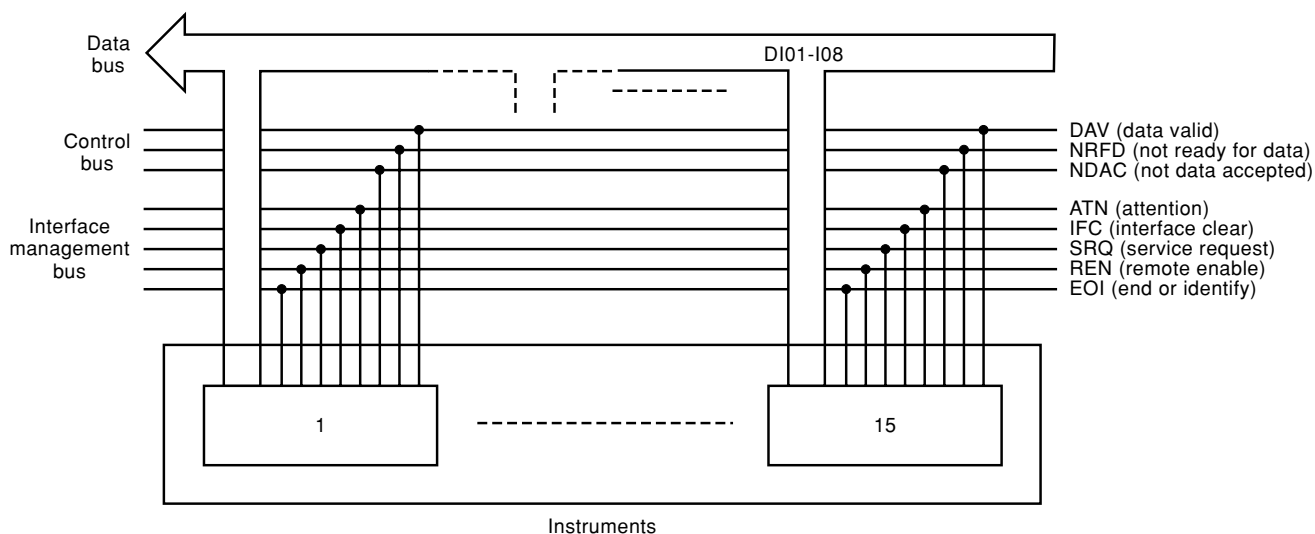


Figure 16. The IEEE-488 or the GPIB bus is an industry standard for interface medium. It has 8 data lines, 3 control lines, and 5 general interface management lines. In noisy environments the maximum length of cable is recommended to be not more than 20 m.

In many industrial applications, the current loop digital communication is used. This communication is similar to analog current loop systems, but the signal is transmitted in digital form, with 20 mA signifying logic 1 and 0 mA representing logic 0. Depending on the external noise sources in the installation environment, the current loop can be extended up to 2 km.

When data are transmitted distances greater than those permitted by the RS-232 or current loop, the modem, microwave, or radiofrequency (RF) transmissions are used. In this case, various signal modulation techniques are necessary to convert digital signals to suitable formats. For example, most modems, with medium-speed asynchronous data transmission, use frequency-shift keyed (FSK) modulation. The digital interface with modems uses various protocols such as MIL-STD-188C to transmit signals in simplex, half-duplex, or full duplex forms depending on the directions of the data flow. The simplex interface transmits data in one direction, whereas full-duplex transmits it in two directions simultaneously.

As far as industrial applications are concerned, several standards for digital data transmission are available, commonly

known as *fieldbuses* in the engineering literature. For example, WordFIP and Profibus have been developed and Foundation Fieldbus is under development to increase the performance of the 20 mA current loop. New devices allow for an increase in the data rates (e.g., National Instruments chips and boards operating with high-speed protocol HS488 for 8 Mbytes/s transfer rate). A new standard is under discussion at the IEEE by the working group for higher performance IEEE Std. 488.1, with a very high increase in the data rate.

Concerning the design software, there are important tools that help implement control (application) software for automatic measuring equipment, such as LabWindows and LabVIEW from National Instruments and VEE from Hewlett-Packard.

In various applications, many instruments (say over a thousand) may be used to monitor and control the process as in the case of computer integrated manufacturing (CIM). In these cases, instruments are networked either in groups or as whole via a central computer or group of computers. Appropriate network topologies (e.g., star, ring, field bus) may be employed to enable the signal flow between the instruments and computers, among the instruments themselves, or between instruments and control panels.

Table 5. RS-232 Pin Connections

Pin Number	Direction	Function
1	—	Frame ground
2	Out	Transmitted data (–TxD)
3	In	Received data (–RxD)
4	Out	Request to send (RTS)
5	In	Clear to send (CTS)
6	In	Data set ready (DSR)
7	—	Signal ground (SG)
8	In	Received line signal detector (DCD)
9	Out	+ Transmit current loop data
11	Out	– Transmit current loop data
18	In	+ Receive current loop data
20	Out	Data terminal ready (DTR)
22	In	Ring indicator (RI)
25	In	– Receive current loop return

INDUSTRIAL MEASURING INSTRUMENTS

In industry, instruments are used to sense and maintain the functions of the process. Because the requirements of diverse industries are different, the instruments are made quite differently to suit applicational differences from one industry to another. Here, instruments specific to some industries will be discussed briefly.

The process industry uses instruments extensively for on-line monitoring and off-line analysis. Specific instruments are used commonly for sensing variables such as temperature, pressure, volumetric and mass flow rate, density, weight, displacement, pH levels, color, absorbency, viscosity, material flow,

dew point, organic and inorganic components, turbidity, solid and liquid level, humidity, and particle size distribution. The selection and use of these instruments constitute an important part of process engineering, which is a discipline in its own right. Additional information can be found in the Bibliography (e.g., Sydenham et al., 1989).

In medical technology, there are three basic types of instruments—imaging, physiological measurements, and laboratory analysis. In imaging and physiological measurements, the instruments are closely linked with patients. Some examples of these instruments are X-ray tomography, nuclear magnetic resonance (NMR) and nuclear spin tomography, ultrasound imaging, thermography, brain and nervous system sensors, and respiratory sensors. Many instruments are based on the radiation and sound, force and tactile sensing, electromagnetic sensing, and chemical and bioanalytical sensors.

Power plants are instrumented for maximum availability, operational safety, and environmental planning. Therefore, their measurements must be as accurate as possible and reliable. Instruments are used for temperature, pressure, flow, level, vibration measurements, and water, steam, and gas analysis. For example, gas analysis requires instruments to measure carbon compounds, sulfur and nitrogen compounds, and dust and ash contents.

Environmental monitoring requires a diverse range of instruments for air, water, and biological monitoring. Instruments are used for measuring various forms of radiation, chemicals hazards, air pollutants, and organic solvents. Many sophisticated instruments are also developed for remote monitoring via satellites, and they operate on optical, microwave, and RF electromagnetic radiation principles.

In automobiles, instruments are used to assist drivers by sensing variables such as cooling, braking, fuel consumption, humidity control, speed, travel route monitoring, and position sensing. Instruments also find applications for safety and security purposes, such as passenger protection and locking and antitheft systems. Recently, with the advent of micromachined sensors, many diverse instruments such as engine control, fuel injection, air regulation, and torque sensing are developed.

The manufacturing industry, especially automated manufacturing, requires a diverse range of instruments. Machine diagnosis and process parameters are made by instruments based on force, torque, pressure, speed, temperature, and electrical parameter-sensing instruments. Optics, tactile arrays, and acoustic scanning instruments are used for pattern recognition. Distance and displacement measurements are made by many methods (e.g., inductive, capacitive, optical, and acoustic techniques).

Aerospace instrumentation requires an accurate indication of physical variables and the changes in these variables. Instruments are designed to suit specific conditions of operations. Some of the measurements are gas temperature and pressure, fluid flow, aircraft velocity, aircraft icing, thrust and acceleration, load, strain and force, position, altitude sensing, and direction finding.

BIBLIOGRAPHY

1. J. P. Bentley, *Principles of Measurement Systems*, 2nd ed., Burnt Mill, UK: Longman Scientific and Technical, 1988.
2. E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.

3. W. Gopel, J. Hesse, and J. N. Zemel, *Sensors—A Comprehensive Survey*, Weinheim, Germany: WCH, 1989.
4. J. P. Holman, *Experimental Methods for Engineers*, 5th ed., New York: McGraw-Hill, 1989.
5. P. H. Sydenham, N. H. Hancock, and R. Thorn, *Introduction to Measurement Science and Engineering*, New York: Wiley, 1989.
6. W. J. Tompkins and J. G. Webster, *Interfacing Sensors to the IBM PC*, Englewood Cliffs, NJ: Prentice-Hall, 1988.

HALIT EREN

Curtin University of Technology

MEASUREMENT ERRORS

In applied science and engineering, it is agreed that all observations contain errors. This article discusses how these errors are described today. It also discusses how this description is used to compute the effect of the errors upon the measurement result, and how this effect is reduced or even minimized by rational use of the observations and by experimental design.

For the description of observations, use is made of a mathematical model. This is a mathematical expression intended to describe the observations fully. It will be supposed throughout this article that the mathematical model is parametric, and that the parameters are the quantities to be measured. For example, the parametric model may be a sinusoidal function of time with unknown amplitude, frequency, and phase. Then, these three quantities are the parameters of this model. Yet the model of the observations is incomplete without including errors. If there is reason to assume that the errors in the observations are nonsystematic and additive, they are taken into consideration by adding a term representing them to the expression for the sinusoidal function. Then, the resulting sum is the mathematical model of the observations. It is supposed to fully describe the observations. Nonsystematic errors may loosely be defined as errors that vary if the experiment is repeated under the same conditions and are equal to zero if averaged over many experiments. They are modeled as stochastic variables with an expectation equal to zero. The term expectation is the abstract mathematical term for the mean value. It will be used throughout to avoid confusion with averages, such as time averages, which are measurements. If, in the example, the errors are stochastic variables, so are the observations. Since each observation is equal to the sum of the sinusoidal function and the stochastic error, the expectation of an observation is the value of the function at the time instant concerned. Therefore, the expectation or, equivalently, the function value represents the hypothetical errorless observation.

Since the observations are stochastic variables, they are described by probability density functions. These define for discrete stochastic variables, such as counting results, the probability of occurrence of a particular discrete outcome. For continuous stochastic variables, the probability density function defines the probability of occurrence of an observation within a particular range of values. The probability density function also determines the expectation of the observations. Since this expectation is equal to the function value, the probability density function of the observations depends on the parameters of the function, and thus, the measurement problem has become

a statistical parameter estimation problem. This observation has important consequences. It implies that for measurement, use can be made of the extensive theory and methods of statistics. It will be seen that this offers a number of exceptional advantages. A description of these advantages requires some familiarity with a number of notions from statistics. Therefore, these will first be introduced. References 1–4 are useful general texts on statistics.

In statistics, the function of the observations with which a parameter is estimated is called an estimator. Using the same observations, for a particular parameter, different estimators can be defined. Since the observations are stochastic variables, so is the estimator. Therefore, the estimator has a probability density function, an expectation, and a standard deviation. If the expectation of the estimator is equal to the hypothetical true value of the parameter to be estimated, the estimator is called unbiased. Otherwise, it is biased. The deviation of the true value from the expectation is called bias.

Bias is the systematic error. It is, therefore, equivalent to the concept accuracy. There are two essentially different sources of bias. In the first place, the expectation of the observations may be different from the function model assumed. In the above mentioned sinusoidal example, a trend may be present in addition to the sinusoid, while the model assumed and fitted to the observations consists of the sinusoidal function only. This will, of course, always result in a systematic deviation of the estimated parameters, even in the hypothetical complete absence of nonsystematic errors. The remedy is to include the trend in the model fitted. The inclusion of the trend has the effect that two additional parameters, the slope and the intercept of the trend, have to be estimated. It will be discussed later that there are various reasons to keep in measurement the number of parameters to be estimated as small as possible. Therefore, classical measurement measures to avoid errors such as trends, day-and-night cycles, and background are always preferable to including these contributions in the model. The second source of bias is of a completely different nature. It is produced by and is characteristic of the estimator, itself. It may, therefore, also occur if the assumed model of the observations and that of the expectations are the same. Two estimators of the same parameters from the same observations may have different bias. If the bias vanishes as the number of observations increases, the estimator is called asymptotically unbiased. An effective method to remove bias of this kind is described by Stuart and Ord (4).

The standard deviation of an estimator represents precision. It is the spread of the measurement result if the experiment is repeated under the same conditions. In the example of the estimation of the parameters of the sinusoid, the amplitude, frequency, and phase have to be estimated simultaneously. This estimator is, therefore, vector valued. A vector estimator has a covariance matrix associated with it. The diagonal elements of this matrix are the variances of the estimators of each of the elements of the vector of parameters. The off-diagonal elements represent the covariances of the estimators of different elements. Bias and standard deviation are statistical key properties of estimators for practical measurement purposes. They demonstrate the practical feasibility, clarity, and generality of the model based statistical parameter estimation approach to the treatment of errors in observations. In addition to these desirable properties, model based parameter

estimation has a number of advantages which will now be discussed.

It has been mentioned earlier that for the measurement of the parameters of the same model from the same set of observations, use may be made of different estimators. These estimators will generally have different standard deviations, that is, have different precision. The question may, therefore, be asked which estimator is most precise? Or put somewhat differently, what precision is attainable if any estimator may be used? The answer to this question may be given using the concept of Fisher information. The Fisher information with respect to the model parameters is computed from the probability density function of the observations. If the model has one parameter, the quantity computed is called the Fisher information amount. If more than one parameter is measured, such as in the sinusoidal example, it assumes the form of a matrix, the Fisher information matrix. This is a symmetrical matrix of the same order as the number of parameters. The elements of the Fisher information matrix are dependent on the probability density of the observations and on the model of the expectations. They are independent of any method of estimation. With any set of observations, considered as stochastic variables, a Fisher information matrix with respect to the unknown parameters is associated. The inverse of the Fisher information matrix is called the Cramér Rao lower bound. It can be shown that the diagonal elements of the covariance matrix of any unbiased vector estimator, the variances of the elements, cannot be smaller than the corresponding diagonal elements of the Cramér Rao lower bound. Therefore, any unbiased estimator is at best as precise as a hypothetical estimator of the same parameters having the Cramér Rao lower bound as its covariance matrix. Thus, the Cramér Rao lower bound is a standard to which the precision of an unbiased estimator may be compared. This is the reason why the ratio of a diagonal element of the Cramér Rao lower bound to the variance of an estimator is called the efficiency of the estimator. Cramér Rao theory also extends to functions of the estimated parameters. For example, suppose that the height, width, and location of a Gaussian pulse or spectral peak are estimated, but that the quantities to be measured ultimately are the location and the area. Then, the Cramér Rao lower bound for the height, width, and location combined with the expression for the area in terms of the width and height may be used to compute the Cramér Rao lower bound for the location and area. The resulting expressions exactly describe the propagation of the Cramér Rao lower bound for the original parameters to that for location and area. Thus, they show exactly the sensitivity of the Cramér Rao lower bound for the area parameter to the various elements of the Cramér Rao lower bound for the original parameters. This means that an instrument has been found to compute the propagation of errors in the observations to errors in the parameters and, subsequently, to errors in functions of the parameters.

For the engineer and applied scientist, an important question is how to make the influence of errors in the observations upon the measurement result as small as possible. This is equivalent to the question of how to find the method that produces the most precise measurement result from the available observations. For calibration purposes, precision itself may be the ultimate purpose. In other applications, precision is often pursued not for its own sake but to make the conclusions drawn from the measurement result more reliable. The extent to which it is possible to find an efficient estimation method

depends on the available a priori knowledge of the probability density function of the observations. As has been discussed earlier, this probability density function is parametric in the hypothetical exact values of the unknown parameters. This dependence of the probability density function of the observations on the parameters may be used to derive a so-called maximum likelihood estimator of the parameters. First, the numerical values of the available observations are substituted for the corresponding independent variables of the probability density function. Next, the true values of the parameters are considered to be variables. The function thus obtained is called the likelihood function of the parameters. Finally, the likelihood function is maximized with respect to the parameters. The values of the parameters at the maximum are the maximum likelihood estimates of the parameters. This procedure shows the first advantage of the maximum likelihood estimator: it is easily found. In addition, the maximum likelihood estimator has a number of favorable statistical properties. The most important of these is that under general conditions, it attains asymptotically the Cramér Rao lower bound. This means that for a large number of observations, it is most precise. The elements of the Cramér Rao lower bound depend on experimental variables. For example, in the estimation of the parameters of the sinusoid and those of the Gaussian pulse, these are the number of observations and their location. If experimental variables such as these may to a certain extent be freely chosen, this freedom may be used to minimize the Cramér Rao lower bound and, thus, the asymptotic variance. This manipulation of the covariance matrix using experimental variables is called experimental design. From a practical point of view, experimental design may be very attractive since it may lead to a more precise measurement result with the same effort or even less. For practical measurement, the invariance property of maximum likelihood estimators is also important: functions of maximum likelihood estimators are maximum likelihood estimators themselves. Generally, maximizing the likelihood function with respect to the parameters is a nonlinear optimization problem which can only be solved using an iterative numerical optimization method. For a long time, this has been a serious impediment to the application of maximum likelihood, but today, excellent optimization methods and software are available which make the method accessible to any user.

If the observations are normally distributed, the maximum likelihood estimator of the parameters can be shown to be the weighted least squares estimator with the inverse of the covariance matrix of the observations as weighting matrix. If the observations are linear in all parameters to be estimated, the least squares estimator is a relatively simple closed form expression linear in the observations. In addition, if the observations are not normally distributed, the weighted least squares method with the inverse covariance matrix as weighting matrix still has the smallest variance among all estimators that are both linear in the observations and unbiased. If the observations are nonlinear in one or more of the parameters to be estimated, the least squares estimator is, as a rule, no longer a closed form and has to be evaluated using an iterative numerical method. However, effective, specialized, and reliable numerical methods and software are available that make the use of nonlinear least squares straightforwardly. As a result, least squares has become a major tool in the handling of observations subject to error in general and not only of normally distributed observations.

EXPECTATIONS OF OBSERVATIONS

The reduction or minimization of the effect of errors in the observations upon the measurement result requires a mathematical model of the observations. In this article, additive nonsystematic errors in the observations will be modeled as stochastic variables with an expectation equal to zero. This implies that the observations are also stochastic variables, and that the expectations of the observations are the hypothetical exact or errorless observations. Thus, these expectations constitute the model underlying the observations. It will be assumed throughout that this model is a parametric function, and that parameters of this model are the quantities to be measured, or that the quantity to be measured can be computed from these parameters.

Example 1. The multiexponential model. Multiexponential observations are observations with expectations

$$y_n(\gamma) = \alpha_1 \exp(-\beta_1 x_n) + \dots + \alpha_L \exp(-\beta_L x_n) \quad (1)$$

with $n = 1, \dots, N$ where N is the number of observations, and the $2L \times 1$ vector γ is defined as $(\alpha_1 \dots \alpha_L \beta_1 \dots \beta_L)^T$, where the amplitudes α_ℓ and the decay constants β_ℓ are the parameters to be measured, and the superscript T denotes transposition. The measurement points x_n , $n = 1, \dots, N$ are supposed known. If, different from Eq. (1), there is a linear trend in the observations, this deterministic contribution has to be included in the model of the expectations of the observations

$$y_n(\eta) = \alpha_1 \exp(-\beta_1 x_n) + \dots + \alpha_L \exp(-\beta_L x_n) + \lambda x_n + \mu$$

where $\eta = (\alpha_1 \dots \alpha_L \beta_1 \dots \beta_L \lambda \mu)^T$. In this expression, λ and μ are the slope and the intercept of the trend, respectively. These parameters have to be estimated along with the parameters α_ℓ and β_ℓ . This means that the number of parameters to be estimated has increased by two. It will be shown below that this is not only disadvantageous from a computational point of view, but it also unfavorably influences the precision with which the α_ℓ and β_ℓ can be measured. Therefore, it is worthwhile to keep the number of parameters as small as possible. As a consequence, changing the experimental conditions to remove the trend is always preferable to including it in the model of the expectations. On the other hand, if trends cannot be avoided, they have to be included since otherwise, the model of the expectations is wrong. Then, values for the amplitudes and decay constants are found systematically deviating from the α_ℓ and β_ℓ , even in the hypothetical case that nonsystematic errors in the observations are absent.

THE DISTRIBUTION OF THE OBSERVATIONS

The mathematical model of the observations is completed by a description of how the observations are distributed about their expectations. This is done in the form of the joint probability density function of the observations. If $w = (w_1 \dots w_N)^T$ is the vector of the N available observations, their probability density function may be described as $p(w)$. Then, the expectation

$E[w] = (E[w_1] \dots E[w_N])^T$ is defined as

$$E[w] = \int \dots \int wp(w)dw \quad (2)$$

where $dw = dw_1 dw_2 \dots dw_N$, and the integrations are carried out over all possible values of w . Then,

$$E[w] = y(\theta) \quad (3)$$

where $y(\theta) = [y_1(\theta) \dots y_N(\theta)]^T$, and $y_n(\theta)$ is the function parametric in the unknown parameters θ defining the errorless observations such as the exponential model described by Eq. (1). For engineering and applied science, two probability density functions are particularly important. These are the normal probability density function and the Poisson probability density function.

Example 2. The normal probability density function. The observations $w_1, \dots, w_N$ are said to be normally distributed if their probability density function is described by

$$p(w) = \frac{1}{(2\pi)^{N/2}(\det W)^{1/2}} \exp \left[-\frac{1}{2}(w - E[w])^T W^{-1}(w - E[w]) \right]$$

where the $N \times N$ matrix W is the covariance matrix of the observations defined by its (i, j) -the element $\text{cov}(w_i, w_j)$ and $\det W$ and W^{-1} are the determinant and the inverse of W , respectively. This probability density function and many others are discussed in Ref. 3. Equation (3) defines the functional dependence of the normal probability density function on the parameters of the function modelling the expectations. For what follows, the logarithm of $p(w)$ as a function of the parameters θ is needed. After substituting $y(\theta)$ for $E[w]$, it is described by

$$-\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(\det W) - \frac{1}{2}[w - y(\theta)]^T W^{-1}[w - y(\theta)] \quad (4)$$

Notice that both first terms of this expression are independent of the parameter vector θ , while the last term is a quadratic form in the elements $w_n - y_n(\theta)$ of $w - y(\theta)$. Observations in practice are often, but not always, normally distributed. One of the reasons is that if the nonsystematic errors are the sum of a number of nonsystematic errors from independent sources, their distribution tends to normal as described by the central limit theorem discussed in Ref. 2.

Example 3. The Poisson probability density function. This probability density function concerns counting statistics. It is described in Ref. 3. Examples of Poisson distributed stochastic variables are radioactive particle counts and pixel values in electron microscopes. The number of counts is Poisson distributed if the probability that it is equal to w_n is given by

$$p(w_n) = \exp(-\lambda_n) \frac{\lambda_n^{w_n}}{w_n!} \quad (5)$$

Simple calculations show that $E[w_n] = \lambda_n$, and that the standard deviation of w_n is equal to $\sqrt{\lambda_n}$. If $w_1, \dots, w_N$ are independent, as is often assumed in applications, their joint probability

density function is equal to the product the probabilities described by Eq. (5)

$$p(w) = \prod_n p(w_n) \quad (6)$$

with $n = 1, \dots, N$. Since $E[w] = \lambda$ with $\lambda = (\lambda_1 \dots \lambda_N)^T$ and $\lambda_n = y_n(\theta)$, the logarithm of the probability density function defined by Eq. (6) may be written

$$\sum_n -y_n(\theta) + w_n \ln[y_n(\theta)] - \ln(w_n!) \quad (7)$$

Notice that the last term in this expression is independent of the parameter vector θ .

From Example 2 and Example 3, the general approach to establishing the dependence of the probability density function of the observations on the parameters, that is, the quantities to be measured, is now clear. First, the expectation of the observations w_n is computed. Then, the result is substituted for the relevant quantities in the probability density function. The probability density function thus obtained is parametric in the parameters of the expectations, that is, of the hypothetical errorless observations. This is the form of the probability density function that will be used hereafter for two purposes. First, it will be used for the computation of the highest attainable precision with which the parameters can be measured from the available observations. It will also be used to find the most precise method to estimate the parameters from the observations.

ATTAINABLE MEASUREMENT PRECISION IN THE PRESENCE OF MEASUREMENT ERRORS

Suppose that a number of N observations $w_1, \dots, w_N$ is available and that the expectations of the observations are described by the multiexponential model defined by Eq. (1). If this model is fitted to the observations with respect to its parameters, the amplitudes, and the decay constants, one could choose the sum of the squares of the deviations of the model from the observations as a criterion of goodness of fit. Then, this criterion could be minimized with respect to the parameters, and the parameter values for which the criterion would be minimum would be the solution. This is the well-known ordinary least squares solution. Alternatively, one could have chosen the values of the parameters for which the sum of the absolute values of the deviations would be minimum. This is the least absolute values or least moduli solution. Then, if the experiment could be repeated sufficiently often, the experimenter could compare the results of both methods and could decide which of both would be most precise. Seeing that the one method is more precise than the other, the experimenter might wonder what the highest attainable precision from these observations is with any method. It has been found that under general conditions, this question may be answered using the concept Fisher information. For a discussion of Fisher information, see Ref. 4. For the computation of the Fisher information, the probability density function of the observations $p(w; \theta)$ is used. This is done as follows. First, the logarithm of $p(w; \theta)$ is taken. For the normal probability density function and for the Poisson probability density function, the result of this operation is described by Eqs. (4) and (7), respectively. Next, the gradient vector of $\ln p(w; \theta)$ with respect

to the elements of θ is calculated. It is defined as

$$\frac{\partial \ln p(w; \theta)}{\partial \theta}$$

If θ is a $K \times 1$ vector, so is the gradient vector. Its k -th element is $\partial \ln p(w; \theta) / \partial \theta_k$. Next, the $K \times K$ matrix

$$\frac{\partial \ln p}{\partial \theta} \frac{\partial \ln p}{\partial \theta^T} \quad (8)$$

is computed where, for simplicity, the arguments of $p(w; \theta)$ have been left out, and $\partial \ln p / \partial \theta^T$ is the transpose of $\partial \ln p / \partial \theta$. The $K \times K$ Fisher information matrix is defined as the expectation of Eq. (8)

$$M = E \left[\frac{\partial \ln p}{\partial \theta} \frac{\partial \ln p}{\partial \theta^T} \right] = \int \dots \int \frac{\partial \ln p}{\partial \theta} \frac{\partial \ln p}{\partial \theta^T} p dw$$

It is not difficult to show that M may alternatively be written

$$M = -E \left[\frac{\partial^2 \ln p}{\partial \theta \partial \theta^T} \right] \quad (9)$$

In this expression, $\partial^2 \ln p / \partial \theta \partial \theta^T$ is the Hessian matrix of $\ln p$ defined by its (q, r) -th element $\partial^2 \ln p / \partial \theta_q \partial \theta_r$.

Example 4. The Fisher information matrix for normally distributed observations. If the observations are normally distributed, the logarithm of the probability density function as a function of the parameters is described by Eq. (4). Then, elementary computations making use of the fact that $E[w_n - y_n(\theta)] = 0$ yield

$$M = \frac{\partial y^T}{\partial \theta} W^{-1} \frac{\partial y}{\partial \theta^T} \quad (10)$$

In this expression, the $N \times K$ matrix $\partial y / \partial \theta^T$ is the Jacobian matrix of $y(\theta)$ with respect to θ . Its (n, k) -th element is equal to $\partial y_n(\theta) / \partial \theta_k$. Therefore, for the multiexponential model, the elements of the Jacobian matrix are of the form $\exp(-\beta_\ell x_n)$ or $-\alpha_\ell x_n \exp(-\beta_\ell x_n)$ with $\ell = 1, \dots, L$.

Example 5. The Fisher information matrix for independent Poisson distributed observations. If the observations are independent and Poisson distributed, the logarithm of the probability density function of the observations as a function of the parameters is described by Eq. (7). This expression may be used to show that here, the information matrix is also described by Eq. (10), but with $W = \text{diag}(y_1 \dots y_N)$, where $y_n = y_n(\theta)$.

The importance of the Fisher information matrix is that from it the Cramér Rao lower bound may be computed. This is a lower bound on the variance of all unbiased estimators of parameters or of functions of parameters. An estimator t is said to be unbiased for the parameter θ if its bias, defined as

$$E[t] - \theta$$

is equal to the null vector. Otherwise, it is biased. In measurement terminology: if the model of the expectations of the observations is correctly specified and the estimator used is unbiased

for the parameters, the measurement result has no systematic error.

Next, suppose that $t(w)$ is any unbiased estimator of the vector of parameters θ from the observations w . Then the Cramér Rao inequality states that

$$\text{cov}[t(w), t(w)] \geq M^{-1} \quad (11)$$

In this expression, $\text{cov}[t(w), t(w)]$ is the covariance matrix of the estimator $t(w)$. That is, the (p, q) -th element of this matrix is defined as the covariance of the p -th element $t_p(w)$ and the q -th element $t_q(w)$. Therefore, the diagonal elements are the variances of $t_1(w), \dots, t_K(w)$, respectively. In Eq. (11) expresses that the difference of the matrix $\text{cov}[t(w), t(w)]$ and the matrix M^{-1} is positive semidefinite. A property of positive semidefinite matrices is that their diagonal elements cannot be negative. Therefore, the diagonal elements of $\text{cov}(t(w), t(w))$, that is, the variances of the elements of the estimator $t(w)$, must be larger than or be equal to the corresponding diagonal elements of M^{-1} . Consequently, the latter diagonal elements are a lower bound on the variances of the elements of the estimator $t(w)$. The matrix M^{-1} is called the Cramér Rao lower bound. For normally distributed observations and for Poisson distributed observations, the Cramér Rao lower bound may be computed by inverting the Fisher information matrix defined by Eq. (10) with appropriate matrix W , respectively. Notice that the main ingredients are simply the derivatives of the model $y_n(\theta)$ with respect to the parameters in each measurement point. These are quantities that are usually easy to compute.

The Cramér Rao lower bound would be of theoretical value only if there would not exist estimators attaining it. Later in this article, estimators will be introduced that do so, at least asymptotically. Therefore, the Cramér Rao lower bound may be used as a standard to which the precision of any estimator may be compared. Notice that the Cramér Rao lower bound is not related to a particular estimation method. It depends on the statistical properties of the observations, the measurement points, and in most cases, the hypothetical true values of the parameters. This dependence on the true values looks, at first sight, as a serious impediment to the practical use of the bound. However, the expressions for the bound provide the means to compute numerical values for it using nominal values of the parameters. This provides the experimenter with quantitative insight in what precision may be achieved from the available observations, an insight that without the bound would be absent. Thus, using the bound, the experimenter gets a detailed insight in the sensitivity of the precision to the values of the parameters. The experimenter also gets impression if the experimental design, that is, the values and the number of the measurement points x_n , is adequate for the purposes concerned. This means an impression if the precision is sufficient to make conclusions possible. If not, there is no other choice than to change the experimental design. If this is not possible, it is to be concluded that the observations are not suitable for the purposes of the measurement procedure.

In many applications, some of the quantities to be measured are functions of the parameters and not the individual parameters. A simple example is the following.

Example 6. Measurement of peak area and location. Suppose that a number of error corrupted observations $w_1, \dots, w_N$

has been made on a spectral peak described by

$$\alpha \exp \left[-\frac{1}{2} \left(\frac{x - \beta}{\gamma} \right)^2 \right] \quad (12)$$

where the parameters α , β , and γ are the peak height, location, and half-width, respectively. Suppose that only the peak location and the peak area are of interest. Then these are described by β and $(2\pi)^{1/2}\alpha\gamma$, respectively.

Fortunately, the Cramér Rao lower bound of functions of the parameters follows relatively easily from the Cramér Rao lower bound for the parameters. Let $r = [r_1(w) \dots r_L(w)]^T$ be an unbiased estimator of the vector function $\rho(\theta) = [\rho_1(\theta) \dots \rho_L(\theta)]^T$, that is, $E[r] = \rho(\theta)$. Furthermore, let M be the information matrix for θ . Then it can be shown that

$$\text{cov}[r(w), r(w)] \geq \frac{\partial \rho}{\partial \theta^T} M^{-1} \frac{\partial \rho}{\partial \theta}$$

where $\partial \rho / \partial \theta^T$ is the $L \times K$ Jacobian matrix with (p, q) -th element $\partial \rho_p / \partial \theta_q$. Therefore, the Cramér Rao lower bound for unbiased estimation of ρ is described by

$$\frac{\partial \rho}{\partial \theta^T} M^{-1} \frac{\partial \rho}{\partial \theta} \quad (13)$$

with M^{-1} the Cramér Rao lower bound for α , β , and γ .

Example 7. The Cramér Rao lower bound for peak area and location. The vector $\rho(\theta)$ for Example 6 is described by $\rho(\theta) = [\beta(2\pi)^{1/2}\alpha\gamma]^T$. Then, the Jacobian matrix of $\rho(\theta)$ with respect to $(\alpha \ \beta \ \gamma)^T$ is defined as

$$\frac{\partial \rho}{\partial \theta^T} = \begin{pmatrix} 0 & 1 & 0 \\ (2\pi)^{1/2}\gamma & 0 & (2\pi)^{1/2}\alpha \end{pmatrix}$$

where $\rho = \rho(\theta)$. The Cramér Rao lower bound for unbiased estimation of ρ is then computed from Eq. (13).

The premultiplication and postmultiplication of M^{-1} in Eq. (13) describe what is conventionally called error propagation. To see how this works, suppose that $\rho = [\rho_1(\theta) \ \rho_2(\theta)]^T$, $\theta = (\theta_1 \ \theta_2)^T$, and let the Cramér Rao lower bound for θ be

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$$

Then, the diagonal elements of the Cramér Rao lower bound for ρ_1 and ρ_2 are equal to

$$\left(\frac{\partial \rho_i}{\partial \theta_1} \right)^2 c_{11} + 2 \left(\frac{\partial \rho_i}{\partial \theta_1} \right) \left(\frac{\partial \rho_i}{\partial \theta_2} \right) c_{12} + \left(\frac{\partial \rho_i}{\partial \theta_2} \right)^2 c_{22}$$

with $i = 1, 2$, respectively. This expression shows how the variances c_{11} and c_{22} and the covariance c_{12} of a hypothetical estimator that attains the Cramér Rao lower bound for θ propagate to the variances of a hypothetical estimator of ρ_1 and ρ_2 that also attains the Cramér Rao lower bound. Similar error propagation schemes are proposed in the literature for covariance matrices

of functions of estimators in general, for example in reference 5. These schemes are approximations using the linear Taylor polynomial instead of the nonlinear functions. Equation (13), on the other hand, is exact.

Next, suppose that M is the information matrix for the estimation of $\theta = (\theta_1 \dots \theta_K)^T$, and that an additional parameter θ_{K+1} is to be estimated. For example, θ_{K+1} may be a constant term added to the spectroscopic line model described by Eq. (12) to model a constant background contribution. Then, M has to be augmented with one row and one column corresponding to θ_{K+1} . If the augmented information matrix is inverted, all first K diagonal elements can be shown to be larger than or equal to the corresponding diagonal elements of M^{-1} . Equality occurs only if the nondiagonal elements of the $(K+1)$ -th row and $(K+1)$ -th column of the augmented information matrix happen to be equal to zero. Generally, it is not difficult to show that, typically, the first K diagonal elements of M^{-1} are monotonously increasing with the number of parameters in excess of K .

PRECISELY MEASURING FROM ERROR CORRUPTED OBSERVATIONS

The a priori knowledge of the experimenter about the observations and the extent to which this a priori knowledge is used may considerably influence the precision and accuracy of the measurement result. This concerns both systematic and non-systematic errors in the observations. Systematic errors in the observations are deviations of the assumed parametric model of the expectations from the true model of the expectations. Even in the absence of nonsystematic errors, discrepancy between both models produces systematic errors, that is, inaccuracy in the measurement result. Since no model fitted will be perfect, there will always be a certain amount of systematic error. Non-systematic errors are described by their distribution about the expectations of the observations. This distribution is not always known, but if it is, this knowledge may contribute substantially to the reducing of the nonsystematic error in the measurement result, that is, in the parameters estimates.

Suppose that observations $w_1, \dots, w_N$, are available and that their probability density function $f(w_1, \dots, w_N; \theta)$ is known where θ is the vector of unknown parameters and $w_1, \dots, w_N$ are the independent variables corresponding to the observations $w_1, \dots, w_N$, respectively. Assume that $w_1, \dots, w_N$ are substituted for $\omega_1, \dots, \omega_N$ in $f(\omega_1, \dots, \omega_N; \theta)$, respectively, and that the fixed true parameters θ are replaced by the vector of corresponding variables t . Then, the resulting function $f(w_1, \dots, w_N; t)$ of t is called the likelihood function of the parameters t , given the observations $w_1, \dots, w_N$. The maximum likelihood estimate of the parameters θ is defined as the value $\tilde{t}$ of t that maximizes the likelihood function.

The maximum likelihood estimator has a number of very favorable properties. In the first place, its definition shows that it is simple to find from the known probability density function of the observations. Furthermore, it can be shown to converge under general conditions in a statistically well-defined way to the true values of the parameters as the number of observations increases. Moreover, under general conditions, the covariance matrix of the maximum likelihood estimator approaches asymptotically the Cramér Rao lower bound. Then, the maximum likelihood estimator is asymptotically most precise. Also, a function of a maximum likelihood estimator is the maximum

likelihood estimator of the function. This is called the invariance property of maximum likelihood.

Two of these properties are asymptotic; they apply to an infinite number of observations. If they also apply to a finite or even small number of observations can often only be assessed by estimating from artificial, simulated observations. These simulations may reveal that maximum likelihood estimation applied to small numbers of observations may lead to bias, that is, systematic error in the measurement result. This kind of bias, or the major part of it, is usually inversely proportional to the number of observations and may be removed as follows. Let $\tilde{t}_N$ be the biased maximum likelihood estimate of θ obtained from $w_1, \dots, w_N$, and let $\bar{t}_{N-1}$ be the average of the N different maximum likelihood estimates computed from the N different sets of $N - 1$ observations obtained by omitting one observation from the set $w_1, \dots, w_N$. Then, it may be shown that

$$N\tilde{t}_N - (N-1)\bar{t}_{N-1}$$

is an estimator of θ which may only be biased to order $1/N^2$. This is the so-called Quenouille correction, today called jackknife. A favorable property of this correction is that it hardly affects the variance of the estimator.

Example 8. Maximum likelihood estimation of peak height, width and location from Poisson distributed observations. Suppose that observations $w_1, \dots, w_N$ are available made on the spectral peak of Example 6, and that these observations are independent and Poisson distributed. Then it follows from Example 3 that the likelihood function of the parameters is described by

$$\sum_n -y_n(t) + w_n \ln[y_n(t)] - \ln(w_n!) \quad (14)$$

with

$$y_n(t) = a \exp \left[-\frac{1}{2} \left(\frac{x_n - b}{c} \right)^2 \right]$$

where $t = (abc)^T$. To obtain the maximum likelihood estimate of α , β , and γ , Eq. (14) must be maximized with respect to t . This is a nonlinear optimization problem which has to be solved numerically. If the peak area is computed from the maximum likelihood estimates $\tilde{a}$ and $\tilde{c}$ as $(2\pi)^{1/2}\tilde{a}\tilde{c}$ this is, by the invariance property, a maximum likelihood estimate as well.

Example 9. Maximum likelihood estimation from observations disturbed by normally distributed errors. If the errors and, therefore, the observations are normally distributed, Eq. (4) shows that the likelihood function of the parameters is described by

$$-\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(\det W) - \frac{1}{2} [w - y(t)]^T W^{-1} [w - y(t)] \quad (15)$$

Since both first terms of this expression do not depend on the vector of parameters t , maximizing Eq. (15) is equivalent to minimizing

$$[w - y(t)]^T W^{-1} [w - y(t)]$$

with respect to t . This shows that with normally distributed observations, maximum likelihood estimation is equivalent to a weighted least squares measurement with W^{-1} as weighting matrix.

Least squares estimation is also often used if the distribution of the observations is not known or is known to be not normal. Then, the general expression for the least squares criterion is

$$[w - y(t)]^T \Omega [w - y(t)] \quad (16)$$

where Ω is a positive definite weighting matrix to be chosen by the experimenter.

Linear Least Squares

First, as an important special case, models linear in the unknown parameters θ are considered. Then

$$E[w] = y(\theta) = X\theta$$

with X a known $N \times K$ matrix, that is,

$$y_n(\theta) = x_{n1}\theta_1 + \dots + x_{nK}\theta_K$$

Notice that

$$x_n = (x_{n1} \dots x_{nK})^T$$

is the vector independent variable corresponding to the n -th observation w_n .

Example 10. Straight line fitting. If the observations w_n are made on a straight line $y = \alpha x + \beta$ at the points $x_{11}, \dots, x_{N1}$, then X and θ are described by

$$\begin{bmatrix} x_{11} & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ x_{N1} & 1 \end{bmatrix} \quad \text{and} \quad \theta = (\alpha \ \beta)^T$$

respectively.

The least squares solution $\hat{t}_\Omega$ for θ is

$$\hat{t}_\Omega = (X^T \Omega X)^{-1} X^T \Omega w \quad (17)$$

It is observed that this solution is a linear combination of the observations. As a result, the propagation of the errors in the observations to the measurement result is perfectly clear. Furthermore, since $E[w] = X\theta$, $E[\hat{t}_\Omega] = \theta$ and, hence, $\hat{t}_\Omega$ is an unbiased estimator of θ . Notice that $\hat{t}_\Omega$ has these properties for any distribution of the observations w . It is easily shown that the covariance matrix $\text{cov}(\hat{t}_\Omega, \hat{t}_\Omega)$ is equal to

$$(X^T \Omega X)^{-1} X^T \Omega W \Omega X (X^T \Omega X)^{-1} \quad (18)$$

The conclusion from this expression is that this covariance matrix and, therefore, the variances of $\hat{t}_\Omega$ depend on the choice of Ω . The question is then which Ω minimizes the covariance

described by Eq. (18). The answer has been found to be $\Omega = W^{-1}$. For this choice,

$$\text{cov}(\hat{t}_\Omega, \hat{t}_\Omega) \geq \text{cov}(\hat{t}_{W^{-1}}, \hat{t}_{W^{-1}})$$

As a consequence, the variances of the elements of $\hat{t}_\Omega$ for any choice of Ω are never smaller than those of the corresponding elements of $\hat{t}_{W^{-1}}$. The estimator $\hat{t}_{W^{-1}}$ is called the best linear unbiased estimator. Equation (17) shows that it is described by

$$\hat{t}_{W^{-1}} = (X^T W^{-1} X)^{-1} X^T W^{-1} w$$

Among all estimators that are both linear in the observations and unbiased, it is called best since it has smallest variance. Notice that only the expectation and the covariance matrix of the observations are specified, not their probability density function. Also notice that $\hat{t}_{W^{-1}}$ is optimal within the class of estimators that are both linear in the observations and unbiased. Therefore, there may be better, that is, more precise estimators among those that are not linear in the observations or are biased.

The covariance matrix of $\hat{t}_{W^{-1}}$ is equal to

$$(X^T W^{-1} X)^{-1} \quad (19)$$

If for normally distributed observations, the maximum likelihood estimator is computed, it is found to be identical to the best linear unbiased estimator $\hat{t}_{W^{-1}}$ and, consequently, to have a covariance matrix equal to the one given by Eq. (19). If next, the Cramér Rao lower bound is computed for the same observations, it is found to coincide with Eq. (19). The conclusion is that for normally distributed observations, the best linear unbiased estimator is identical with the maximum likelihood estimator and attains the Cramér Rao lower bound for any number of observations.

In measurement practice, the weighting matrix Ω of $\hat{t}_\Omega$ is often taken as the identity matrix. The reason may be that the covariance matrix W is unknown. Another reason may be the amount and the complexity of numerical computation involved since with $\Omega = I$, the estimator simplifies to the ordinary least squares estimator

$$\hat{t}_I = (X^T X)^{-1} X^T w \quad (20)$$

which is clearly easier to compute than $\hat{t}_\Omega$. The corresponding ordinary least squares criterion is described by

$$(w - Xt)^T(w - Xt)$$

which is simply the sum of the squares of the deviations

$$w_n - x_{n1}t_1 - x_{n2}t_2 + \dots - x_{nK}t_K$$

Notice that $\hat{t}_I$ is only the best linear unbiased estimator if the covariance matrix W is equal to $\sigma^2 I$, that is, if the observations are uncorrelated and have equal variance σ^2 . The estimator $\hat{t}_I$ is the maximum likelihood estimator and achieves the Cramér Rao lower bound if, in addition, the observations are normally distributed. Therefore, if these conditions are not met, the use of $\hat{t}_I$ may mean an exchange of precision for simplicity.

Finally, it is emphasized that Eq. (20) is a formal description of the ordinary linear least squares estimator. It is not a recipe for its numerical evaluation. Special numerical methods have been designed taking care of the fact that the set of linear equations described by Eq. (20) may be ill-conditioned. References 6 and 7 provide the details.

Nonlinear Least Squares

Nonlinear least squares is the most frequently used method for estimation of the parameters of nonlinear models. The criterion used is described by

$$[w - y(t)]^T[w - y(t)] = \sum_n [w_n - y(x_n; t)]^2 \quad (21)$$

which is Eq. (16) with weighting matrix $\Omega = I$. Notice that generally, the solution $\hat{t}_I$ for t minimizing the least squares criterion defined by Eq. (21) is only the maximum likelihood estimator if the observations are independent and identically normally distributed. This means normally distributed with covariance matrix $\sigma^2 I$. For other distributions, $\hat{t}_I$ is generally not the maximum likelihood estimate since it does not maximize the pertinent likelihood function. As compared with linear least squares, the amount of theory concerning nonlinear least squares is limited. However, if the observations are independent and identically distributed, then under general conditions, the least squares estimator $\hat{t}_I$ is asymptotically normally distributed with covariance matrix

$$\sigma^2 \left(\frac{\partial y^T}{\partial \theta} \frac{\partial y}{\partial \theta^T} \right)^{-1} \quad (22)$$

where $y = y(\theta)$. This result is due to Jennrich (8). Notice that the computation of this covariance matrix requires the parameters to be known. In practice, this is not the case, and nominal or estimated values are substituted for the exact ones. Also notice that for independent and identically normally distributed observations, Eq. (22) is equal to the Cramér Rao lower bound. The general form of the elements of the matrix $(\partial y^T / \partial \theta)(\partial y / \partial \theta^T)$ is

$$\sum_n \frac{\partial y(x_n; \theta)}{\partial \theta_p} \frac{\partial y(x_n; \theta)}{\partial \theta_q}$$

This expression shows the dependence of the elements of this matrix upon the values of the independent variable x . Therefore, if the experimenter has some freedom in the choice of the measurement points, it may be used to manipulate the covariance matrix described by Eq. (22) in a desired way. This usually concerns the diagonal elements, that is, the variances and is an example of experimental design: the manipulation of the variances by selecting free experimental variables.

The gradient of the nonlinear least squares criterion with respect to the parameter vector t is equal to

$$-2 \sum_n [w_n - y_n(t)] \frac{\partial y_n(t)}{\partial t} \quad (23)$$

A necessary condition for a point to be a minimum is that the gradient is equal to the null vector. If Eq. (23) is equated to the null vector, this produces a set of K nonlinear equations in

K variables. This set must be solved by an iterative numerical method since, typically, it cannot be solved in closed form. For this problem, specialized numerical methods have been developed. Most frequently used are the Gauss–Newton method and the Levenberg–Marquardt method. These are described in references 8 and 6, respectively. Software for their practical implementation is found in references 6 and 7.

Many nonlinear models in engineering practice are linear in some of their parameters. How this special property may be exploited in nonlinear least squares estimation is illustrated in the following example.

Example 11. Least squares estimation of the parameters of a multiexponential model. Suppose that in a least squares estimation problem, the model fitted is described by

$$y_n(t) = a_1 \exp(-b_1 x_n) + \dots + a_L \exp(-b_L x_n)$$

where $t = (a^T b^T)^T$ with linear parameters $a = (a_1 \dots a_L)^T$ and nonlinear parameters $b = (b_1 \dots b_L)^T$. Then Eq. (23) shows that at the minimum of the least squares criterion, the derivatives with respect to the linear parameters a must satisfy

$$\sum_n [w_n - a_1 \exp(-b_1 x_n) + \dots + a_L \exp(-b_L x_n)] \exp(-b_\ell x_n) = 0$$

with $\ell = 1, \dots, L$. This may be considered a set of L linear equations in L unknowns a_ℓ . The solution for these unknowns is a function of the unknown nonlinear parameters b_ℓ and is denoted as $a_\ell(b)$. Substitution of the $a_\ell(b)$ for the a_ℓ in the least squares criterion yields

$$\sum_n [w_n - a_1(b) \exp(-b_1 x_n) + \dots + a_L(b) \exp(-b_L x_n)]^2$$

Thus, the least squares criterion has become a function of the nonlinear parameters b only. Minimization of it with respect to b yields the solution $\hat{b}$ for β and the solution $\hat{a} = a(\hat{b})$ for α .

Nonlinear least squares problems of the kind described in Example 11 are called separable nonlinear least squares problems since the linear and the nonlinear parameters are estimated separately. Notice that in Example 11, the number of parameters involved in the iterative numerical minimization is reduced by a factor of two. This also means that the number of initial values for the procedure is reduced correspondingly.

HANDLING MEASUREMENT ERRORS IN NONSTANDARD PROBLEMS

Complex Parameter Estimation

Many practical measurement problems concern complex valued parameters or mixtures of real and complex valued parameters. In particular, these problems are found in measurement in the frequency domain. Such complex parameter estimation problems can always be transformed into real parameter estimation problems by splitting a complex parameter into its real and imaginary part and estimating these real quantities separately. This, however, leads to unnecessarily complicated expressions for the estimator and, as a result, to complicated

numerical procedures. This is avoided by leaving quantities complex if they are complex by nature.

The most important tool in the formulation of complex parameter measurement from error corrupted observations is the following. Suppose that in a measurement problem there are $K + 2L$ parameters

$$\theta = (\eta_1 \dots \eta_K \alpha_1 \beta_1 \dots \alpha_L \beta_L)^T$$

of which the η_k are intrinsically real, and the α_ℓ and β_ℓ are the real and imaginary parts of the complex parameters $\gamma_\ell = \alpha_\ell + j\beta_\ell$ with $j = \sqrt{-1}$. Then γ_ℓ and its complex conjugate γ_ℓ^* on the one hand and α_ℓ and β_ℓ on the other are connected by the linear transformation

$$\begin{bmatrix} \gamma_\ell \\ \gamma_\ell^* \end{bmatrix} = J \begin{bmatrix} \alpha_\ell \\ \beta_\ell \end{bmatrix}$$

where

$$J = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix}$$

Therefore, the mixed real complex parameter vector

$$\zeta = (\eta_1 \dots \eta_K \gamma_1 \gamma_1^* \gamma_L \gamma_L^*)^T$$

and θ are connected by

$$\zeta = B_{K+2L} \theta \quad (24)$$

where B_{K+2L} is the $(K + 2L) \times (K + 2L)$ block diagonal matrix

$$B_{K+2L} = \text{diag}(I_K \quad A_{2L})$$

with I_K the identity matrix of order K and A_{2L} the $2L \times 2L$ block diagonal matrix

$$A_{2L} = \text{diag}(J \dots J)$$

The theory, methods, and techniques presented up to now concerned the estimation of real parameters from error corrupted observations. Using the linear transformation described by Eq. (24), transformation of the pertinent expressions into those for estimating a mixed real complex parameter vector is relatively easy. All that is required is observing the mathematical rules governing linear transformation of coordinates in general. For what follows, it is important to notice that Eq. (24) implies that α_ℓ and β_ℓ are transformed into both γ_ℓ and γ_ℓ^* . Also, the definition of the covariance matrix of a vector of complex stochastic variables is needed. Let z be a vector of stochastic variables. Then, the covariance matrix of z is defined as

$$E[(z - E[z])(z - E[z])^H]$$

where the superscript H denotes complex conjugate transposition. The Fisher information matrix defined by Eq. (9) after the transformation of parameters described by Eq. (24) is given by

$$M = -E \left[\frac{\partial^2 \ln f}{\partial \zeta^* \partial \zeta^T} \right]$$

and the corresponding Cramér Rao lower bound on the variance of unbiased estimators of ζ is equal to M^{-1} . Again using Eq. (24), the Cramér Rao lower bound for a vector of real and complex functions $\phi(\zeta)$ of the mixed real complex parameter vector is found to be

$$\frac{\partial \phi}{\partial \zeta^T} M^{-1} \frac{\partial \phi^H}{\partial \zeta^*}$$

where for brevity, the argument of $\phi(\zeta)$ has been omitted. A further example is the weighted least squares estimator defined by Eq. (17). After transformation of θ into ζ and ℓ_Ω into $\hat{\zeta}_\Omega$, respectively, it becomes

$$\hat{\zeta}_\Omega = (R^H \Omega R)^{-1} R^H \Omega w$$

where the complex $N \times (K + 2L)$ matrix R is equal to XB^{-1}_{K+2L} . If Ω is equal to W^{-1} , this is the best linear unbiased estimator. Finally, suppose that the real complex $(P + 2Q) \times 1$ vector of observations u is composed of the elements of the real $(P + 2Q) \times 1$ vector of observations w as follows

$$u = B_{P+2Q} w$$

with $P + 2Q = N$. Then, u is a vector of real and complex observations described by

$$(w_1 \dots w_P w_{P+1} + jw_{P+2} w_{P+1} - jw_{P+2} \dots w_{P+2Q-1} + jw_{P+2Q} w_{P+2Q-1} - jw_{P+2Q})^T$$

and

$$\hat{\zeta}_\Psi = (S^H \Psi S)^{-1} S^H \Psi u$$

where S and Ψ are equal to $B_{P+2Q} R$ and $B^{-H}_{P+2Q} \Omega B^{-1}_{P+2Q}$, respectively. The covariance matrix of the mixed real complex observations is defined as $E[(u - E[u])(u - E[u])^H]$ and is, therefore, equal to $B_{P+2Q} W B^H_{P+2Q}$. Hence, the estimator $\hat{\zeta}_\Psi$ with $\Psi = (B_{P+2Q} W B^H_{P+2Q})^{-1}$ is the best linear unbiased estimator.

The iterative numerical optimization of likelihood functions and nonlinear least squares criteria of mixed real complex parameters may be carried out directly with respect to the vector of mixed real complex parameters. This is discussed in reference 9. In particular, use may be made of the complex gradient. Specifically, the complex gradient of the logarithm of the likelihood function $\ln f$ with respect to the complex parameter vector z is defined as $\partial \ln f / \partial z$. An important property of this complex gradient is that the real gradient $\partial \ln f / \partial t$ is equal to the null vector if and only if the complex gradient is equal to the null vector. Therefore, the complex gradient may be used to find maxima of the likelihood function and minima of the nonlinear least squares criterion in the same way as the real gradient.

Nonstandard Fourier Analysis

Estimation of Fourier coefficients from error disturbed observations made on periodic functions is an important problem in dynamic system identification in general and in specialized applications as crystal structure reconstruction. Suppose that the

problem is to estimate the Fourier coefficients γ_k , $k = 0, \pm 1, \dots, \pm K$ and, possibly, the period δ of the real periodic function

$$y_n(\zeta) = \sum_k \gamma_k \exp(-j2\pi k x_n / \delta) \quad (25)$$

from error corrupted observations $w = (w_1 \dots w_N)^T$ where the measurement points x_n are known, and the vector of unknown parameters ζ is either equal to $\gamma = (\gamma_0 \gamma_1 \gamma_{-1} \dots \gamma_K \gamma_{-K})^T$ or to $(\gamma^T \delta)^T$, where γ_0 and δ are real, while the remaining γ_k are complex and satisfy $\gamma_{-k} = \gamma_k^*$ since the $y_n(\zeta)$ are real. It will not be supposed that the measurement points are equidistant, nor if they are, that the period is a known integer multiple of the sampling interval, and an integer number of periods is observed. The purpose of this section is to formulate the estimation of the parameters ζ as a complex statistical parameter estimation problem and to describe the special conditions under which this problem simplifies to the standard Discrete Fourier Transform, the DFT.

Example 12. Estimation of Fourier coefficients from Poisson distributed observations. Suppose that observations $w_n \geq 0$, $n = 1, \dots, N$ are available with expectations described by Eq. (25) and that these observations have a Poisson distribution. Then, by Eq. (14), the likelihood function of the parameters is

$$\sum_n -y_n(z) + w_n \ln[y_n(z)] - \ln(w_n!) \quad (26)$$

with

$$y_n(z) = \sum_k c_k \exp(j2\pi k x_n / d)$$

where the elements of $z = (c^T d)^T$ correspond to those of ζ , and those of $c = (c_0 c_1 c_{-1} \dots c_K c_{-K})^T$ correspond to those of γ . Then, the complex gradient of Eq. (26) with respect to z is

$$-\sum_n \left(1 - \frac{w_n}{y_n(z)} \right) \frac{\partial y_n(z)}{\partial z} \quad (27)$$

Since the maximum of the likelihood function is, by definition, a stationary point, the $2K + 2$ elements of the maximum likelihood estimate $\hat{z}$ of ζ must satisfy the $2K + 2$ nonlinear equations in z obtained by equating Eq. (27) to the null vector. The numerical solution for z representing the absolute maximum of the likelihood function described by Eq. (26) is the maximum likelihood estimate of the Fourier coefficients.

It follows from Eq. (25) that the expectations $y_n(\zeta)$ of the observations w_n are described by

$$y_n(\zeta) = X\gamma$$

where the n -th row of X is defined as

$$(1 \exp(j2\pi x_n / \delta) \exp(-j2\pi x_n / \delta) \dots \exp(j2\pi K x_n / \delta) \exp(-j2\pi K x_n / \delta))$$

If the observations are normally distributed with covariance matrix W , then the maximum likelihood estimator of ζ has to minimize

$$[w - y(z)]^T W^{-1} [w - y(z)] \quad (28)$$

Hence, if only the Fourier coefficients are unknown, their maximum likelihood estimator is

$$\hat{c}_{W^{-1}} = (X^H W^{-1} X)^{-1} X^H W^{-1} w \quad (29)$$

For observations with a distribution different from normal, this estimator is no longer maximum likelihood but is still best linear unbiased. If, in addition, the period is unknown, the estimation problem is recognized as a separable nonlinear least squares problem. The model is linear in the $2K + 1$ Fourier coefficients and nonlinear in the period d . This means that, in addition to Eq. (29), one further equation must be satisfied. This is the equation resulting from equating the derivative of the least squares criterion with respect to d to zero. If, in this equation, Eq. (29) is substituted for the Fourier coefficients c , a scalar nonlinear equation is obtained in the period d only. Hence, all that needs to be done is real, scalar root finding to estimate the period d and substitute the estimate in the closed form of Eq. (29) for the Fourier coefficients. The estimates thus obtained are maximum likelihood if the observations are normal with covariance matrix W . They are weighted complex nonlinear least squares estimates with other error distributions.

If the covariance matrix W is unknown, the ordinary least squares estimator

$$\hat{c}_I = (X^H X)^{-1} X^H w \quad (30)$$

may be chosen, possibly combined with root finding for the period. This is a maximum likelihood estimator only if the w_n are independent and identically normally distributed about their expectations. For other distributions, it is a best linear unbiased estimator if only the Fourier coefficients are to be estimated, and the observations are uncorrelated with equal variance. In other cases, it is simply the ordinary least squares estimator.

A special case occurs if only the Fourier coefficients are to be estimated, the measurement points x_n are equidistant with interval Δ , the period is a known integer multiple of Δ , and an integer number of periods is observed. Under these conditions, the elements of $\hat{c}_I$ described by Eq. (30) may be shown to be equal to the DFT

$$\frac{1}{N} \sum_{n=1}^N w_n \exp[-j2\pi k(n-1)/M]$$

where $M\Delta$ is the period. Under the restrictive conditions mentioned, the DFT is, therefore, the maximum likelihood estimator if the observations are independent and identically normally distributed about their expectations. For other distributions, it is best linear unbiased if the observations are uncorrelated and have equal variance.

Measurement Errors and Resolution

Like precision and accuracy, resolution is a key notion in applied science and engineering. It is used in fields as diverse as radar, sonar, optics, electron optics, seismology and various

forms of spectroscopy. An extensive review of resolution is presented in reference 10.

The most important form of resolution is two-component resolution.

Example 13. Rayleigh two-component resolution. As discussed in reference 10, Rayleigh considers observations described by

$$\alpha \{\text{sinc}^2[2\pi(x - \beta_1)] + \text{sinc}^2[2\pi(x - \beta_2)]\}$$

with $\text{sinc}(x) = \sin(x)/x$. This is a pair of sinc-square components of equal height and located at β_1 and β_2 , respectively. As the difference in location decreases, the components increasingly overlap and become increasingly difficult to distinguish visually. According to Rayleigh, the components are resolvable if the absolute difference of β_1 and β_2 exceeds 0.5. At this distance, the maximum of the one component coincides with the first zero of the other, and the component sum has two maxima and a relative minimum in between. Then, the ratio of the value at the relative minimum to that at the maxima may be shown to be 0.81. Later, this ratio has been generalized to other component functions, and the distance corresponding to this ratio has been called generalized Rayleigh resolution limit.

From this example, it is clear that this classical resolution limit and comparable ones proposed later are, in fact, measures of component width. Since in definitions such as Rayleigh's, the component functions are known, and the observations are exact, today, the model could be exactly fitted numerically to the observations with respect to the locations, the result would be exact, and there would in fact be no obvious limit to resolution. The reason why in practice unlimited resolution cannot be achieved is that observations exactly describable by two component functions do not occur. Therefore, it is not the distance of the components, but it is the errors in the observations, systematic and nonsystematic, that ultimately limit resolution. During the last decades, a number of measurement error based resolution limits have been proposed in the literature reviewed in Ref. 10. One of the most recent ones will now be described.

Suppose that a number of two-component observations $w = (w_1 \dots w_N)^T$ has been made, and that the two-component model

$$a[h(x; b_1) + h(x; b_2)] \quad (31)$$

is fitted to these observations with respect to a , b_1 , and b_2 . Then, depending on the set of observations available, two essentially different types of solutions for a , b_1 , and b_2 may occur. In the first type, the solutions for b_1 and b_2 are distinct. This implies that the two components in Eq. (31) are resolved from the observations. In the second type of solution, the solutions for b_1 and b_2 exactly coincide. Then, the model corresponding to this solution is $2a h(x; b)$ with $b_1 = b_2 = b$. Thus, it is concluded that a one-component model is found as solution. This one-component solution is, of course, not found from exact two-component observations of the same functional family as the model fitted. However, it may result from error corrupted, two-component observations if the components seriously overlap.

At first sight, exactly coinciding solutions may look highly improbable. However, their coincidence is not caused by mere chance but by a structural change of the criterion of goodness of fit under the influence of the set of observations. In the

N -dimensional Euclidean space of the observations, where the n -th coordinate axis corresponds to the n -th observation w_n , a set of observations is represented by a single point. If two-component models are fitted, this space may be divided into two parts. For observations in the one part, the criterion has an absolute minimum with $b_1 \neq b_2$. For observations in the other part, only a minimum can be shown to exist with $b_1 = b_2$. The boundary of both parts separates the sets of observations from which the components can be resolved from those from which they cannot. Therefore, it is this boundary that constitutes the limit to resolution in terms of the observations. Of course, hypothetical, errorless two-component observations to which a two-component model of the same family is fitted are on the side of the boundary corresponding to resolution. However, nonsystematic and systematic measurement errors may move this point to the other side of the boundary, where resolution is impossible since the solutions coincide. Systematic measurement errors influence the location of the point around which sets of observations are distributed. This point represents the expectations of the observations. The systematic errors may move this point close to the boundary. The kind of distribution of the nonsystematic errors defines how the sets of observations are distributed around this point. Therefore, the probability of resolution is determined by both types of errors combined.

BIBLIOGRAPHY

1. C. Chatfield, *Statistics for Technology*, 3rd ed., London: Chapman and Hall, 1995.
2. A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed., Auckland: McGraw-Hill, 1987.
3. A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics—Vol. 1 Distribution Theory*, London: Arnold, 1994.
4. A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics—Vol. 2 Classical Inference and Relationship*, London: Arnold, 1991.
5. Anonymous, *Guide to the Expression of Uncertainty in Measurement*, 1st ed., Geneva: International Organization for Standardization, 1993.
6. W. H. Press et al., *Numerical Recipes in Fortran; the Art of Scientific Computing*, 2nd ed., New York: Cambridge University Press, 1992.
7. A. Grace, *Optimization Toolbox for Use With MATLAB™*, USCL's Guide, South Natick, MA: The Math Works, 1990.
8. R. I. Jennrich, *An Introduction to Computational Statistics*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
9. A. van den Bos, Complex gradient and Hessian, *IEEE Proceedings Vision and Image Signal Processing*, **141** (6): 380–382, 1994.
10. A. J. den Dekker and A. van den Bos, Resolution—A survey, *J. Opt. Soc. Am.*, **14** (3): 547–557, 1997.

ADRIAAN VAN DEN BOS
Delft University of Technology

AUTOMATIC TEST EQUIPMENT

Automatic test equipment (ATE) embraces a very broad array of test equipment ranging from hand-held meters to multiple racks of sophisticated instrumentation and *unit under test* (UUT) handling equipment all controlled by test programs. Automatic test equipment is applicable to the full spectrum from invisible light to direct current (dc). It involves an approach to testing rather than any particular configuration. Thus this

treatise addresses the principles and philosophies of automatic testing rather than specific designs of ATE. Various ATEs have incorporated just about every imaginable variety of instrument in existence. Design parameters for instruments used in ATE vary little from those of stand-alone instruments other than in their method of control. They follow industry developments, becoming smaller, more complex, and requiring less power as the years pass. Thus, examining the attributes of such instruments would be far beyond the scope of this chapter. Specifics of instrument design are covered in many other sources, including other chapters of this encyclopedia.

EARLY HISTORY OF AUTOMATIC TEST EQUIPMENT

The decade of the 1960s spawned a new class of test equipment that had its roots in the prototypes of the late 1950s. Known as ATE, these testers used the programmable feature of digital computers to store “intelligence” pertaining to a testing procedure or process. This stored program, coupled to a variety of programmable instruments, provided a capability to perform a wide variety of complex tests on a device or UUT with a minimum of operator knowledge or intervention. Automatic test equipment was first developed and introduced to meet field maintenance needs of the military.

By the mid-1960s, military electronics equipment and systems had increased in complexity to a point where well-trained technicians were required to maintain them. In the Navy, for example, the Class A electronics school course was 42 weeks long. Additional schooling was required to qualify a technician in specialized equipment. By the time a technician was adequately trained and gained a year or two of field experience, the enlistment was over. Most chose to accept a discharge rather than reenlist, because civilian work was less demanding and paid better. Thus the military faced a real problem in meeting the maintenance needs of ever increasingly complex electronics equipment and weapons.

Automatic test equipment offered a solution to the military maintenance problem. Accordingly, the army pioneered the development of ATE by funding prototype designs of ATE. One of the first testers was designed to test Atlas missiles. Other testers followed for communications equipment and advanced missile system components and subassemblies. By the late 1960s military applications of ATE abounded. The thrust then was to build very capable and therefore large, complex ATE with a very wide variety of testing devices (often referred to as building blocks) capable of testing almost anything in the current electronics inventory and even provide for testing systems yet to be developed. This multipurpose test equipment became systems with up to a dozen 5 ft racks of electronics and costing several million dollars per copy. Still they were called simply, ATE.

The promises of military ATE technology were fantastic. Suppliers claimed that an unskilled person with only a couple of weeks of training could now do the work of a skilled technician with several years of training and experience. This, of course, greatly reduced the cost of maintenance over manual testing procedures.

What vendors failed to mention was the enormous cost of developing a set of test programs and interface devices required to test that wide variety of UUTs. A test program had to incorporate a complete qualification testing process (referred to

as the *go chain*) as well as a complex network of fault isolation tests to diagnose the UUT when any go-chain test failed. This is analogous to anticipating all probability of failures during the lifecycle of a product and incorporating corrective actions even though in reality many of the failure modes might never be experienced. It is like buying a very expensive insurance policy and paying for it all at once rather than by modest premiums. It soon became evident that the claim of reduced cost of testing, at least during the early life of a product, was a myth; ATE was not nearly as cheap as believed. Yet, while not being as simple to use as claimed, it did greatly expand the capability of a trained technician to perform tests much more complicated than typical skills and time permitted. Thus ATE, in spite of its cost and impact on logistics requirements, became a necessary evil for maintaining complex military systems—ATE remains the backbone of modern weapon systems support.

In the 1970s ATE found a niche in the commercial manufacturing world. Here the challenge was not so much to test a wide diversity of complex assemblies, but rather to test large quantities of electronic components such as transistors and memory devices very rapidly and cheaply by minimizing costs and increasing manufacturing yield. As components increased in complexity ATE became an essential part of the manufacturing process.

For many years maintenance testing remained in the military domain, while manufacturing testing dominated the commercial world. Maintenance testing approaches the problem *top down*, starting at the system level and progressing to major assemblies, subassemblies, modules, and finally to the component level. Manufacturing testing generally works from the *bottom up*, beginning with basic materials and progressing to the finished product. Because of these basic differences and very different testing philosophies, there has been little cross-pollination between maintenance and manufacturing testing. Yet there is much to be gained from the common experiences of these seemingly independent areas of interest, because, broadly speaking, they do share a single technology. The basic principles apply to both maintenance testing and manufacturing testing. Differences and similarities are pointed out as applicable.

MAINTENANCE VERSUS MANUFACTURING TESTING

The basic principles of ATE and functional elements as described apply generally to both maintenance and manufacturing testing. The greatest differences lie in the testing philosophies and the design of the ATE hardware. Maintenance testing assumes that the UUT has been properly manufactured, and it once operated satisfactorily so the objective is to return the product to its original operating condition. Manufacturing testing is designed to assure that the product is properly assembled using tested components and that it is ready for packaging and sale. The failure modes of these differing processes are very different. The organization of the ATE hardware is also very different. Maintenance ATEs are durably constructed and fairly simple to operate in a field environment. Manufacturing ATEs are designed to be fast, flexible, reconfigurable, fully automatic, and suitable for factory environments. The most salient differences between maintenance and manufacturing ATE are identified in Table 1.

BASIC PRINCIPLES OF AUTOMATIC TEST EQUIPMENT

Fundamentally any combination of instrumentation controllable by a programmed sequence of events or algorithm can be considered to be an ATE. This sequence is commonly referred to as a test program. Once the system has been properly energized and initially set up, control of the instruments and the testing process is vested in the test program. In the early days of ATE, the control logic that executed the programmed instructions was of a design unique to the special needs of a given ATE. Soon, however, digital computers were generally utilized as the control system. They offer both improved flexibility and reduced cost over specially designed controllers. Today computers find multiple usage in ATEs, not only as the central control unit but also in many “building blocks” or instruments embodied in a system.

In its basic form, an ATE consists functionally of five subsystems: control, stimulus, measurement, switching, and system software. An ATE also contains various peripheral devices to facilitate user communication with the system. They include but are not limited to maintenance panels, input keyboard, printers, and output panel for connecting to the UUT interface device (ID). A simplified block diagram is shown in Fig. 1.

Control is generally provided by one or more general purpose digital computer(s) with associated memory. Switching consists of a network of electrically actuated paths interconnecting all of the internal elements of the ATE and the interface to the outside world. It is through the switching subsystem that the control subsystem *instructs* and actuates each ATE function as to when and how to execute the programmed instructions. It also provides paths for the ATE and its associated input/output devices, such as control panels and bulk memory storage devices. The stimulus subsystems comprises all of the power and signal sources required to energize the instruments internal to the ATE and the device or unit being tested. The measurement subsystem comprises all of the instrumentation required to measure both internal signal sources and responses from the UUT. The interface subsystem consists of all control panels, operator interface devices and linkages to the outside world, to which the UUT is connected directly or through an ID.

COMPARISON OF MANUAL AND AUTOMATIC TESTING

Manual and automatic testing have much in common. In both instances, the objective is to evaluate the item to be tested to verify proper performance or to detect and isolate a fault. Both approaches require several elements:

1. A way to energize and stimulate the UUT
2. Signal routing devices such as wire, switches, or relays
3. One or more devices or meters to condition signals and measure the signals involved
4. Controls and displays to allow interaction between the test technician and the hardware
5. A procedure, written or intuitive, that the technician follows in the testing process
6. The item or unit being tested

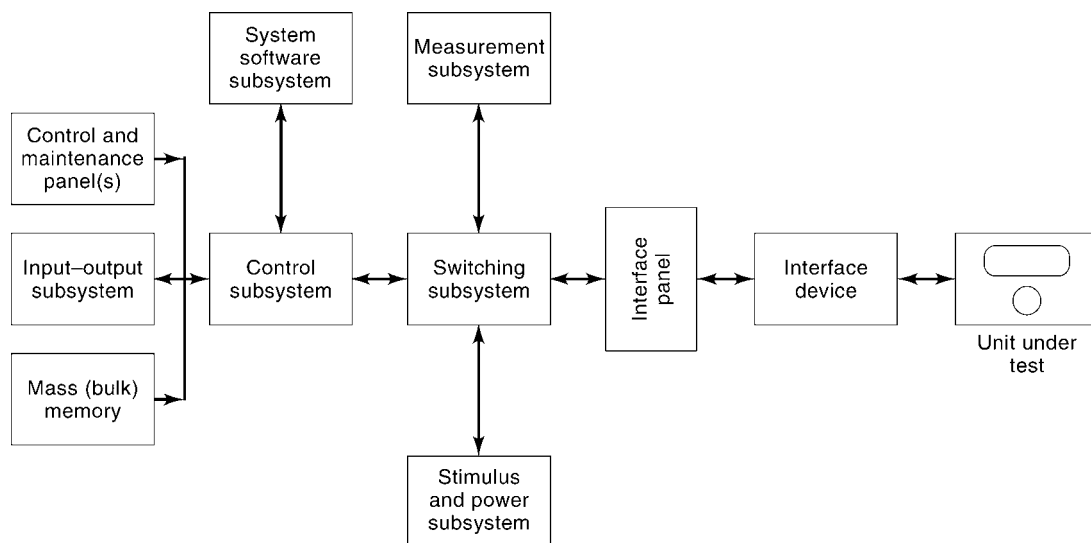
In the manual testing process these elements tend to be discrete instruments and components tied together by

Table 1. Comparison of Manufacturing and Maintenance Testing

	Manufacturing Test	Maintenance Test
First applications	Late 1960s	Late 1950s
Product perspectives	Speed	Integrity
	Cost to test	Life cycle cost
	Cost to produce	Cost of ownership
	Marketability	Reliability
Product mix	Large quantities	Small quantities
	Small variety	Large variety
	Inexpensive	Very expensive
Operating environment	Factory	Field
Test philosophy	Test to sell	Test to repair
	Discard	Salvage and repair
	Stand alone	Hierarchical
	Structural	Functional
Test logic	Bottom up	Top down
	Component to unit	Unit to component
Failure modes	Multiple	Single
	Out of tolerance	Catastrophic
Test techniques	Boundary scan	Logic tree
	Iddq	Guided probe
	Pseudorandom inputs	Truth table
ATE hardware	Standard commercial	Custom
Test languages	C, C++, custom	ATLAS, BASIC
Compilers	Incremental	Full pass
Test specification	Performance parameters	Test requirements
	original	traceable
Operators	Engineers	Technicians
Documentation	Listings	Flow charts
		User instructions
Validation	Compare golden unit	Insert faults
UUT handlers	Automated	Manual
Common needs:	Verification, validation	
Common tools:	Test design tools, compilers, simulators, automatic test generators, software support stations	

cables or wires. In an ATE most of these elements are embodied in the tester. If a good match exists between testing requirements and tester capability (referred to as assets), the only external elements are the interfacing connectors or cables connecting the tester to the UUT. The most signifi-

cant difference is that with the ATE the procedure used in testing is neither dependent on intrinsic knowledge or sequence of actions by a technician nor on a written procedure. It is embodied in the ATE memory and is called a test program.

**Figure 1.** Simplified block diagram of an ATE.

CONTROL SUBSYSTEM

The control subsystem is a functional entity rather than a single physical unit. In the early days of ATE it consisted of the system input and output devices such as a perforated tape or punched card reader, an operator control and display panel, a printer, and a uniquely designed logic network for interpreting and executing instructions fed into the system through the input device. Some of the earliest testers did not contain an internal memory but simply executed instructions directly from the input device. Later, specially designed computers with bulk and random access memories were developed for controlling the ATE. By the late 1960s commercial, general-purpose computers became popular because they offered greater versatility and reduced design and manufacturing costs. Some military ATEs used rather sophisticated, medium-scale computers, whereas those used in commercial manufacturing applications tended to use basic, 8-bit computers.

Once loaded with necessary software and a test program the role of the control subsystem is to execute all programs, both system level and application (test) programs. It works very much like any general-purpose computer, but instead of simply doing clerical work it actuates the various paths and instruments of the system as it interprets programmed instructions.

STIMULUS SUBSYSTEM

The most elaborate and costly subsystem of an ATE is the stimulus subsystem. It consists of all the power supplies, signal generators, and signal conditioners required to energize the UUT during dynamic testing, that is, operating the UUT in as close to its normal operational modes as practicable. In elaborate, general-purpose ATEs the stimulus subsystem can occupy several racks of signal generators, depending on the range of and complexity of the family of UUTs it is intended to test. Special-purpose testers require only a few signal sources, because the application is limited to a small variety of UUTs.

In early maintenance testers, the signal generators consisted of building bricks or chasses of standard test equipment with programmable controls. In recent years many of the signal sources have been reduced in size to plug in cards. These signal sources (building bricks) are often coupled together to generate complex signals from basic circuits. For example, instead of using three signal generators to produce square, triangular, and sawtooth waves, a single signal synthesizer is utilized, thus saving space and cost. Some building bricks, such as a frequency multiplier, can serve for either signal generation or as a measurement resource. Hence, the actual role of building bricks can vary depending on how they are programmed and for what purpose.

MEASUREMENT SUBSYSTEM

The measurement subsystem provides the capability to measure the responses from a UUT during the testing process. Collectively the elements of the measurement subsystems serve as a programmable multimeter. The measured responses from the UUT are compared to the programmed upper and lower limit stored in the test program. The control subsystem then

directs the testing sequence according to the results of the measurement.

SWITCHING SUBSYSTEM

The switching subsystem consists of all the programmable paths in the ATE required to connect the tester to and from the UUT through the interface panel, select the proper stimuli, choose the measurement instrumentation, and switch all of the tester resources as required for each test in sequence. Some testers incorporate a network of relays for internal signal routing, some use solid-state switching networks, and some have patch panels with wires connected as appropriate for the given UUT. Some testers use combinations of all of these switching mechanisms. Some utilize a *universal* routing system for maximum flexibility by allowing any tester resource to be connected to any other and to any pin in the ATE/UUT interface panel.

Some type of connecting mechanism is required to connect the ATE to the UUT. It can be as simple as a cable or as complex as a network of components within an enclosure. The general term for the interface hardware is the ID. In maintenance applications the ID typically consists of an assembly that mounts onto a patch panel. In manufacturing applications the ID can actually be a complex electromechanical unit for automatically feeding UUTs such as integrated circuit (IC) chips into test receptacles and sorting the tested chips into bins based on test results.

SYSTEM SOFTWARE SUBSYSTEM

As with general-purpose computers, an ATE requires several software subsystems or modules to effect its operation. The operating system provides all of the facilities required for operator interface, communicating with input-output devices, allocation of the system resources, and execution of various mathematical subroutines commonly used during the testing process.

No ATE is complete without its software. The software consists of a hierarchy of programs (or software models), most of which are intrinsic to the ATE design and not modifiable by the user. Among the software programs are the operating system, the self-test system, and the user libraries and test program. The operating system controls all of the basic operations such as loading the test program, calling and executing subroutines, executing the testing sequence, and driving the peripheral equipment, such as operator displays and printers.

System software is built up in hierarchical layers, much like the cross-section of an onion, as illustrated in Fig. 2. The core of the system consists of the basic machine instructions of the computer that is the nerve center of the control subsystem. The next innermost layer is the operating system followed by the utilities that facilitate *booting* up the system and input-output operations. Next is the on-line interpreter that converts input code to executable code during program execution. Next is the set of self-test programs that consist of a hierarchy of programs as described later. Next is a library of user subroutines and/or macros that can be initiated by the test program during run time to perform various, standard testing and/or to test set-up functions such as power on sequence of combining several stimuli outputs to generate a complex signal or waveform. Finally there is the specific test program for the UUT to be tested.

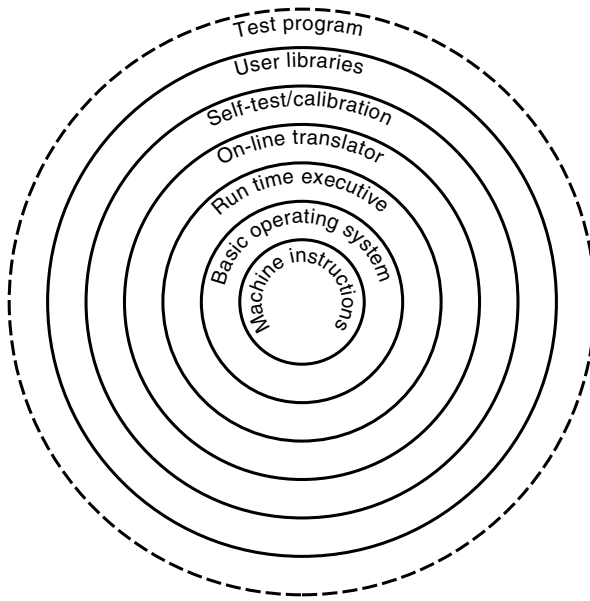


Figure 2. System software hierarchy.

MACHINE INSTRUCTIONS

Computers are designed to execute a set of basic instructions such as add, subtract, store, and get, that provide the capability to sequentially perform arithmetic functions and data manipulations. These functions are performed by logic circuitry that interprets *words* composed of different combinations of ones and zeros. It is only through the addition of various software programs that ascribe specific meanings to these words that the user is able to communicate with the computer in an English-like language.

OPERATING SYSTEM AND RUN TIME EXECUTIVE

The operating system is a program that is fundamental to controlling the activity of the computer. It is a fixed part of the ATE design that the user cannot modify. As in most computer systems the operating system is the *traffic controller* for the various activities needed to allocate the system resources such as the arithmetic unit, executable and bulk memory, and input and output devices. In an ATE there often is also a *run-time* executive program that can be considered part of the operating system. It interprets programmed test statements and sends appropriate signals to the ATE resources.

TRANSLATORS

Translators simply interpret the source language statements, check them for proper format, and generate equivalent object language statements without attempting to evaluate the validity of the statements to perform the appropriate action in the ATE. Usually translators are used only to make changes to a test program during program validation rather than to generate the original object program. Hence they are often resident on the ATE as a auxiliary tool called into service through the maintenance panel when needed. Some ATEs utilize an on-line

translator that converts the object program to executable code during test time. Its use is automatic and invisible to the user. It is one of the functions performed by the run time executive program.

COMPILERS

Interpretation of English-like (source language) test statements are accomplished by either compilers or translators. Compilers are not generally resident in the ATE. Rather they are hosted on a powerful, off-line data processing system. Hence no compiler is shown to be part of the ATE system software in Fig. 2. Compilers do much more than translate source language to ATE language. They generally screen test procedures to assure that the target ATE resources are available to perform the functions called out, that none of the tests programmed violate safe rules for using the ATE, and sometimes even automatically allocates ATE resources such as signal sources and generates the linkages to apply the signal to a specific set of pins on the ATE external connection panel.

SELF-TEST

One of the features of an ATE is the ability to test itself. To accomplish self-testing, however, certain features must be incorporated in the stimulus and measurement building bricks, and the diagnostic programs must be designed and incorporated in the ATE. The stimulus signal sources must have self-test outputs that can be internally routed to the ATE measurement subsystem. These outputs can be used to evaluate stimulus assets during self-test or used to measure current signal settings prior to being applied to the UUT. Likewise, the measurement subsystem must have entry points for the selected stimulus signals to be routed to the proper measurement circuitry. These features are intrinsic to the ATE system hardware design. They cannot be added on later or implemented effectively with software only unless a network of programmable switches are provided to route signals internally. That is a poor alternative because of the signal losses or contamination that is introduced by long signal paths and switching devices.

When the required hardware features are provided to internally link the stimulus and measurement subsystems, a set of self-test programs can be incorporated in the ATE. Self-test capabilities range from a simple set of system *health checks* to a multilevel hierarchy of self-test programs, some of which are automatically executed during UUT test time or as more detailed tests selected at the user's option. A comprehensive set of self-tests are described in conjunction with the illustration of Fig. 3. Not all ATEs will provide such a complete set of test programs but they are desirable features to look for in selecting an ATE. Although these features add to the design cost of an ATE, that cost is amortized across all of the ATEs of a given design and not borne by a single customer.

SELF-TEST HIERARCHY

Just as system software is built up from a hierarchy of interdependent software modules, self-testing is also a hierarchical set of software modules. Figure 3 shows an expansion of the self-test function indicated in Fig. 2. The software modules shown

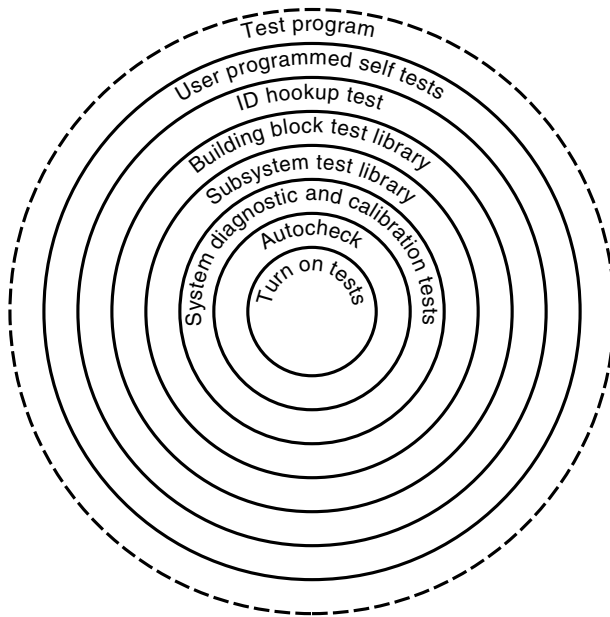


Figure 3. Self-test software hierarchy.

in Fig. 3 are typical but not always organized in the fashion shown.

TURN-ON TESTS

Automatic test equipment is designed to provide maximum versatility to the user. This versatility can lead to misuse and possible damage to the hardware or even harm to the operator. Accordingly, a well-designed ATE incorporates safeguards that preclude damage to the hardware or danger to the user. These are sometimes called *safe to turn on* tests. For example, they check to see that the hook up between the ATE and UUT will not draw excessive current or that high voltages are not connected to low-voltage terminals or where they are exposed to operator controls.

AUTOCHECK

After a safe turn on, the system automatically performs a set of basic tests of the ATE subsystems to assure that the system is *healthy* enough to be used. These tests, called *autocheck* continually monitor the health of the ATE during normal testing of a UUT. They perform routine measurements of stimuli and check for proper operation of measurement circuits. If any test fails, testing of the UUT is halted and the operator is advised that there is a problem. Generally the fault is not isolated by autocheck, but it does instruct the operator as to which set of diagnostic tests should be performed.

SYSTEM DIAGNOSTIC AND CALIBRATION TESTS

Some ATEs include secondary standards, such as standard voltage cells. These, together with the measurement subsystem, can be used to calibrate many of the system resources. Tests can be incorporated in the basic system design to con-

tinuously or periodically check the calibration of the most sensitive circuits of the ATE. These tests would automatically be executed without the need for the UUT test program being impacted unless a capability required falls below acceptable tolerance. Hence calibration tests are at the very core of the self-test hierarchy.

These consist of a library of tests generally resident in the ATE but initiated by the operator in the event of an autocheck indication of failure. They check out the system in more detail to automatically adjust calibration of the ATE resources or, when not corrected by calibration, to indicate which subsystem has failed. The operator or maintenance person can then initiate testing with the indicated subsystem diagnostic tests from the subsystem test library.

SUBSYSTEM TEST LIBRARY

This is a library of test routines grouped according to the ATE subsystems. Because the system diagnostic tests indicate which subsystem tests should be performed a logical linkage exists. It may be necessary to load in the appropriate subsystem diagnostic test program or it could be resident in the ATE so that it is selected from a maintenance panel. Normally the operator must call in an authorized maintenance person, because the maintenance panel can only be activated by a key. Once loaded the subsystem test program identifies the faulty assembly (or ATE building brick) requiring repair. Ideally the faulty assembly can be replaced with a spare, and the ATE can be returned to service. The diagnostic test must be repeated to assure that the maintenance action solved the problem. The faulty assembly can then be tested as a UUT using building-brick tests.

BUILDING-BRICK TEST LIBRARY

The building-brick test library is a set of test programs comprised of individual test programs for each ATE building brick. They are hooked up to the ATE through the standard interface panel and the appropriate ID just as any UUT would be tested.

USER LIBRARIES

As experience is gained through significant use of an ATE, commonly required tests of specific elements of the system are developed into standard test sequences. They can be implemented as subroutines or macros. The purposes vary but can include self-tests for specific building blocks known to be critical to a given test procedure or simply sets of instructions often required to perform complex measurements such as for gain measurement. This library is generally developed by the test design team and developed over a period of time as new ways are found to reduce test design time and/or increase standardization. Such a library is limited only by the ingenuity of the users.

THE TEST PROGRAM

At the top of the software hierarchy is the test program. It is the real workhorse of the ATE. All other software in the lower rings of the system are there to facilitate testing. The effectiveness of an ATE is dependent on a good test program design.

TEST PROGRAM SET DESIGN AND PRODUCTION

A test program set (TPS) consists of a test program, an ID, and a test program instruction. Producing a TPS for other than a trivial UUT is a complex, time-consuming, tedious process. It often involves a team effort with a variety of skills and a whole host of specialized hardware and software tools as well as the target ATE. However, it is divided into five fundamental processes performed more or less in sequence, but with multiple feedback loops when things don't go right the first time—and they rarely do. The five processes are test design, program production, ID design and fabrication, validation, and demonstration and acceptance.

TEST DESIGN

Test design is the most challenging part of the test program development process, requiring a skilled, technically competent individual with both design and maintenance experience. It requires years of training and experience to become adept in the science and art of test design. Neither this document nor any other can compensate for a lack of testing experience. However, an overview is offered here with some valuable tips and recommendations based on many years of TPS development experience applicable to maintenance testing.

Test design is the process of developing the testing requirements of the UUT and the logical sequence or flow of the intended process of testing. For maintenance applications the testing information required is derived from many sources. The performance tests, referred to as the go-chain tests, are usually derived from the factory acceptance test (FAT) procedure. It should provide basic information on the acceptable performance requirements for a properly operating UUT. However, the FAT is most likely performed with test equipment other than the target ATE and the test process is designed to quickly check if the UUT has been properly assembled and tries to detect typical manufacturing errors such as solder splashes, open connections, or improperly mounted components. So although the go-chain tests might be usable in maintenance testing, the fault detection and isolation tests are not. And the go-chain tests most likely will have to be modified to take into account signal deterioration due to signal routing in the ATE as well as making the test limits less demanding to account for acceptable performance degradation or UUTs in service.

Sometimes a test requirements document (TRD) is provided by the UUT manufacturer. It identifies input and output signal requirements for normal testing of the UUT as well as some fault mode information. It may also recommend the proper sequence of testing. Test requirements documents are not cheap and often are hastily produced and contain numerous errors and voids. Making the producer of a TRD accountable for its completeness and accuracy is not easy. Ultimately the burden for proper test design rests on the test design engineer, regardless of how good the data package may be.

Software tools are sometimes available to assist in designing some tests. Such a tool is often called an automatic test program generator (ATPG). It consists of a powerful, special simulation program operable on a general-purpose computer. The program automatically generates test patterns based on a defined digital logic design. These patterns must then be integrated into the remainder of the test program. Some ATPGs

can handle limited hybrid circuit simulation, but to date none provide a comprehensive analog circuit simulation capability.

The test designer need not know exactly how the ATPG works, but must become adept at interacting with it because ATPGs are not really as automatic as they profess to be. Specialized training is required to become adept at using an ATPG, and the availability of expert consultants may also be required. Even with the aid of an ATPG, much of the test-design process depends on the ingenuity of the test designer.

The objective of the test-design process is to develop a detailed, unambiguous, step-by-step procedure that becomes the controlling program used by the ATE to test a UUT. It is up to the test designer to "think as a machine thinks," while generating tests consisting of a series of interrelated tests pertinent to the UUT involved. Here, an individual test is defined as a set of instructions needed to set up the ATE instrumentation (sometimes referred to as ATE assets) as required to energize the UUT, route the ATE signals to the proper ATE output panel and interface device connectors, provide a signal response path from the UUT to the selected ATE measurement instrument, compare the response to stipulated upper and lower value limits, and direct the ATE to the next test to be performed based on the results of the current evaluation. The sequence of tests and all alternative branch paths between tests is referred to as the test-flow process. It is best recorded as a test-flow diagram showing all go-chain tests with appropriate branching to fault isolation tests. A typical test involves proper setting of all the parameters needed for a single measurement of a UUT response followed by a single decision in a three way branch in the program logic, based on a *high*, *low* or *go* test result. If the measured value is within the programmed limits the go branch is taken. If the measured value exceeds the upper limit, the high branch is followed. If less than the lower limit, the low branch is followed.

ESTABLISHING TEST LIMITS

One of the more difficult yet critical tasks of the test engineer is establishing upper and lower limits for each test in the program. The recommended approach is to begin the calculation with the *when-new* or manufacturer's specified limits for normal operation. These are the ideal limits which must be relaxed or made less demanding based on anticipated adverse impact due to expected environmental conditions of UUT usage, expected aging due to component deterioration while in use, and measurement inaccuracies of the ATE (including distortion and crosstalk). The final upper and lower limit values should be as broad as allowable for acceptable operation of the UUT in its normal environment but never beyond.

The sequence of tests begins with checks to assure that the UUT has been properly connected to the ATE via the proper ID. These tests are followed by several safe-to-turn-on tests that assure there is nothing in the setup that could damage the tester. The UUT tests begin with static tests to assure that there are no undesired open or short circuits in the UUT. Static tests are conducted without power being applied to the UUT. Having safely passed these tests power and signals are applied to the UUT to approximate normal operating conditions. Tests executed while the UUT is energized are called dynamic tests. They comprise the bulk of the testing process and require the greatest design skill.

Dynamic tests are generally grouped according to similar input signal requirements to minimize setup and signal settling times. It is also good practice to provide *break points* in the testing sequence. These are places in the testing process where it is safe to enter the process without having to begin with the first test. Break points assure that all ATE parameters are properly set. They are very handy during test design validation and also when attempting to repeat selected tests during a UUT retest.

Sometimes the test operator must interact with the testing procedure by performing manual actions. For example, the UUT may have operator controls or mode-setting switches. The test is not considered complete until each position of the switches or other UUT controls have been tested. The test designer must anticipate the need for manual intervention and program a stop or *pause* instruction and send a message to the operator regarding the action to be taken. When the manual action is completed, the operator initiates a *continue* action and the automatic testing sequence resumes.

When all parameters of a normally operating UUT have been programmed in a sequence of tests, the performance tests (referred to as go-chain tests) are considered complete. It is often necessary to redesign some of the tests based on actual results observed during the validation process as is discussed later. The entire go-chain should be designed before attempting to develop the fault isolation tests (referred to as the *no-go tests*). Results of each test are generally printed out together with a final judgment as if the UUT is okay or faulty.

Fault detection and isolation tests can be partially derived from deductive logic based on which go-chain tests fail. However, most no-go tests must be determined from anticipated field failures. Rarely is such information provided to the test designer, so developing it becomes one of the greatest challenges for the test design engineer. For this task there is no substitute for practical experience and, if available, help from the UUT designer.

When all anticipated failure modes (ways in which the UUT is anticipated to fail) have been accounted for in the test design the no-go chain is completed. Often the information concerning realistic field failures is not given in the source documentation on which the testing is to be based. The preferred solution is to hook up the UUT in a laboratory environment and induce failures in the UUT one at a time to determine the results of each anticipated failure. Without such bench testing the fault isolation tests are theoretical and are often found to be defective during validation when the UUT is attempted to be tested on the ATE. At the end of the fault isolation tests the ATE indicates any failed tests as well as notifying the operator what the most probable component in the UUT has failed. Once repaired, the UUT must again be tested to assure that it is now operating correctly. Usually only one fault is found at a time so it may take several passes in the ATE before the UUT is *ready for issue*.

TEST PROGRAM SET PRODUCTION

Test program production involves generating the object language test program (the language acceptable as an input to the ATE) that dictates the testing process, assembling the ID from existing hardware or building a prototype (called a brass board), and generating the TPS documentation. Generating the object language statements can involve two steps:

(1) Translating an English-language test procedure, such as a TRD, into test programming language (referred to as the source language). (2) Translating the source language statements into the object language. Many programming languages have been used for source code. The IEEE standard language ATLAS (automatic test language for all systems) has been used most for military applications. The airlines maintenance community generally uses a specialized version of the ATLAS language. Commercial manufacturing applications tend to utilize more general-purpose and flexible programming languages such as Fortran or C+ but they are not test-oriented languages so they require development of a specialized test statement dictionary developed by the user so they require a greater depth of programming skill. A good test programming language allows the test engineer to compose a program in engineering English statements, such as set supply A to 15 V, dc or connect supply A to pin 3 of J24. ATLAS provides such a capability but many versions or dialects of ATLAS have been used so that there are special instructions required for each dialect. The proper dialect for any given tester is generally defined in a test programming guide.

Once encoded in source language the source program must be converted to the object language peculiar to the ATE. Conversion is usually performed on a software generation station consisting of a general-purpose computer containing a compiler. A compiler is a program that not only translates source code to object code, but also performs a number of validity checks on the programming logic, allocates memory spaces, and performs numerous other *housekeeping* chores needed to handle test program execution during testing (referred to as run time).

Compiler design is a highly specialized technology relegated to skilled programmers and not the domain of the test engineer. However, the test engineer must become adept at using the compiler and resolving problems *flagged* in the compiler print-out or what is called the *object code*. The program is not considered completely compiled until all flagged statements have been corrected or resolved. It may take several passes through the compiler to get a clean object code listing that can be used during program debug and validation.

INTERFACE DEVICE DESIGN AND FABRICATION

The nomenclature of ID may be misleading because interfacing a UUT to an ATE is generally a complicated process. The term *interface device* was coined many years ago when a simple adaptive device such as a multipin connector or cable was all that was required to hook up a UUT to one of the early ATEs. Today a fairly large box is often required that contains a group of connectors, cabling, and some electrical components. Such an ID is typically used to interface a related group of circuit modules. In military terms the modules are generally called *shop replaceable assemblies* (SRAs). These consist of quite complex circuit boards or assemblies taken from larger UUTs called *weapon repairable assemblies* (WRAs) or *line replaceable units* (LRUs). Typically in aviation maintenance applications WRAs are 30 to 40 lb assemblies removable from an aircraft at the flight line, whereas SRAs are removed for testing in a maintenance shop. A WRA can contain 10 or more SRAs and a typical military aircraft can contain 80 or more WRAs.

Interface requirements linking UUT to the ATE are best determined and recorded concurrently with the test design

process. A standard ID wiring form is recommended with ATE panel connections at the left side and UUT connection points on the right. Adequate space should be provided for drawing wiring and component schematic symbols. Because an ID is often shared among several UUTs, it is important that there be no interference between diverse connections required by each UUT. Unique connections required by UUTs can usually be accommodated by programmable switches in the ATE or, if unavailable, relays may have to be incorporated in the ID design. If specialized signals are required that are not available from the ATE, signal conditioning circuitry may have to be incorporated in the ID. But every attempt should be made to avoid the use of active components in the ID, because they become likely candidates for failure.

Once the ID requirements are documented, they are submitted to a shop for manufacturing. A "brass board" preliminary design is recommended until the TPS has been validated, since changes are likely. After validation the ID should be manufactured as a deliverable item using good design and manufacturing practice.

The ID design is normally represented by a schematic drawing generated by the test designer at the same time as the test procedure is programmed in source language. It shows the connections between the ATE and the UUT with any intervening parts. Ideally only standard cables or patch panel connectors are required when all signals from and to the ATE are compatible with the ATE. Often, however, there are some components that must be included in the ID. They might include load resistors or special signal conditioning circuits. It is up to the test designer to specify the parts and the design of any nonstandard circuits. Assembly and testing of the ID is usually done by a technician.

In maintenance testing applications IDs are custom designed by the test design engineer to meet unique needs of a given UUT or set of UUTs. In manufacturing applications of ATE, interfacing the item to be tested to the tester is accomplished with a standard design *handler* that is part of the ATE system. Interfacing the tester to the items being tested typically involves a complex and expensive handler that automatically accepts each item to be tested from a *hopper* containing hundreds or thousands of items such as computer chips and feeds them in proper orientation to a standard test jig. There each is tested and fed to separate collecting chambers according to how it passed the tests. For testing circuit boards the interfacing is often accomplished by a *bed-of-nails* set of probes that make all contacts with the circuit board in a single operation. The tests are then performed and the next board is automatically loaded to the test position. Test results are automatically printed out and keyed to the pertinent board by serial number.

TEST PROGRAM INSTRUCTIONS

Test program set documentation is referred to as the test program instructions (TPI). Documentation requirements can vary widely depending on the needs and skill of the user. As a minimum it consists of source and object program listings, ID schematic, and instructions for the ATE operator. It may also include a flow diagram of the testing process and general operating procedures for the ATE operator.

During test design and ID design the test designer should also prepare notes on instructions that must be given to the

test operator for use while testing the UUT. Generally, this includes initial hook-up instructions for the ID, any precautions to be taken, and associated manual actions required of the operator to start the testing process. This information when properly edited becomes the TPI. Traditionally this is treated as a formal, hard-copy publication, deliverable as an element of the TPS. A better approach is to incorporate all instructions to the test operator within the test program. Initialization of the standard testing process can be built into the ATE in a resident *boot strap* program that is automatically executed when a start button is pressed on the ATE control panel. Information to be passed to the operator during testing, whether requiring operator intervention or simply informing the operator as to what is happening or relaying test results, should also be incorporated in the test program so that there is a constant communication and a symbiotic interaction between the operator and the ATE. With current technology and equipment there is no need to depend on hard-copy documentation for any part of the UUT testing process or even the trouble-shooting and repair of the ATE.

Once the test program listings, the ID with its schematic, and the draft copy of the TPI are available the test designer can begin TPS validation on the ATE.

VALIDATION

Validation is the process of testing a TPS on the target ATE to confirm that the testing process is valid and accomplishes the intended purpose. This process should be distinguished from the process called verification, which simply assures that the program is a faithful representation of what the test designer stipulated in source language statements during test design; not that the test design and ID are doing what is really required to test or fault-isolate the UUT. Verification can be performed by a good compiler but usually requires some *desk checking* of the compiler printout by the test designer to be sure that the program is ready to be tried on the ATE with the UUT hooked up and activated.

Once the compilation has been verified and the ID brass board is assembled and bench tested, the TPS is ready to be validated on the ATE. A known good (properly operable) UUT is required for the process of validating the go-chain. At least one additional known good UUT should be tested on the ATE to be sure that the testing process and operating limits of the program are generally valid and not unique to only one UUT of the type being tested. Go-chain validation should be completed before any attempt is made to validate the fault isolation (no-go) tests. Once the go-chain has been validated a sample number of anticipated field failures should be induced in the UUT, one at a time. Careful analysis of the UUT design is required to assure that induced faults will not damage the UUT. What failures are considered reasonable field type or operational failures and what percentage of anticipated failures should be induced during the validation process is a tough judgment call. It depends on the degree of confidence that the customer is willing to pay for. Formulas have been developed to assist in the process of computing the degree of confidence and the level of fault isolation that is desired. Some guidance along these lines is provided in future chapters.

It is important that the test designer has a clean compiler listing, tested ID and its schematic, and UUT documentation

prior to attempting to validate a TPS. Auxiliary external test equipment may also be needed to examine signals at interface points when the signals programmed or received from the UUT are in doubt. If a lot of validating is planned, it is a good idea to have an experienced trouble-shooter to assist in the validation process. Time on the ATE is very precious, so validation should proceed as rapidly as possible. That is not the time to design or redesign tests. One should note each problem and move on to the remaining tests unless precluded by the impact of previous failures. Problem solving and test redesign should be done off line. Validation can then be resumed at a later time when machine time can be scheduled. Any problems experienced with the ATE should be recorded in a validation logbook.

A TPS cannot be considered fully validated until a complete printout is obtained indicating all values measured with *go* results and printouts showing measured values for UUT failures with appropriate results and instructions printed for the test operator.

Demonstration and Acceptance. Once the TPS has been debugged and validated, some sort of demonstration is required to *sell off* the TPS to a customer. In the military sell off is a formal, well-documented, and explicit process. In a factory application it may be simply demonstrating to the production manager that the test program and ID operate as required. In a formal sell off, the documentation package describing the testing process is reviewed and accepted. Next the TPS operation on the ATE is demonstrated. The *go-chain* is executed to show that an operating UUT can be properly tested as operational. Then an agreed upon set of faults are inserted into a UUT, one at a time, to show that faults are identified by the testing process. Once an acceptable number of faults have been shown to be covered by the process the customer's representative signs the acceptance document.

Built-in Test (or Built-in Self Test)

In recent years, as components and systems have become much more complex, considerable attention has been given to designing devices, modules, and systems so that they are more readily testable. Since the cost of testing logic chips in the manufacturing process represents as much as 30% of the production cost, it has become cost effective to include built-in-test (BIT) features, which reduce manufacturing test time and the overall cost of producing the product. Typically approximately 10% of the *real estate* (surface area) of a chip is devoted to testing circuitry. A popular technique is called boundary scan. This methodology provides program-controlled access to the internal circuits of a chip to allow sampling internal circuits without the need for an external probe.

In maintenance applications, the BIT features introduced in the manufacturing process is utilized to isolate faults in the product's components. In large systems such as aircraft, system-level BIT is usually incorporated in the design to assist in maintenance testing. BIT becomes a subsystem of the design and provides in-flight health monitoring as well as indications of particular failures. Identification of a fault allows the pilot to select alternate modes of operation during flight. Once landed the BIT subsystem can be queried to determine which subassemblies need to be replaced to restore the aircraft to full operation. In the maintenance shop the information gained from the BIT circuitry can be used to help isolate faulty

components and thus facilitate the repair of the failed units removed from the aircraft. Automobiles are also being equipped with some BIT features.

Simulation as a Test Tool

Simulation provides a means for modeling a circuit, network of circuits, or even an entire system on a computer. Simulation finds greatest utility in the design process but can also help in test design. For many years software programs have been used to simulate digital circuits and networks and automatically generate test patterns needed for testing. More recently some simulators have incorporated test generation features for hybrid networks (a mix of analog and digital circuitry). Simulation of complex analog networks has not been very successful due to the complexity of developing effective models. Simulators used in circuit design generally do not provide for simulating fault modes as required for maintenance testing, so they offer only limited utility for test design. Future design simulators very likely will begin to address the needs of the test designer as well as the product designer.

Golden Card as a Testing Standard

Some manufacturing applications have utilized a circuit card or subassembly known to be operating properly as a standard against which newly manufactured cards of the same design can be tested by comparing input and output signals. The standard card is referred to as the *golden card*. This approach works well in some applications and is easier to use than a software simulator. The danger lies in the possibility that the golden card could fail and result in erroneous testing of the new cards. A more effective technique is to develop a software model of the golden card to use as a standard since once software is properly designed and tested it does not fail.

Guided Probe

Some commercial ATE includes a feature called *guided probe*. It is a system software module that helps in trouble-shooting a circuit card or assembly. In the probe mode of operation the ATE directs the operator where to put a signal probe in each of a series of points while the tester evaluates the signal from each point probed. Thus by use of a trouble-shooting algorithm the ATE leads the person to identify the faulty component. This technique can be more effective than relying only on access to the circuitry from input and output connectors, as is the case with many testers.

BIBLIOGRAPHY

Although ATE is utilized extensively in both manufacturing and maintenance applications, surprisingly few textbooks have been published on the subject. A search through the Library of Congress index of published books reveals that by 1997 only four have been published that are devoted to ATE since the first was published in 1974 by IEEE Press. The three that could be classified as textbooks are entirely devoted to manufacturing applications of ATE with no mention of the multibillion dollar market for maintenance testing utilized principally by the military and the airline industry.

The principal sources of documentation are the periodic symposia, conferences, and workshops dedicated to various ATE

specialties. They provide both invited and volunteered papers which are subject to peer review. Copies of recent issues are generally sold directly by the technical societies sponsoring the events. Some conferences offer tutorial sessions that provide valuable prerequisite information covering basic principles or advanced techniques. These meetings, with their respective publications, are the best source of technical information, because they are directed at a wide variety of applications and tend to contain the latest developments in ATE technology. The International Test Conference is devoted almost exclusively to manufacturing test and AUTOTESTCON is almost exclusively devoted to maintenance testing.

There are many periodicals, generally published monthly, that provide a medium for publishing volunteered papers of variable quality. Industry studies are common but they stress marketing data rather than providing a source of technical information. Other publications, such as user guides, deal with the specifics of particular ATEs so they are of limited value for general education since all ATEs are different.

- K. Brindley, *Automatic Test Equipment*, London: Butterworth-Heinemann, 1991.
 - R. Knowles, *Automatic Testing Systems and Applications*, New York: McGraw-Hill, 1979.
 - F. Liguori (ed.), *Automatic Test Equipment: Hardware, Software and Management*, New York: IEEE Press, 1974.
 - A. C. Stover, *ATE: Automatic Test Equipment*, New York: McGraw-Hill, 1984.
- Some useful conference records are as follows.
- Automatic Support Systems Symposium for Advanced Supportability*, New York: IEEE, published annually between 1965 and 1975 except for 1971.
 - AUTOTESTCON*, New York: IEEE, published annually from 1976 to the present.
 - International Test Conference* (successor to the Cherry Hill Conference), New York: IEEE, published annually from 1970 to the present.

FRED LIGUORI
ATE Consulting Services

SENSORS AND TRANSDUCERS

ELECTRIC SENSING DEVICES

In a measuring chain or a measuring instrument, the sensor is the element that is directly affected by the measurand (1). Specifically, an electric sensing device converts the quantity to be measured into an electric output signal (generally a voltage) (2). Today electric measurement systems have replaced most nonelectric systems. Only in environments where ionizing radiation or explosive atmospheres are present, nonelectric (mechanical, hydraulic, pneumatic) or optical signals may be more suitable (3). The advantages of using electric sensing devices are numerous and are related above all to the current technology in electrical signal elaboration, recording, and transmission. There are, in fact, even sensors that incorporate integrated circuits for electric signal conditioning or modification in a single package and consequently provide outputs more suitable for the following elements of the measurement chain. Moreover, in the automatic control system where the sensor's outputs have to be elaborated by a numerical processor, it is very easy to convert an analog electric signal to a numerical form.

It is possible to design an electric sensor for measuring any nonelectric quantity. Because of the electronic structure of matter, by selecting appropriate materials, any variation in a nonelectric parameter yields a variation in the electric parameter (3).

Measurand quantities may be grouped on the basis of the form of energy in which the signal is received by the sensor: mechanical, thermal, electrical, magnetic, radiant (optics), or chemical. In Table 1 a limited number of significant examples belonging to each class are reported (4). Often more than one transformation step is required to generate an electric output signal. It is possible to (2) define *primary measurand* quantities directly sensed by electric sensor whereas *secondary measurands* include a combination of primary measurands (5). There are several physical effects that generate electric signals in response to nonelectric influences. It is very difficult to group the countless types of sensing devices. Table 2 gives an interesting summary of the most widely used physical principles of sensing with the electric output and the typical measured quantities (6).

Table 1. Examples of Measurand Grouped by Their Domain

Type of Sensor	Physical Quantities
Mechanical	Displacement, length, strain, velocity, acceleration, mass flow, force, torque, pressure
Thermal	Temperature, heat, heat flow, entropy, humidity
Electrical	Voltage, current, power, charge, resistance, inductance, capacitance, dielectric constant, electric field
Magnetic	Field intensity, flux density, magnetic moment, permeability
Radiant	Intensity, phase, wavelength, polarization, reflectance, transmittance, refractive index
Chemical	Composition, concentration, reaction rate, pH, oxidation

MATERIALS USED FOR SENSING DEVICES

Metals

Metals are characterized by high electric and thermal conductivity, high optic reflectivity, high deformability, plasticity, mechanical tenacity, and high electronic extraction potential. Alloys or intermetallic compounds conserve these qualities to a high degree. Consequently, many sensors operate on the basis of the variation of physical proprieties of pure metals or alloys following variations of external quantities. In general, from the standpoint of sensor design, there are two classes of metals: nonferrous and ferrous. Ferrous metals, like steel, are often used in combination with a magnetic sensor to measure motion, distance, magnetic field strength, etc. Nonferrous metals offer a wide variety of mechanical and electric proprieties but are permeable to magnetic fields and used whenever these fields are not involved. The major use of nonferrous metals is to produce thermoresistances and thermocouples (5). Metals also exhibit piezoresistant effects, but because they are more sensitive to temperature, the production of strain gauges by using metal alloys rather than pure metals is preferred.

When selecting a metal for a sensor design, one must consider its physical proprieties, and also the relative ease of mechanical processing. For example, copper has excellent thermal and electric proprieties, yet is difficult to work with. So in many cases aluminum is considered a compromise alternative (5).

Ceramics

Ceramic oxide materials play an important, steadily increasing role in almost all fields of electronics. In sensor technologies, ceramics are very useful because of their structural strength, thermal stability, light weight, resistance to many chemicals, ability to bond with other materials, and excellent electric properties (5). There are, moreover, abundant opportunities for optimizing performance and tailoring to specific demands by modifying their chemical composition and/or varying their microstructure by changing parameters in manufacturing. There are ceramic materials with piezoelectric, pyroelectric, and ferroelectric proprieties (4).

Glass

Glass is characterized by transparency, hardness, and resistance to most chemicals. The main type of glass used in sensor technologies is lead alkali, which is a good electric insulator and is used for manufacturing optical windows, prisms, and nuclear radiation shields. Furthermore, optical fibers can be realized from glass (5). Optical fiber sensors can be separated into extrinsic and intrinsic types. In extrinsic fiber-optic sensors, sensing process takes place externally to the fiber, and the fiber itself plays a passive role as a light conduit (characterized by a very low attenuation factor). In intrinsic transducers, however, the measurand interacts locally with the light in the fiber by changing a parameter of the fiber, such as the refractive index, and the fiber in turn modulates the light beam propagating through its core (4,7).

Table 2. Physical Principles of Electric Sensing Devices

Physical Principle	Typical Application	Measurand	Output
Resistive The variation of the sensing element electric resistance depends on the measurand.	Thermistor or resistance thermometer	Temperature	Resistance
	Potentiometer	Displacement, force, pressure	Resistance
	Hot-wire anemometer	Flow	Resistance
	Resistive hygrometer	Humidity	Resistance
	Chemioresistor	Presence of gas	Resistance
Capacitive The sensing element capacitance depends on the measurand.	Parallel-plate capacitor sensor	Displacement, force, liquid level, pressure	Capacitance
	Rotary-plate capacitor sensor	Displacement, force, angular position, torque	Capacitance
	Differential capacitor	Small displacement	Capacitance
	Capacitance manometer	Very low pressure	Capacitance
	Humidity sensor	Moisture	Capacitance
	Capacitive diaphragm	Pressure	Capacitance
Inductive The sensing element inductance depends on the measurand.	Linear variable differential transformer	Displacement, torque	Inductance
	Self inductance sensor	Displacement, torque, liquid level	Inductance
	Eddy current sensor	Position, conductivity, thickness, cracks in materials	Inductance
Reluctive The variation in the reluctance path between two or more coils depends on the measurand.	Linear variable differential transformer	Linear displacement	Voltage
	Rotary variable differential transformer	Angular rotation	Voltage
	Microsyn	Angular displacement	Voltage
	Resolver	Position	Voltage
	Syncro	Position, torque	Voltage
	Reluctive diaphragm	Pressure	Reluctance
Piezoresistive effect Resistance of the sensing element depends on the strain.	Strain gauge	Stress, strain, fluid pressure, displacement, force	Resistance
Electromagnetic In any circuit capturing a magnetic flux, whenever the flux changes an electromotive force is induced (Faraday law).	Linear velocity sensor	Linear velocity	Voltage
	Flowmeter	Flow	Voltage
	Tachometer generator	Angular speed	Voltage
	Torque sensor	Torque	Voltage
Superconducting Josephson effect When a superconductor is placed inside a magnetic field, the field is completely expelled from the interior of the superconductor.	RF SQUID	Magnetic field	Voltage
	DC SQUID	Magnetic field	Voltage
Hall effect If the sensing element, carrying current, is put in a magnetic field, a difference in electric potential among its sides is generated.	Gaussmeter	Magnetic field, displacement	Voltage
	Wattmeter	Power	Voltage
Magnetoresistive effect Resistance of the sensing element depends on the strain.	Magnetoresistor	Magnetic field, linear and angular displacement, proximity, position	Resistance
Piezoelectric effect Subjecting the sensing element to stress there is a generation of electric charge.	Vibration cables	Vibration	Voltage or charge
	Active and passive force sensor	Force	Voltage or charge
	Piezoelectric microphone	Ultrasonic waves	Voltage or charge
	Piezoelectric temperature sensor	Temperature	Voltage or charge
Pyroelectric effect The sensing element generates an electric charge in response to heat flow.	Heat flowmeter	Change in the temperature	Voltage
	Pyroelectric sensor		
Thermoelectric effect When there is a difference in temperature between two junctions of different metals, a difference of electric potential is generated.	Thermocouples, thermopiles, infrared pyrometer	Difference of temperature	Voltage
Ionization effect The sensing element when exposed to the measurand becomes ionized.	Electrolytic sensor	Electrical conductivity, pH	Current
	Vacuum gauges	Pressure	Current
	Chemical ionizer	Atomic radiation	Current
Photoresistive The electric resistance of the sensing element is caused by the incidence of optical radiation.	Photoresistor, photodiode, phototransistor, photofet	Light, position, motion, sound flow, force	Resistance
Photovoltaic effect When the sensing element is subject to a radiation it generates an electric potential	Flame photometer	Light intensity	Voltage
	Light detector	Light, position, motion, sound flow, force	
	Pyrometers	Temperature	
Acoustooptic effect The interaction of an optical wave with an acoustic wave produces a new optical wave	Acoustic optic deflection, Bragg cell	Physical vibration	Phase modulated voltage
Doppler effect The apparent frequency of a wave train changes depending on the relative motion between the source of the train and the observer.	Remote sensor of linear velocity, Doppler radar, laser Doppler velocimeter	Relative velocity	Frequency
Thermal radiation An object emits thermal radiation, whose intensity is related to its temperature	Pyrometer	Temperature	Voltage

Plastics

The most widely used polymers in sensor-related applications are thermoplastic polymers, such as polyester, phenolic, alkyd, allyl, and epoxy (5). These are characterized by high flexibility and stability to mechanical stress and can be formed into any desirable shape. Because plastics are excellent electrical isolators, different methods are used to provide them with electrical conductive properties required for shielding: lamination with metal foil, painting with conductive paint, metallization, mixing plastics with conductive additives, and building composite plastic parts incorporating metal mesh. For example, piezoelectric plastics are realized by poling them either in high voltage or by corona discharge and depositing metal electrodes on both sides of the plastic film.

Furthermore, polymers are used together with glass or alone to produce optical fiber. The plastic fibers (plastic-plastic or plastic-glass) are more commonly used than glass because it is possible to realize fibers with any refractive index.

Semiconductors

There are relatively few types of intrinsic semiconductors. Those made of germanium and gallium arsenide are the most widely used (8). Adding impurities to these materials, forms extrinsic semiconductors, also called doped semiconductors, which are characterized by an increased number of carriers (9). In sensor designs the semiconducting materials (intrinsic and extrinsic) are used both as active materials and as passive materials. In some cases, the sensing element is constituted by the semiconductor itself (active material). The semiconducting materials and, in particular, silicon actually exhibit a great number of physical effects which are quite useful for sensor application, for example, photovoltaic, photoelectric, photoconductive, piezoresistive, magnetoresistant, and ion sensitive. On the other hand, when a semiconducting material does not display the proper effect, it is possible to deposit layers of materials, that have the desired sensitivity on top of the silicon substrate, which provide either a mechanical structure or electrical connection to a sensing device. Silicon, in fact, exhibits very useful mechanical proprieties and is a relatively inert element. Most acids, except hydrofluoric acid, do not affect it, but it is affected by halogens and dilute alkali. The large use of semiconducting materials is also due to the fact that these materials are inexpensive and now are produced and processed in controlled conditions to unparalleled standards of purity and perfection.

SENSOR CHARACTERISTICS

Because sensors provide an interface between the measurement system and the outside world, the choice of a sensor involves evaluating its input-output relationship and also all links with external systems (6). In Fig. 1 the physical systems directly connected to the sensing device are reported, and the information path and interaction between the sensor and the other systems are indicated. Naturally, upstream there is the system to be measured and down stream is the user system. By user system we mean the system to which the sensor provides information. It may of course perform other elaborations on the sensor output. In the environment system all interfering sources are summarized, and an auxiliary

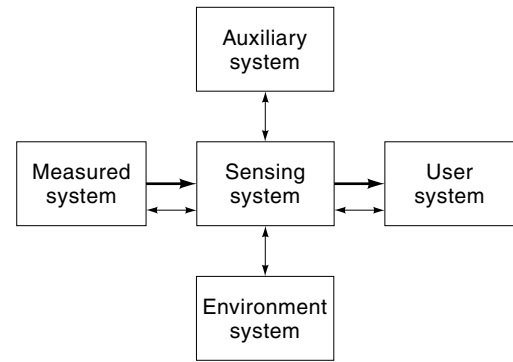


Figure 1. A sensing device and its interactions with the outside world.

system is required to operate the sensor (e.g., for the power supply) (6).

In the following, the most important characteristics of sensors are reported, organized on the basis of the systems to which they are related. The analysis and comparison of these characteristics among different sensors can help the user to choose the most suitable sensor for each particular application. Naturally, depending on the application requirements, some characteristics may be more important than others.

Measured System

Input Range, Overload, Overrange. The sensor's nominal range indicates the lower limit and the upper limit values of the measurand between which the sensor is designed to operate. The span refers to the modulus of the difference between the two limits of the nominal range. In some fields, this is called input full scale (FS). (1,2). Often the overload values or the overrange value are also provided. The overload values are the extremes of the safety field of the measurand, whereas the overrange value is the maximum value safety field of the measurand, implying that the minimum coincides with the lower limit of the measured field.

Loading Effects. The introduction of any measuring device into a measured system always results in modifying the characteristics of the measurand, thereby changing its value from its undisturbed state, and thus making a perfect measure theoretically impossible (10). For example, introducing a temperature sensor into a vessel of liquid may change its temperature (7). Consequently, the sensor manufacturer has to indicate the sensor loading effects. In particular, the sensor input impedance indicates the electric loading. In general, to limit the electric loading effects, the input impedance of the sensor must be high compared to the output impedance of the system to which the sensor is connected (10).

Mechanical Coupling. From this viewpoint, sensors can be divided into two very broad categories: contacting sensors and noncontacting sensors (2). These categories indicate whether or not the sensor must be in direct contact with the measurand. Inability to connect suitably may depend on physical requirements (the fingers cannot reach the point of interest or it is impossible to see inside the container holding the measurand) or on safety considerations (the measurand or its environment may be hazardous) (7).

Sensing System

Sensor performance characteristics are generally categorized as static and dynamic. The first describe performance at room conditions with very slow changes in the measurand and without any shock, vibration, or acceleration. The latter are related to the response of the sensor to variations of the measurand with time (11). In most measurement systems the quantity to be measured changes so slowly that it is only necessary to know the static characteristics of sensors. But the same sensor in the same operating conditions can be defined in a static or a dynamic regime according to the accuracy required (3).

Finally it is necessary to emphasize that we regard the sensor as a black box and that we are only concerned with the relationship existing between input and output quantities, even if more than one conversion step is involved from the input to the output.

Static Characteristic

Accuracy, Uncertainty, Bias, Repeatability, Reproducibility. Accuracy indicates the closeness of agreement between the result of a measurement and the value of the measurand (1).

Uncertainty characterizes the dispersion of values that could reasonably be attributed to the measurand. Thus, it expresses the fact that, for a given measurand and a given response to it, there is not one value but an infinite number of values dispersed around the result that are consistent. Even if the uncertainties are small, there is no guarantee that accuracy is high. Thus the uncertainty does not necessarily indicate the likelihood that the sensor response is close to the value of the measurand (12).

The bias of a sensor is the constant error that exists for the full range of its measurements (6).

Repeatability indicates the closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement. These conditions are called repeatability conditions and include the same measurement procedure; the same observer; the same location; and repetition over a short period of time. Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the results (1).

Reproducibility is the sensor's ability to indicate identical values of the measurand each time a measurement is made, assuming that all environmental conditions are the same for each measurement. It is defined as the closeness of agreement between the results of measurements of the same measurand carried out under changed conditions of measurement. Changes in conditions may include the principle of measurement; the method of measurement; the observer; and the same location, time, and condition of use. Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results, and it is necessary to specify the conditions changed (1).

Response Characteristic, Static Calibration. The response characteristic for a sensing device is the relationship between the input quantity (also called stimulus) and the corresponding output value (also called response) for defined conditions. This relationship may be expressed in the form of a mathematical equation, a numerical table, or a graph. An ideal or theoretical response characteristic exists for every sensor, and if the sensor follows this ideal behavior, its output always represents the value of the stimulus. But because of variations of

materials, design errors, manufacturing tolerances, workmanship and other limitations, the real response characteristic rarely coincides with the ideal. Consequently, it is necessary to establish a relationship between the sensor response and the value, with its uncertainty, to be assigned to the measurand. A calibration process helps determine this relationship. This is a test during which known values of the measurand, also called calibration points, are applied to the sensor, and the corresponding output values are recorded. The input is changed very slowly over the entire range, first increasing and then decreasing the measurand and keeping all other influential quantities constant. A single performance of this test is called a calibration cycle. A complete calibration process usually comprises two or more calibration cycles. The choice of the number and location of the calibration points and of the number of calibration cycles is very important because they may affect the achievable accuracy for a given total number of measurements. The sensor responses are suitably fitted against the input values to form the calibration curve (or calibration diagram). On the basis of all of the gathered calibration points an uncertainty band, which can vary along with any variation of the measurand, can also be added to the calibration curve. The value of measurand (y_0) corresponding to a sensor response (x_0) is obtained as the ordinate of the point of intersection between the calibration curve and the straight line parallel to the ordinate axis passing through x_0 . The uncertainty given to the input (Δy_0) is determined by the width of the segment obtained by the intersection of the same line with the uncertainty band (see Fig. 2) (1–4,9,11).

In the calibration process the value of the measurand can be determined by using two different methods: direct comparison and indirect comparison (see Fig. 3). In the first case, a standard generator furnishes a known stimulus to the sensor being calibrated. The values of these stimuli should be at least ten times more accurate than the sensor. In the indirect comparison, the calibration test consists of comparing the calibrating sensor outputs to the outputs of a standard sensor. In this case an approximate level of stimulus source is required, but the time stability and the spatial uniformity of the source must be guaranteed (7).

Sensitivity, Offset, Resolution, Dead Band, Threshold, Saturation

Sensor sensitivity refers to the sensor's ability to generate an output response to a given change in the measurand (1). It is expressed as the change in the response divided by the corresponding change in the stimulus ($\Delta y/\Delta x$). If the response

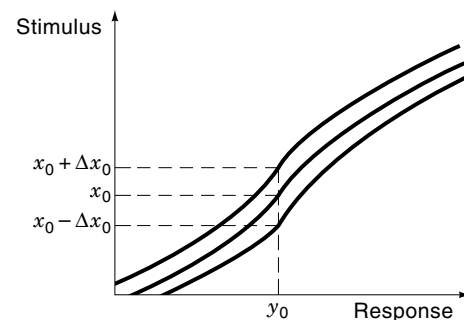


Figure 2. An example of a calibration diagram.

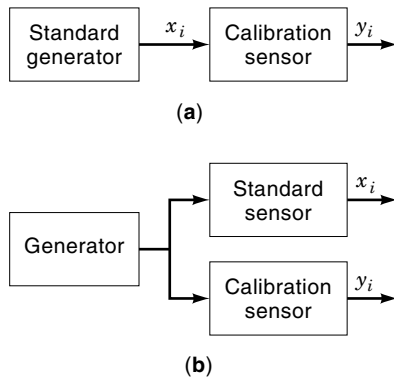


Figure 3. Techniques for calibration: (a) direct comparison; (b) indirect comparison.

characteristic of the sensor is not a straight line, the sensitivity varies with the value of the stimulus (1,2,10).

Offset is the deviation of the output signal from zero when the input is zero (6).

Resolution is defined as the smallest difference between sensor readings that can be clearly distinguished. In other words it is the minimal change of the input necessary to produce a detectable change at the output. When the output signal is digital, the resolution represents the smallest change in analog input that generates a change of one bit. The resolution is sometimes expressed as a fraction of the maximum input value (1,3,6).

The dead band describes the insensitivity of a sensor in a specific range of the input signal. It is defined as the maximum interval through which a stimulus changes in both directions without producing a change in response. It may depend on the rate of the change. The dead band is sometimes deliberately widened to prevent any change in response for small changes in the stimulus (1,9).

The sensor threshold is the largest change in a null-stimulus that produces no detectable change in the response of the sensor. The threshold may depend, for example, on noise, friction, and also on the value of the stimulus (1).

Almost any sensor has its operating limits. Further increase of the stimulus does not produce an increase (or decrease) of the output, and the sensor goes into a saturation zone.

Linearity. Sometimes, rather than expressing the calibration curve as a numerical table or a graph, its linear approximation is furnished. There are different methods for constructing the line (2,3,11) (see Fig. 4):

Least-Square Linearity. The straight line is defined by the least square criterion. Consequently the sum of the squares of the residuals is minimized. This is the method that usually gives the lowest value of nonlinearity error.

Zero-Based Linearity. The straight line is also defined by the least square but with the additional restriction that it passes through zero.

Theoretical-Slope Linearity. The straight line is referenced to the theoretical line defined in the sensor design phase.

Terminal-Based Linearity. The straight line is defined as the line that passes through the output corresponding to the lower input and the theoretical output when the higher input is applied.

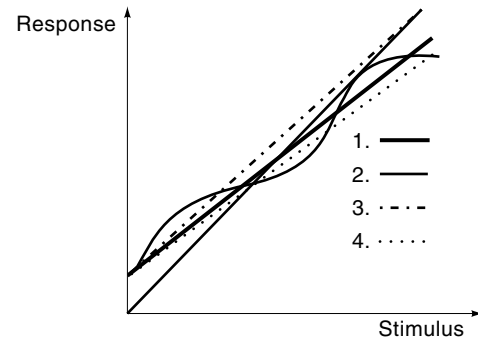


Figure 4. Different strain lines: 1. least-square; 2. theoretical slope; 3. terminal-based; 4. end points.

End-Point Linearity. The straight line is defined by the real output when the upper and lower input range are applied.

The nonlinearity error is defined as the maximum deviation of any calibration point from the corresponding point on a specified straight line. Normally it is defined as a percentage of the output range.

Another term sometimes used is conformance (or conformity), which indicates the closeness of a calibration curve to a specific curve (normally the theoretical curve) for a nonlinear sensor (4,5).

Hysteresis. Hysteresis is the difference between two output values that correspond to the same stimulus depending on the direction (increasing or decreasing) and whether that value is reached. There is a chance that the output corresponding to a given input depends on whether the previous input was higher or lower than the present one. That is similar to the magnetization of ferromagnetic materials (3). Typical causes for hysteresis are friction and structural change in the material of the sensing element. Hysteretic error is defined as the difference between the measured values obtained when the measured quantity is increasing and when decreasing to that value. Hysteresis is usually quantified in terms of the maximum hysteretic error as a percentage of the full-scale output (6,9).

Stability, Creep, Drift, Zero-Shift, Sensitivity Shift. Stability indicates the sensor's ability to maintain constant metrological characteristics in time (1). Short-term stability is manifested as changes in the sensor performance within minutes, hours, or even days. The long-term stability may be related to aging of the sensor materials, which causes an irreversible change in its material properties. Aging depends greatly on environmental storage and operating conditions and how well the sensor components are isolated from the environment. Stability may be quantified in several ways, for example, in terms of the time over which a metrological characteristic changes by a stated amount or in terms of the change in a characteristic over a stated time (5).

A sensing device shows drift if there is a gradual change in output over a period of time which is unrelated to any change in input (6).

Zero-shift is a change in the zero measurand over a specific time interval at room conditions. The zero-measurand output is the output of the sensor under room conditions with nominal excitation and zero measurand applied (1).

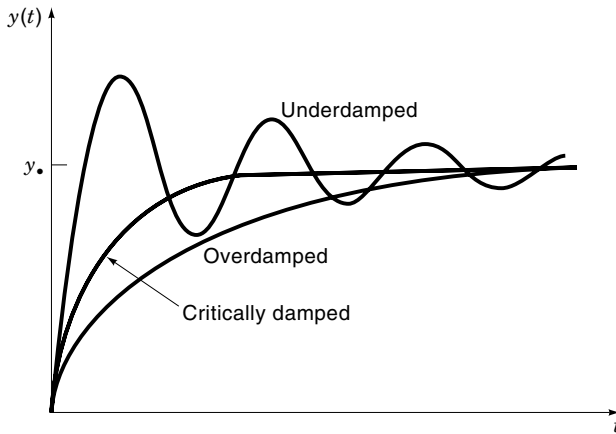


Figure 5. Different damping characteristics.

Sensitivity shift is a change in the sensor's sensitivity over a specific period at room conditions (1).

Dynamic Characteristics. When an input stimulus varies suddenly, the sensor response does not follow this variation with perfect fidelity. The sensor does not respond instantly to the stimulus change. The sensor's dynamic characteristics can be stated in term of speed of response, velocity limit, slew rate, and recovery time (2,5,11). Speed of response indicates how fast the sensor reacts to changes in the output variable. Slew rate is the maximum rate of change with time of the input for which the output can keep up with the change. Recovery time is the time interval necessary after a specified event for the sensor to resume functioning as specified.

For linear sensors other parameters are used to describe the sensor behavior in dynamic conditions. They refer to the time and frequency domains.

Time Domain. The most commonly used dynamic characteristics in the time domain are the step response and the natural response. In response to a step in a measurand, the sensor may have different damping characteristics (see Fig. 5):

- An underdamped sensor oscillates around its final value (y_∞) before coming to rest at that value.
- An overdamped system comes to a final value without overshoot.
- A critically damped system is at the point of change between underdamped and overdamped conditions.

The ratio of the actual damping to the degree of damping required for critical damping is the damping factor. A damping ratio of 1.0 indicates critical damping. Damping ratios larger than 1.0 signify overdamping. Finally, underdamping is indicated by a damping ratio of less than 1.0 (11).

The parameters of an overdamped or critically damped sensor are defined as follows:

Dead time is the time interval between the instant of step application (t_0) and the instant (t_1) in which the response exceeds a threshold value (y_1), usually defined as a percentage of (y_∞).

Response time describes the time that has to elapse before the sensor fully responds to the change. It is defined as the length of time required for the output to rise to a specified percentage of (y_∞) (the percentage is typically 95% or 98%).

A special term has been assigned to 63.2% response time: the time constant (τ).

Having fixed two percentages of the final output value, the rise time is the interval necessary for the output to rise from the smaller percentage to the larger. Unless otherwise specified, the percentages should be assumed to be 10% and 90% of the final value, respectively.

For underdamped systems other parameters are also used:

The overshoot is the difference between the maximum output value ($y_{\max}$) and y_∞ .

After having fixed a value band at around y_∞ (normally it is equal to 5% of y_∞), the settling time is the time interval between t_0 and the instant in which the output remains limited to the previously mentioned band;

The ringing frequency is the damping oscillation frequency.

With reference to natural response, which is the evolution of the sensor output starting from a not null initial value and without measurand, two parameters are defined, the natural frequency and the natural undamped frequency. The natural frequency is the fundamental sinusoidal component of the natural response (14).

Frequency Domain. In the frequency domain a very important dynamic characteristic is the frequency response. It is the change of the amplitude and phase of the output as a function of the frequency of a unit amplitude sinusoidal input. These two curves are, respectively, the module and the phase of the so-called Fourier transfer function [$G(\omega)$] (4). The frequency response is displayed graphically (see Fig. 6) as the plot as a function of the frequency of the sensor amplitude output (amplitude diagram), normally using a logarithmic scale, and of the phase displacement between the input sinusoid and the output sinusoid (phase diagram). Many synthetic parameters describe the frequency response of a system. The bandwidth, also called frequency range, indicates the range

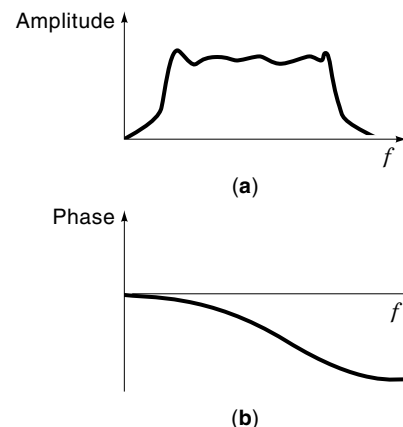


Figure 6. An example of the (a) amplitude and (b) phase frequency response of a sensor.

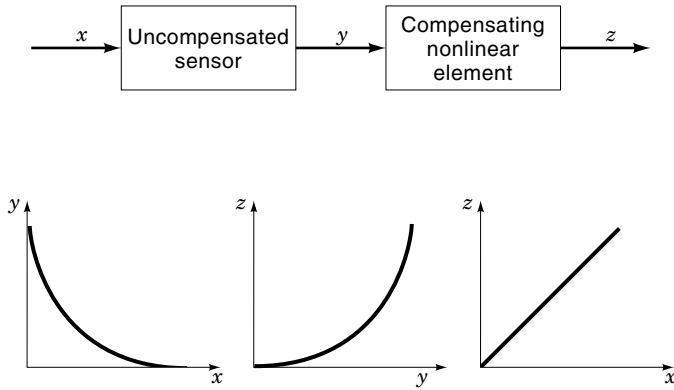


Figure 7. A static compensation technique.

of frequencies over which the sensor can be used. It is defined as the range of frequencies for which the transfer function is within a fixed band (normally 3 dB, 70.7%) of its peak value and is defined by the lower and upper cutoff frequencies. The resonant frequency is the frequency corresponding to which the module of the frequency response has a maximum value (6).

Static and Dynamic Compensation Technique. Generally, the sensor calibration curve is nonlinear, but it may be corrected by a suitable static compensation technique. One of the most common methods consists of introducing a compensating nonlinear element in cascade connection to the sensing device. This method is illustrated in Fig. 7, which also shows the response characteristic of the single blocks and the whole compensated system (13). As a consequence the relationships between response y and stimulus x can be expressed as: $y = ax + b$. Sometimes, to obtain a proportional relationship $y = kx$, another compensation stage is added, which allows compensating the sensor's sensitivity (a) and the shift from zero (b).

When the sensing fails to meet the dynamic specifications, that is, the dead time is too long or equally the frequency range is too small, it is necessary compensate for it (15). Having identified the dominant element in the system, the most obvious method of improving the dynamic response is that of the inherent design, that is, the design parameters are varied to improve the dynamic response of such an element. Two other methods are normally used: open-loop and closed-loop dynamic compensation. In the open-loop technique a linear element, with a transfer function in the Laplace domain $H_c(s)$, is introduced into the sensor system, such that the overall Laplace transfer function $H(s) = G(s)H_c(s)$ fulfils the required condition [Fig. 8(a)]. In the closed-loop, the compensation is obtained by using a high negative feedback [Fig. 8(b)]. The output of the sensing device is amplified by a high-gain amplifier [$H_g(s)$]. The amplifier output is fed back to an element [$H_c(s)$] which provides a balancing force to oppose the input force. In this case the compensated transfer function is given by

$$H(s) = \frac{G(s)}{1 + H_c(s) \cdot G(s)}$$

If the amplifier gain is rendered large such that the condition $G(s)H_c(s)H_g(s) \gg 1$ is satisfied, then $H(s) \sim 1/H_c(s)$. This means that, providing the previous condition is respected, changes in

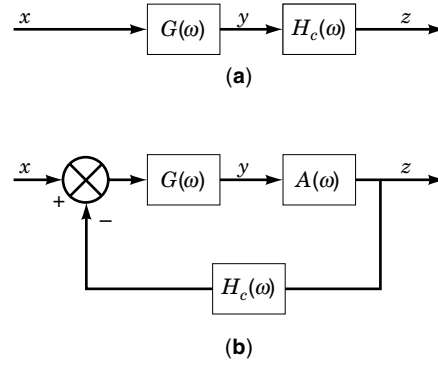


Figure 8. Sensor dynamic compensation, using (a) an open-loop technique; (b) a closed-loop with high negative feedback.

$G(s)$ due to nonlinear effects have a negligible effect on the compensated output (13) (Fig. 9).

Today, compensation is often realized by numerical microprocessors. In static compensation, using a look-up table based on the results of the calibration or using an interpolation technique, the microprocessor gives the corresponding input value for each measured value. Consequently, there is more interest in repeatability than in linearity. As for dynamic compensation,

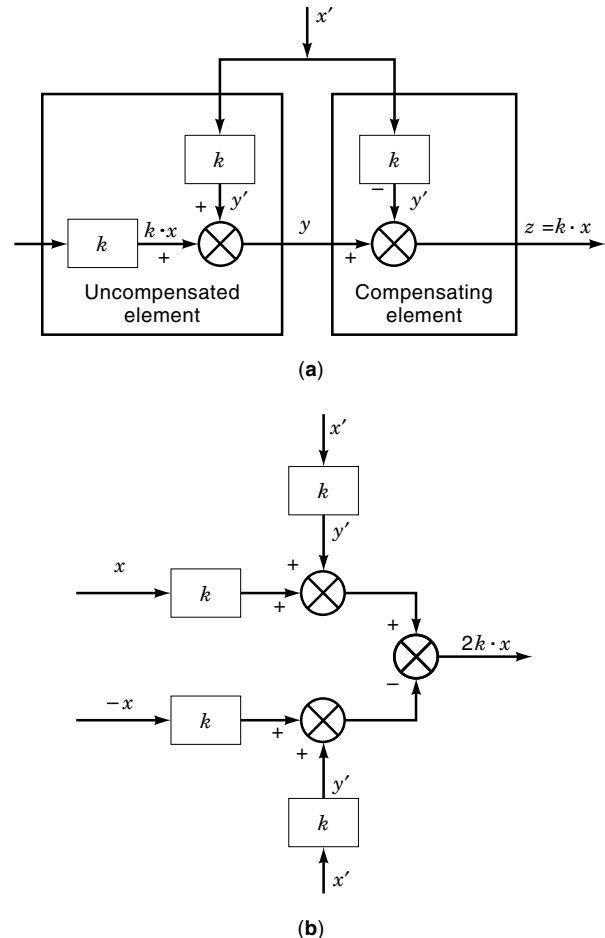


Figure 9. Compensation for interfering inputs (x') using (a) opposing environmental inputs or (b) a differential system for a linear sensor.

the transfer function of the sensor is stored in the processor memory, and using a deconvolution technique and the previous output values, the processor gives the corresponding input value for each output value. The sensor transfer function is obtained by an identification procedure realized in a preliminary dynamic calibration. Naturally, it is necessary that the dynamic behavior of the sensor not vary in time (10).

Reliability Characteristics. Reliability characteristics are those characteristics relating to the useful life of a sensor. These characteristics can be specified in different ways depending on the sensor type and, for a sensor, depending on what aspect has to be highlighted. In particular, the following parameters can be of interest:

Operating Life. It is the minimum length of time over which the sensor will operate, either continuously or over a number of on-off cycles, whose duration is specified, without changing its performance beyond specified tolerances.

Cycling Life. It is the minimum number of measurand full-range excursions (or specified partial range excursions) over which a sensor will operate without changing its performance beyond specified tolerances.

In some cases it may also be necessary to specify or be concerned about a sensor's storage life, the length of time over which it can be exposed to specified storage conditions without changing its performance beyond specified tolerances.

The evaluation of the operating life, also of great interest in the field of industrial applications, is hardly ever performed, because specification for operating life tends to be in terms of years, typically between three and ten years. Sometimes, field experience can be substituted for performing an operating life test, when it can be documented that at least one sensor of the same design, taken randomly from a production lot, has operated maintenance free and within specified tolerances but in its end-use application for a certain number of years. In order to reduce the testing time and consequently the cost, suitable accelerated life tests (tests performed under environmental conditions more severe than normal) can be set up for some kind of sensors. Of course, the results of these kind of tests have to be considered as indicative and have to be verified by a life test under nominal environmental conditions.

Cycling life tests are usually performed as part of a qualification test when a sensor specification call for full-range or partial range cycling life. Equipment has been designed for rapid and automatic cycling of many categories of sensors.

User System

Output Signal. There are three major categories of data signal forms: analog, carrier, and digital or pulse. In the analog sensor the output signal level is directly correlated to the value of the measurand. In the so-called carrier form the measured value modulates the characteristics of a carrier sinusoid. The amplitude (amplitude modulation AM), the frequency (frequency modulation FM) or the phase relationship with a reference signal (phase modulation PM) of the sensor output signal depends on the measurand. The most widely used modulation is frequency modulation, because the analog signal frequency has great advantages for signal transmission, for example, fail-safeness and easy galvanic separation. Furthermore it is

very easy to obtain a digital output from FM (2). A digital signal consists of a series of pulses containing the encoded information corresponding to the input data. The information may be encoded into the amplitude, width, position or frequency of the pulses. Only a few sensors with direct digital output are available (e.g., incremental or coded displacement and angle sensors) because it is mostly necessary to convert the output signals into digital form to interface them with digital system. Many sensing devices also incorporate an analog-to-digital converter.

Output Range. To indicate the output range, the most commonly used parameter is the full-scale output (FSO). This is the algebraic difference between the electric output signal measured, respectively, with the maximum and the minimum input in the input range. The output overload values are the sensor outputs that correspond to input overload values (9).

Output Noise. This is a random fluctuation of the output signal not correlated to any information. Noise at the sensor output is caused by internal and external sources, but when a sensor has to be characterized, only internal noise is considered. Noise is usually furnished in terms of rms value, and its considered bandwidth is also indicated. Sometimes the signal-to-noise ratio is indicated. This is the ratio of the signal level to the internally generated noise level usually expressed in decibels (4,14).

Output Impedance. The sensor output impedance is the impedance between the output terminals of the instrument (6). It may be connected in parallel or in series with the input impedance of the interface circuit. Output impedance generally should be represented in a complex form and may include active and reactive components. When the reactive part is negligible, it is called output resistance. It is important to know its value to better interface a sensor with the user system (5).

Output Power. This is the maximum value of the power that the sensor can supply to the user system without decreasing its own performance. Sometimes, if the sensor output is in voltage, the maximum deliverable current is indicated, whereas, for a current output signal, the maximum deliverable voltage or the maximum load impedance values are specified.

Grounding. Sometimes the sensing device is grounded to establish a common mode among different parts of the system with the requirement that no potential variation may occur along this common node with respect to any point inside the node (4). It is important to know the possible grounding circuit to avoid grounding some user system floating points unintentionally.

Auxiliary System

In considering the need for a power supply, sensors can be classified as modulating (passive) or self-generating (active). Modulating sensors require excitative power from an external source. To produce the output signal, the excitative signal is modified by the sensing element as a function of the measurand. In self-generating sensors, instead, output power comes from the input, that is they produce an electric output signal from their input quantity. For example, a thermocouple is a

self-generating thermal sensor. It produces an electromagnetic force from the difference in junction temperatures, whereas the resistance of a thermistor changes with temperature. To measure its value, it is necessary to pass a current through the thermistor. By using the power supply voltage, it is possible to control overall sensitivity. In fact, normally self-generating sensors produce very low output power, whereas modulating sensors produce much higher output energies. The presence of an auxiliary power source increases the danger of explosion in explosive atmospheres. (2,3,5,9).

Environment System

The static and dynamic performance characteristics of sensors are those which the sensor exhibits at room conditions and in the absence of any external conditions such as temperature, humidity, vibrations, that may affect the sensor's performance (11). The definition of influential quantity includes values associated with measurement standards, reference materials, and reference data, upon which the result of a measurement may depend, and phenomena, such as room temperature, barometric pressure, and humidity (11).

When a sensor is reasonably expected to operate under conditions, called operating environmental conditions, other than those under which it was calibrated, the environmental effects must be known, and the resulting deviations from static performance (environmental errors) must be determined by tests. Such environmental tests may have to be performed on each sensor used. Usually they are performed on a sampling basis (test one of every N sensors of each model and range) but sometimes only on a qualification basis (a test on a representative sensor). In linear sensors, the effect of a particular parameter can be studied in terms of its effect on static and dynamic sensor behavior, keeping the remainder of the parameters constant.

Other operating environmental effects on the behavior of a sensor during its normal operation which should be known and included in specifications include humidity or immersion in liquid; corrosive effects of high salt concentration in the ambient atmosphere; various effects of measured fluids on the sensing elements; the influence of ambient electromagnetic fields on sensor elements and integral circuitry; and the effects of radiation on various internal sensor elements.

Besides operating environmental conditions, there are other conditions to which a transducer may be exposed, but the sensor is not expected to operate within specified tolerances while exposed to them. However, the sensor is expected to perform within specified tolerances after exposure to such environmental conditions, called nonoperating environmental conditions (11).

In the following, the most common environmental effects are presented together with the parameter usually used to express them.

Temperature Effects. The operating range is the range of ambient temperatures, given by their lower and upper extremes within which the sensor is intended to operate and within which all specifications related to temperature effects apply. The maximum temperature is the highest (or lowest) temperature to which a sensor can be exposed without being damaged or subsequently showing performance degradation beyond specified tolerances. A more general and inclusive way of specifying thermal effects on performance characteristics is given by the

temperature error, which is the maximum change in output when the temperature is changed from room temperature to specified temperature extremes. Analogously, a temperature error band is defined as the error band applicable over the operating temperature range. For some sensors, temperature effects are stated only in terms of thermal zero shift and thermal sensitivity shift, which cause a parallel displacement and a slope change, respectively, of the calibration curve. Knowledge of these individual errors is useful when the prevailing temperature during a measurement procedure is known and appropriate corrections to final data are to be made (11).

It is necessary to emphasize that temperature also affects dynamic characteristics. Consequently, the specifications should include corresponding thermal effects.

Pressure Effects. In some sensors calibrated at room barometric pressure and used where the ambient pressure is very low (e.g., at high altitude on aircraft) or where the pressure is very high (far underground or deeply submerged underwater), significant variation of performance may arise. Ambient-pressure error is the maximum change in output at any measurand value within the sensor's range, when the ambient pressure is changed between specified values, usually between room pressure and a lower or higher ambient pressure. The pressure error can be stated in terms of an ambient-pressure error band. Sometimes the pressure error is referred to as altitude error, and ranges of pressures are stated in terms of altitude above sea level (11).

Acceleration Effects. Quasi-steady-state acceleration of internal elements of a sensor may act directly on a mechanical sensing element or its linkage and cause errors in its output. When a sensor is to be used in an application where it experiences acceleration, the possibility of acceleration errors must be considered and tolerances must be established for such error. The acceleration error is defined as the maximum difference between output readings taken without and with the application of specified constant acceleration along specified axes. The acceleration effects, in fact, are typically more evident when the acceleration is applied along one axis of the sensor than when it is applied along other axes. This error can also be reported in terms of acceleration error band (11).

Vibration Effects. Vibration acceleration affects a sensor in the same manner as steady-state acceleration. Stronger effects, however, are connected with the frequencies of vibration. Amplified vibration (resonances) of internal elements can occur at one or more frequencies, and different resonances may be observed for different measurand values, particularly when the sensor incorporates a mechanical sensing element. Vibration error is defined, then, as the maximum change in output when a vibration level of specified amplitudes and ranges of frequency is applied to the sensor. Consequently, it may be necessary to predict the measurand value most likely to be observed by the sensor while it is exposed to the most severe vibration environment, and then to specify and verify vibration errors at that value (11).

Mounting Effects. A sensor's performance may be changed during its installation, for example, when the mounting surface of the sensor is not evenly machined so that the case becomes deformed when all the mounting hardware is tightened

or when the torque applied to the coupling nut on a pressure fitting causes sensing element deformations. Mounting error is the error resulting from mechanical deformation of the sensor caused by mounting the sensor and making all electrical and measurand connections. Mounting error, obviously, is not included in specifications. However, it may be necessary to verify its absence (11).

Environmental Effects Reduction Technique. The most obvious method of reducing the effects of environmental parameters is that of isolation, which allows reducing undesirable electric, magnetic, electromagnetic, and mechanical coupling among various parts of the system and between the system and the outside (4).

The most successful techniques are the opposition method and the differential approach. The first is based on opposing environmental inputs. Suppose that an element is affected by an environmental parameter. Then a second element, subject to the same environmental input [see Fig. 9(a)], is introduced into the system so that the two effects cancel each other. An example of a differential system is shown in Fig. 9(b). As can be seen, two identical sensors are placed to sense the measurand and its opposite. Because both are affected by external quantities which are presented to both with the same sign, by calculating the difference between the two outputs, such effects can be eliminated. Naturally, environmental effects may be also compensated for by using the previously presented high-gain negative feedback and using a feedback system which is unsusceptible to environmental input (13).

Now the technique of computer estimation of measured value can also be used. A good model of the element in the system is required for this method. In fact, knowing the relationship between the sensor output, the environmental parameters, and the values of these parameters, the environmental effects can be numerically eliminated by the sensor output value.

BIBLIOGRAPHY

1. ISO *Guide to the Expression of Uncertainty in Measurement*, Geneva: ISO, 1993.
2. J. R. Carstens, *Electrical Sensors and Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
3. R. Pallàs-Areny and J. G. Webster, *Sensors and Signal Conditioning*, New York: Wiley, 1991.
4. Gopel and J. N. Zemel, *Sensors: A Comprehensive Survey*, New York: VCH, 1989, vol. 1.
5. J. Fraden, *Handbook of Modern Sensors Physics, Designs, and Applications*, 2nd ed., Woodbury, NY: AIP Press, 1997.
6. Bolton, *Electrical and Electronic Measurement and Testing*, Singapore: Longman Scientific & Technical, 1990.
7. C. F. Coombs, Jr., *Electronic Instrument Handbook*. New York: McGraw-Hill, 1995.
8. J. Millman and A. Grabel, *Microelectronics*, 2nd ed., Singapore: McGraw-Hill, 1987.
9. J. W. Gardner, *Microsensors Principles and Applications*, Chichester: Wiley, 1995.
10. E. O. Doebelin, *Measurement Systems Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
11. H. N. Norton, *Handbook of Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
12. ISO *International Vocabulary of Basic and General Terms in Metrology*, Geneva: ISO, 1993.
13. J. P. Bentley, *Principles of Measurement Systems*, 3rd ed., New York: Wiley, 1995.
14. P. H. Mansfield, *Electrical Transducers for Industrial Measurements*, Toronto: Butterworths, 1973.
15. P. H. Sydenham, *Handbook of Measurement Science*, vol. I, II, New York: Wiley, 1982–1983.

MASSIMO D'APUZZO
University of Napoli
CONSOLATINA LIGUORI
University of Cassino

MAGNETIC SENSORS

Magnetic sensors find many applications in everyday life and in industry. They provide convenient, noncontact, simple, rugged, and reliable operations compared to many other sensors. The technology to produce magnetic sensors involves many aspects of different disciplines such as physics, metallurgy, chemistry, and electronics.

Generally, magnetic sensors are based on sensing the properties of magnetic materials, which can be done in many ways. For example, magnetization, which is the magnetic moment per volume of materials, is used in many measurement systems by sensing force, induction, field methods, and superconductivity. However, the majority of industrial sensors make use of the relationship between magnetic and electric phenomenon. A typical application of the phenomenon is the computer memory requiring the reading of the contents of a disc without making any contact between the sensor and the device. In other applications, the position of objects sensitive to magnetic fields (e.g., the metals in the ground) can be sensed magnetically. Magnetic sensors find most sensitive applications in medicine to diagnose human illnesses, as in the case of superconducting quantum interference devices (SQUID) and nuclear resonance magnetic (NMR) imaging.

The magnetic elements in sensors are used in a wide range of forms: toroids, rods, films, substrates, and coatings. Some elements are essentially free standing, whereas others are an integral part of more complex devices. In order to obtain maximum material response in magnetic sensors, the relative orientation and coupling between input measurand and magnetic properties are very important, and they are optimized at the design stages.

Many different types of magnetic sensors are available. These sensors can broadly be classified as primary or secondary. In primary sensors, also known as the magnetometers, the parameter to be measured is the external magnetic field. The primary sensors are used in biological applications and geophysical and extraterrestrial measurements. In secondary sensors, the external parameter is made from other physical variables such as force and displacement. In this article, both the primary and secondary sensors will be discussed. These sensors include inductive, eddy current, transformative, magnetoresistive, Hall-effect, metal-oxide-semiconductor (MOS) magnetic field, and magneto-optical sensors; magnetotransistor and magnetodiode sensors, magnetometers; superconductors;

Table 1. List of Manufacturers

Adsen Tech, Inc. 18310 Bedford Circle La Puente, CA 91744 Fax: 818-854-2776	Motion Sensors, Inc. 786 Pitts Chapel Road Elizabeth City, NC 27909 Tel: 919-331-2080 Fax: 919-331-1666
Analog Devices, Inc. 1 Technology Way P.O. Box 9106 Norwood, MA 02062-9102 Tel: 800-262-5663 Fax: 781-326-8703	Rechner Electronics Industries, Inc. 8651 Buffalo Avenue Niagara Falls, NY 14304 Tel: 800-544-4106 Fax: 716-283-2127
Dynalco Controls 3690 N. W. 53rd Street Ft. Lauderdale, FL 33309 Tel: 305-739-4300 & 800-368-6666 Fax: 305-484-3376	Reed Switch Developments Company, Inc. P. O. Drawer 085297 Racine, WI 53408 Tel: 414-637-8848 Fax: 414-637-8861
Electro Corporation 1845 57th Street Sarasota, FL 34243 Tel: 813-355-8411 & 800-446-5762 Fax: 813-355-3120	Smith Research and Technology, Inc. 205 Sutton Lane, Dept. TR-95 Colorado Springs, CO 80907 Tel: 719-634-2259 Fax: 719-634-2601
Honeywell Dept. 722 11 West Spring Street Freeport, IL 61032 Tel: 800-537-6945 Fax: 815-235-5988	Smith Systems, Inc. 6 Mill Creek Drive Box 667 Brevard, NC 28712 Tel: 704-884-3490 Fax: 704-877-3100
Kaman Instrument Company 1500 Garden of the Gods Road Colorado Springs, CO 80907 Tel: 719-599-1132 & 800-552-6267 Fax: 719-599-1823	Standex Electronics 4538 Camberwell Road Dept. 301L Cincinnati, OH 45209 Tel: 513-871-3777 Fax: 513-871-3779
Kavlico Corporation 14501 Los Angeles Avenue Moorpark, CA 93021 Tel: 805-523-2000 Fax: 805-523-7125	Turck, Inc. 3000 Campus Drive Minneapolis, MN 55441 Tel: 612-553-7300 & 800-544-7769 Fax: 612-553-0708
Lucas 1000 Lucas Way Hampton, VA 23666 Tel: 800-745-8008 Fax: 800-745-8004	Xolox Sensor Products 6932 Gettysburg Pike Ft. Wayne, IN 46804 Tel: 800-348-0744 Fax: 219-432-0828

semiconductors; and magnetic thin films. They are offered by many manufacturers as listed in Table 1.

can be expressed as

$$e = -d\Phi/dt \text{ (V)} \quad (1)$$

INDUCTIVE SENSORS

Inductive sensors make use of the principles of magnetic circuits. They can be classified as passive sensors and self-generating sensors. The passive sensors require an external power source; hence, the action of the sensor is restricted to the modulation of the excitation signal in relation to an external stimuli. On the other hand, the self-generating types generate signals by utilizing the electrical generator principle based on Faraday's Law of Induction. That is, when there is a relative motion between a conductor and a magnetic field, a voltage is induced in the conductor. Or a varying magnetic field linking a stationary conductor produces voltage in the conductor, which

where Φ is the magnetic flux.

In instrumentation applications, the magnetic field may be varying in time with some frequency, and the conductor may be moving at the same time. In many cases, the relative motion between field and conductor is supplied by changes in the measurand, usually by means of a mechanical motion.

In order to explain the operation of the basic principles of inductive sensors, a simple magnetic circuit is shown in Fig. 1. The magnetic circuit consists of a core, made from a ferromagnetic material, and a coil of n number of turns wound on it. The coil acts as a source of magnetomotive force (mmf), which drives the flux Φ through the magnetic circuit. If we assume that the air gap is zero, the equation for the magnetic

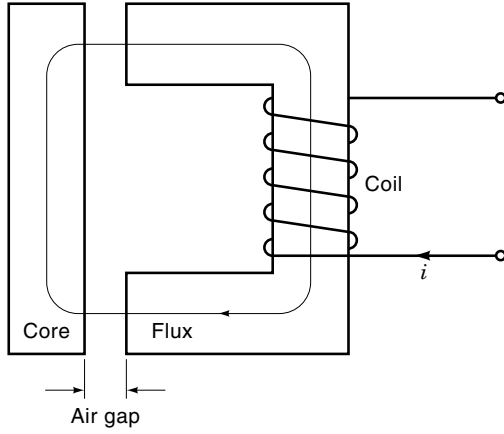


Figure 1. A basic inductive sensor consists of a magnetic circuit made up from a ferromagnetic core with a coil wound on it. The coil acts as a source of mmf, which drives the flux through the magnetic circuit and the air gap. The presence of the air gap causes a large increase in circuit reluctance and a corresponding decrease in the flux. Hence, a small variation in the air gap causes a measurable change in inductance.

circuit may be expressed as

$$\text{mmf} = \text{Flux} \times \text{Reluctance} = \Phi \times \mathcal{R} \quad (\text{A-turns}) \quad (2)$$

such that the reluctance $\mathcal{R}$ limits the flux in a magnetic circuit just as resistance limits the current in an electric circuit. By writing the magnetomotive force in terms of current, the magnetic flux may be expressed as

$$\Phi = ni/\mathcal{R} \quad (\text{Wb}) \quad (3)$$

In Fig. 1, the flux linking a single turn is expressed by Eq. (3). But the total flux linking by the entire n number of the turns of the coil is

$$\Psi = n\Phi = n^2 i/\mathcal{R} \quad (\text{Wb}) \quad (4)$$

Equation (4) leads to self-inductance L of the coil, which is described as the total flux per unit current for that particular coil. That is,

$$L = \Psi/I = n^2/\mathcal{R} \quad (\text{H}) \quad (5)$$

This indicates that the self-inductance of an inductive element can be calculated by magnetic circuit properties. Expressing $\mathcal{R}$ in terms of dimensions as

$$\mathcal{R} = l/\mu\mu_0 A \quad (\text{A-turns/Wb}) \quad (6)$$

where l is the total length of the flux path (meters), μ is the relative permeability of the magnetic circuit material, μ_0 is the permeability of free space ($= 4\pi \times 10^{-7} \text{ H/m}$), and A is the cross-sectional area of the flux path.

If the air gap is allowed to vary, the arrangement illustrated in Fig. 1 becomes a basic inductive sensor. In this case, the ferromagnetic core is separated in two parts by the air gap. The total reluctance of the circuit now is the addition of the reluctance of the core and the reluctance of the air gap. The relative

permeability of air is close to unity, and the relative permeability of the ferromagnetic material can be on the order of a few thousand, indicating that the presence of the air gap causes a large increase in circuit reluctance and a corresponding decrease in the flux. Hence, a small variation in the air gap causes a measurable change in inductance. There are many different types of inductive sensors as will be discussed next.

Linear and Rotary Variable-Reluctance Sensors

The variable-reluctance transducers are based on change in the reluctance of a magnetic flux path. These types of devices find applications particularly in acceleration measurements. However, they can be constructed to be suitable for sensing displacements as well as velocities. They are constructed in many different forms, some of which will be described in this article.

Single-Coil Linear Variable-Reluctance Sensor. A typical single-coil variable-reluctance displacement sensor is illustrated in Fig. 2. The sensor consists of three elements: a ferromagnetic core in the shape of a semicircular ring, a variable air gap, and a ferromagnetic plate. The total reluctance of the magnetic circuit is the sum of the individual reluctances:

$$\mathcal{R}_T = \mathcal{R}_C + \mathcal{R}_G + \mathcal{R}_A \quad (7)$$

where $\mathcal{R}_C$, $\mathcal{R}_G$, and $\mathcal{R}_A$ are the reluctances of the core, air gap, and armature, respectively.

Each one of these reluctances can be determined by using the properties of materials involved as in Eq. (6). In this particular case, $\mathcal{R}_T$ can be approximated as

$$\mathcal{R}_T = \frac{R}{\mu_C \mu_0 r^2} + \frac{2d}{\mu_0 \pi r^2} + \frac{R}{\mu_A \mu_0 r t} \quad (8)$$

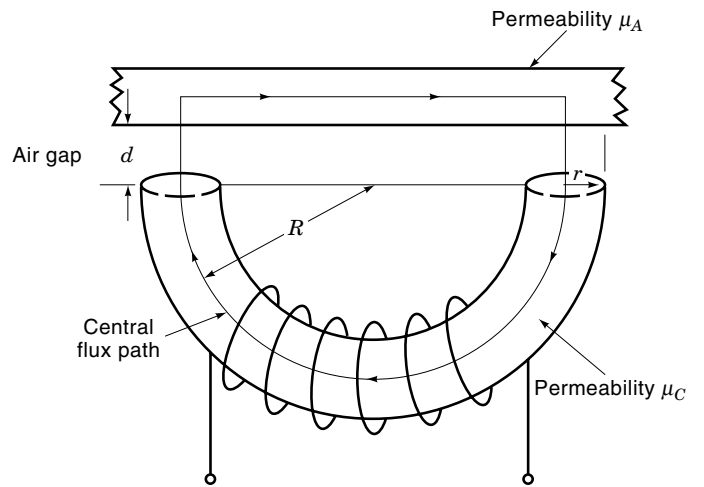


Figure 2. A typical single-coil variable-reluctance displacement sensor. The sensor consists of three elements: a ferromagnetic core in the shape of a semicircular ring, a variable air gap, and a ferromagnetic plate. The reluctance of the coil is dependent on the air gap. Air gap is the single variable, and the reluctance increases nonlinearly with the increasing gap.

In obtaining Eq. (8), the length of the flux path in the core is taken as πR , and the cross-sectional area is assumed to be uniform with a value of πr^2 . The total length of the flux path in air is $2d$, and it is assumed that there is no fringing or bending of the flux through the air gap, such that the cross-sectional area of the flux path in air will be close to that of the cross section of the core. The length of an average central flux path in the armature is $2R$. The calculation of an appropriate cross section of the armature is difficult, but it may be approximated to $2rt$, where t is the thickness of the armature.

In Eq. (8), all the parameters are fixed except the only one independent variable, the air gap. Hence, it can be simplified as

$$\mathcal{R}_T = \mathcal{R}_0 + kd \quad (9)$$

where $\mathcal{R}_0 = R/(\mu_0 r^2 [1/(\mu_C r) + 1/(\mu_A t)])$, and $k = 2/(\mu_0 \pi r^2)$.

By using Eqs. (5) and (9), the inductance can be written as

$$L = \frac{n^2}{\mathcal{R}_0 + kd} = \frac{L_0}{1 + \alpha d} \quad (10)$$

where L_0 represents the inductance at zero air gap and $\alpha = k/\mathcal{R}_0$.

The values of L_0 and α can be determined mathematically. They depend on the core geometry, permeability, and the like, as already explained. As it can be seen from Eq. (10), the relationship between L and α is nonlinear. Despite this nonlinearity, these types of single-coil sensors find applications in many areas, such as force measurements and telemetry. In force measurements, the resultant change in inductance can be made to be a measure of the magnitude of the applied force. The coil usually forms one of the components of an LC oscillator whose output frequency varies with the applied force. Hence, the coil modulates the frequency of the local oscillator.

Variable-Differential Reluctance Sensor. The problem of the nonlinearity may be overcome by modifying the single-coil system into variable-differential reluctance sensors (also known as push-pull sensors), as shown in Fig. 3. This sensor consists of an armature moving between two identical cores separated by a fixed distance of $2d$. Now, Eq. (10) can be written for both coils as

$$L_1 = \frac{L_{01}}{1 + \alpha(d - x)}, \quad L_2 = \frac{L_{02}}{1 + \alpha(d + x)} \quad (11)$$

Although the relationship between L_1 and L_2 is still nonlinear, the sensor can be incorporated into an ac bridge to give a linear output for small movements. The hysteresis error of these transducers is almost entirely limited to the mechanical components. These sensors respond to static and dynamic measurements. They have continuous resolution and high outputs, but they may give erratic performances in response to external magnetic fields. A typical sensor of this type has an input span of 1 cm, a coil inductance of 25 mH, and a coil resistance of 75 Ω . The resistance of the coil must be carefully considered when designing oscillator circuits. The maximum nonlinearity may be limited to 0.5%.

In typical commercially available variable-differential sensors, the iron core is located half way between the two E-shaped

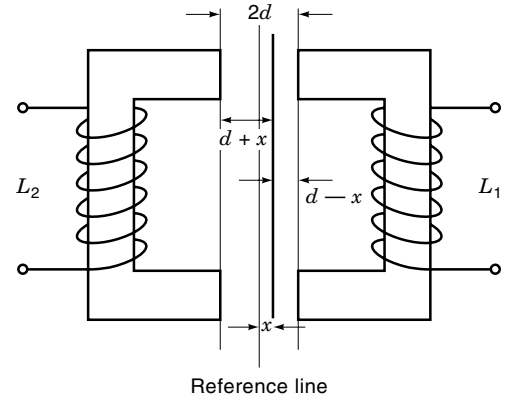


Figure 3. A variable-differential reluctance sensor consists of an armature moving between two identical cores separated by a fixed distance. The armature moves in the air gap in response to the mechanical input. This movement alters the reluctance of coils 1 and 2 thus altering their inductive properties. This arrangement overcomes the problem of nonlinearity inherent in single-coil sensors.

frames. The flux generated by primary coils depends on the reluctance of the magnetic path, the main reluctance being the air gap. Any motion of the core increases the air gap on one side and decreases it on the other side. Consequently, the reluctance changes in accordance with the principles explained previously, thus inducing more voltage on one of the coils than the other. Motion in the other direction reverses the action with a 180° phase shift occurring at null. The output voltage can be modified depending on the requirements in signal processing by means of rectification, demodulation, or filtering. In these instruments, full-scale motion may be extremely small, on the order of few thousandths of a centimeter.

In general, variable-reluctance transducers have small ranges and are used in specialized applications such as pressure transducers. Magnetic forces imposed on the armature are quite large, and this limits the application severely.

Variable-Reluctance Tachogenerators. Another example of the variable-reluctance sensor is shown in Fig. 4. These sensors

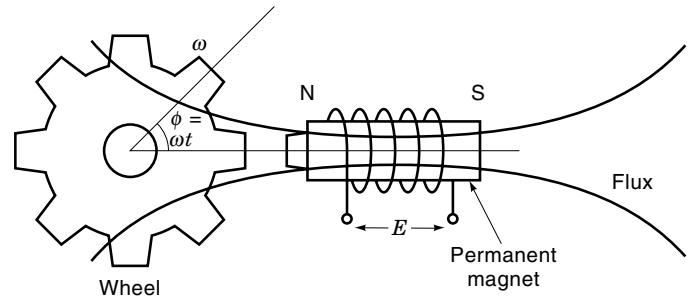


Figure 4. A variable-reluctance tachogenerator is a sensor based on Faraday's Law of Electromagnetic Induction. It consists of a ferromagnetic toothed wheel attached to the rotating shaft and a coil wound onto a permanent magnet extended by a soft iron pole piece. The wheel rotates in close proximity to the pole piece, thus causing the flux linked by the coil to change. The change in flux causes an output in the coil similar to a square waveform whose frequency depends on the speed of the rotation of the wheel and the number of teeth.

are based on Faraday's Law of Electromagnetic Induction; therefore, they may also be referred as electromagnetic sensors. Basically, the induced electromagnetic force (emf) in the sensor depends on the linear or angular velocity of the motion.

The variable-reluctance tachogenerator consists of a ferromagnetic toothed wheel attached to the rotating shaft, and a coil wound onto a permanent magnet, extended by a soft iron pole piece. The wheel moves in close proximity to the pole piece, causing the flux linked by the coil to change, thus inducing an emf in the coil. The reluctance of the circuit depends on the width of the air gap between the rotating wheel and the pole piece. When the tooth is close to the pole piece, the reluctance is at a minimum, and it increases as the tooth moves away from the pole. If the wheel rotates with a velocity ω , the flux may mathematically be expressed as

$$\Psi(\theta) = \Psi_m + \Psi_f \cos m\theta \quad (12)$$

where Ψ_m is the mean flux, Ψ_f is the amplitude of the flux variation, and m is the number of teeth.

The induced emf is given by

$$E = -\frac{d\Psi(\theta)}{dt} = -\frac{d\Psi(\theta)}{d\theta} \times \frac{d\theta}{dt} \quad (13)$$

or

$$E = \Psi_f m \omega \sin n\omega t \quad (14)$$

Both the amplitude and the frequency of the generated voltage at the coil are proportional to the angular velocity of the wheel. In principle, the angular velocity ω can be found from either the amplitude or the frequency of the signal. In practice, the amplitude measured may be influenced by loading effects and electrical interference. In signal processing, the frequency is the preferred option because it can be converted into digital signals easily.

The variable-reluctance tachogenerators are most suitable for measuring angular velocities. They are also used for volume flow rate measurements and the total volume flow determination of fluids.

Microsyn. Another commonly used example of variable-reluctance transducer is the microsyn, as illustrated in Fig. 5. In this arrangement, the coils are connected in such a manner that at the null position of the rotary element, the voltages induced in coils 1 and 3 are balanced by voltages induced in coils 2 and 4. The motion of the rotor in the clockwise direction increases the reluctance of coils 1 and 3 while decreasing the reluctance of coils 2 and 4, thus giving a net output voltage e_o . The movement in the counterclockwise direction causes a similar effect in coils 2 and 4 with a 180° phase shift. A direction-sensitive output can be obtained by using phase-sensitive demodulators.

Microsyn transducers are extensively used in applications involving gyroscopes. By using microsyns, very small motions can be detected giving an output signal as low as 0.01° of changes in angles. The sensitivity of the device can be made as high as 5 V per degree of rotation. The nonlinearity may vary from 0.5% to 1.0% full scale. The main advantage of these transducers is that the rotor does not have windings and slip rings. The magnetic reaction torque is also negligible.

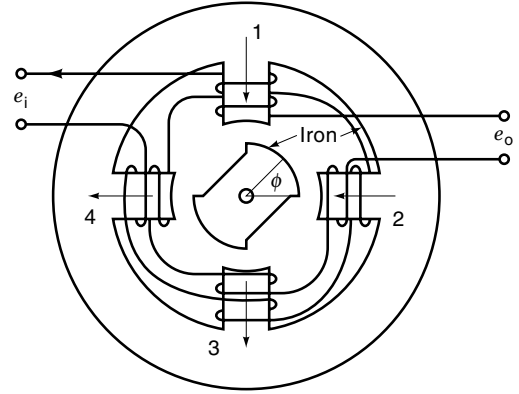


Figure 5. A microsyn is a variable-reluctance transducer that consists of a ferromagnetic rotor and a stator carrying four coils. The stator coils are connected such that at the null position, the voltages induced in coils 1 and 2 are balanced by voltages induced in coils 3 and 4. The motion of the rotor in one direction increases the reluctance of two opposite coils while decreasing the reluctance in others resulting in a net output voltage e_o . The movement in the opposite direction reverses this effect with a 180° phase shift.

Synchros

The term *synchro* is associated with a family of electromechanical devices. They are primarily used in angle measurements and are commonly applied in control engineering as parts of servomechanisms, machine tools, antennas, and the like.

The construction of synchros is similar to that of wound-rotor induction motors, as shown in Fig. 6. The rotation of the motor changes the mutual inductance between the rotor coil and the three stator coils. The three voltage signals from these coils define the angular position of the rotor. Synchros are used in

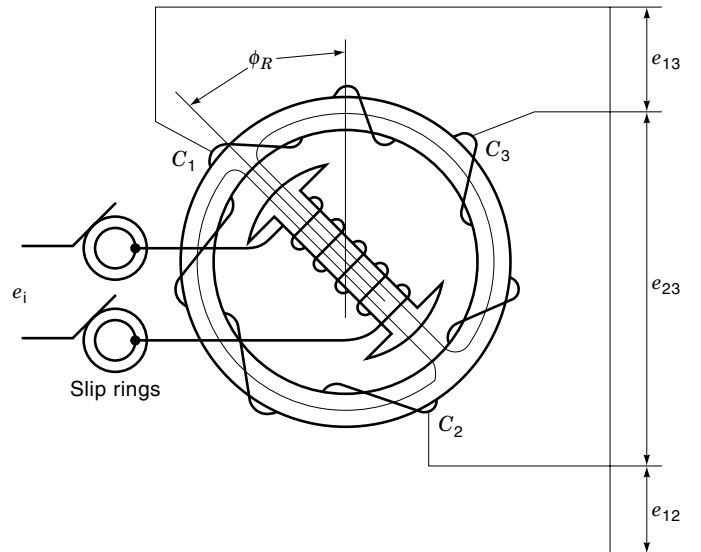


Figure 6. A synchro is similar to a wound-rotor induction motor. The rotation of the rotor changes the mutual inductance between the rotor coil and the three stator coils. The voltages from these coils define the angular position of the rotor. They are primarily used in angle measurements and are commonly applied in control engineering as parts of servomechanisms, machine tools, antennas, and the like.

connection with a variety of devices, such as control transformers, Scott T transformers, resolvers, phase-sensitive demodulators, and analog-to-digital (AD) converters.

In some cases, a control transformer is attached to the outputs of the stator coils such that the output of the control transformer produces a resultant mmf aligned in the same direction as that of the rotor of the synchro. In other words, the synchro rotor acts as a search coil in detecting the direction of the stator field of the control transformer. When the axis of this coil is aligned with the field, the maximum voltage is supplied to the transformer.

In other cases, ac signals from the synchros are first applied to a Scott T transformer, which produces ac voltages with amplitudes proportional to the sine and cosine of the synchro shaft angle. It is also possible to use phase-sensitive demodulations to convert the output signals to make them suitable for digital signal processing.

Linear Variable Inductor

There is very little distinction between variable-reluctance and variable-inductance transducers. Mathematically, the principles of linear variable transducers are very similar to the variable-reluctance type of transducers. The distinction is mainly in the sensing rather than principles of operations. A typical linear variable inductor consists of a movable iron core to provide the mechanical input and the two coils forming two legs of bridge network. A typical example of such a transducer is the variable coupling transducer.

The variable-coupling transducers consist of a former holding a center-tapped coil and a ferromagnetic plunger, as shown in Fig. 7. The plunger and the two coils have the same length l . As the plunger moves, the inductances of the coils change. The two inductances are usually placed to form two arms of a bridge circuit with two equal balancing resistors. The bridge is then excited with ac of 5 V to 25 V with a frequency of 50 Hz to 5 kHz. At the selected excitation frequency, the total transducer impedance at null conditions is set in the 100 Ω to 1000 Ω range. The resistors are set to have about the same value as transducer impedances. The load for the bridge output must be at least ten times the resistance R value. When the plunger is in the reference position, each coil will have equal inductances of value L . As the plunger moves by δl , changes in inductances $+\delta L$ and $-\delta L$ create a voltage output from the bridge. By constructing the bridge carefully, the output voltage may be made as a linear function displacement of the moving plunger within a rated range.

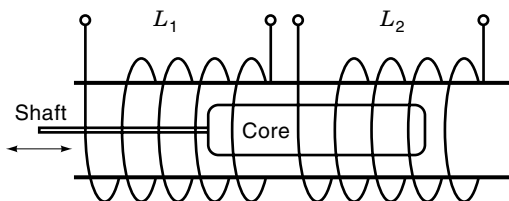


Figure 7. A typical linear variable inductor consists of a movable iron core inside a former holding a center-tapped coil. The core and both coils have the same length l . When the core is in the reference position, each coil will have equal inductances of value L . As the core moves by δl , changes in inductances $+\delta L$ and $-\delta L$ create voltage outputs from the coils.

In some transducers, in order to reduce power losses resulting from the heating of resistors, center-tapped transformers may be used as a part of the bridge network. In this case, the circuit becomes more inductive, and extra care must be taken to avoid the mutual coupling between the transformer and the transducer.

It is particularly easy to construct transducers of this type, by simply winding a center-tapped coil on a suitable former. The variable-inductance transducers are commercially available in strokes from about 2 mm to 500 cm. The sensitivity ranges between 1% full scale to 0.02% in long-stroke special constructions. These devices are also known as linear displacement transducers or LDTs, and they are available in various shapes and sizes.

Apart from linear variable inductors, rotary types are also available. Their cores are specially shaped for rotational applications. Their nonlinearity can vary between 0.5% and 1% full scale over a range of 90° rotation. Their sensitivity can be up to 100 mV per degree of rotation.

Induction Potentiometer

A version of rotary-type linear inductors is the induction potentiometer, as shown in Fig. 8. Two concentrated windings are wound on stator and rotor. The rotor winding is excited with an ac, thus inducing voltage in the stator windings. The amplitude of the output voltage is dependent on the mutual inductance between the two coils, where mutual inductance itself is dependent on the angle of rotation. For concentrated coil-type induction potentiometers, the variation of the amplitude is sinusoidal, but linearity is restricted in the region of the null position. A linear distribution over an angle of 180° may be obtained by carefully designed distributed coils.

Standard commercial induction pots operate in a 50 Hz to 400 Hz frequency range. They are small in size from 1 cm to

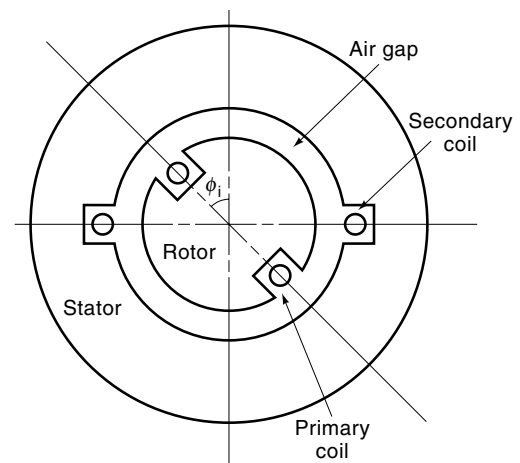


Figure 8. An induction potentiometer is a linear variable inductor with two concentrated windings wound on the stator and rotor. The rotor winding is excited with ac, inducing voltage in the stator windings. The amplitude of the output voltage is dependent on the relative positions of the coils determined by the angle of rotation. For concentrated coils, the variation of the amplitude is sinusoidal, but linearity is restricted in the region of the null position. Different types of induction potentiometers are available with distributed coils, which give linear voltages over an angle of 180° of rotation.

6 cm, and their sensitivity can be in the order of $1\text{ V}/1^\circ$ of rotation. Although the ranges of induction pots are limited to less than 60° of rotation, it is possible to measure displacements in angles from 0° to full rotation by suitable arrangements of a number of induction pots. As in the case of most inductive sensors, the output of the induction pots may need phase-sensitive demodulators and suitable filters. In many inductive pots, additional dummy coils are used to improve linearity and accuracy.

EDDY CURRENT SENSORS

Inductive transducers based on eddy currents are mainly probe types containing two coils, as shown in Fig. 9. One of the coils, known as the active coil, is influenced by the presence of the conducting target. The second coil, known as the balance coil, serves to complete the bridge circuit and provides temperature compensation. The magnetic flux from the active coil passes into the conductive target by means of a probe. When the probe is brought close to the target, the flux from the probe links with the target, producing eddy currents within the target.

The eddy current density is greatest at the target surface and become negligibly small about three skin depths below the surface. The skin depth depends on the type of material used and the excitation frequency. Even though thinner targets can be used, a minimum of three skin depths may often be necessary to minimize the temperature effects. As the target comes closer to the probe, the eddy currents become stronger, causing the impedance of the active coil to change and altering the balance of the bridge in relation to the target position. This unbalance voltage of the bridge may be demodulated, filtered, and linearized to produce a dc output proportional to target displacement. The bridge oscillation may be as high as 1 MHz. High frequencies allow the use of thin targets and provide a good system frequency response.

Probes are commercially available with full-scale ranges from 0.25 mm to 30 mm with a nonlinearity of 0.5% and a maximum resolution of 0.0001 mm. Targets are usually supplied

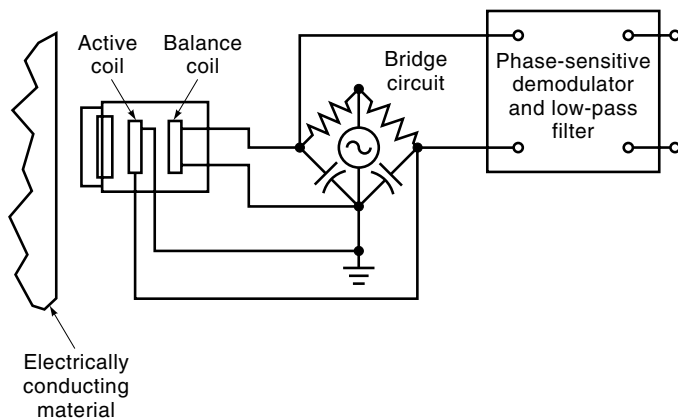


Figure 9. Eddy current transducers are inductive transducers using probes. The probes contain one active and one balance coil. The active coil responds to the presence of a conducting target, whereas the balance coil completes a bridge circuit and provides temperature compensation. When the probe is brought close to the target, the flux from the probe links with the target producing eddy currents within the target, which alter the inductance of the active coil. This change in inductance is detected by a bridge circuit.

by the clients, involving noncontact measurements of machine parts. For nonconductive targets, conductive materials of sufficient thickness must be attached onto the surface by means of commercially available adhesives. Because the target material, shape, and the like influence the output, it is necessary to calibrate the system statistically for a specific target. The recommended measuring range of a given probe begins at a standoff distance equal to about 20% of the stated range of the probe. In some cases, a standoff distance of 10% of the stated range for which the system is calibrated is recommended as standard. A distance greater than 10% of the measuring range can be used as long as the calibrated measuring range is reduced by the same amount.

Flat targets must be the same diameter as the probe or larger. If the target diameter is smaller than the probe diameter, the output drops considerably, thus becoming unreliable. Curved-surface targets may behave similar to flat surfaces if the diameter exceeds about three or four diameter of the probe. In this case, the target essentially becomes an infinite plane. This also allows some cross-axis movement without affecting the system output. Target diameter comparable to the sensor could result in detrimental affects from cross-axis movements.

For curved or irregularly shaped targets, the system needs to be calibrated using an exact target that may be seen in the operation. This tends to eliminate any errors caused by the curved surfaces during the applications. However, special multiprobe systems are available for orbital motions of rotating shafts. If the curved (shaft) target is about ten times greater than the sensor diameter, it acts as an infinite plane and does not need special calibrations. Special care must be exercised to deal with electrical runout resulting from factors such as inhomogeneities in hardness, particularly valid for ferrous targets. However, nonferrous targets are free from electrical runout concerns.

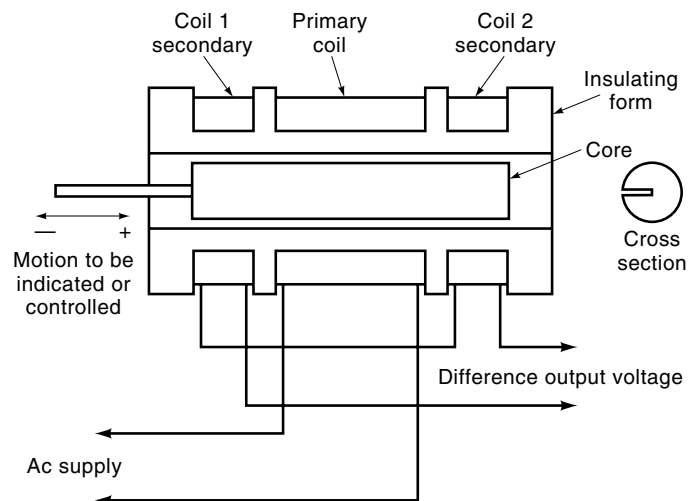


Figure 10. A linear variable-differential transformer is a passive inductive transducer consisting of a single primary winding positioned between two identical secondary windings wound on a tubular ferromagnetic former. As the core inside the former moves, the magnetic paths between primary and secondaries alter, thus giving secondary outputs proportional to the movement. The two secondaries are made as similar as possible by having equal sizes, shapes, and number of turns.

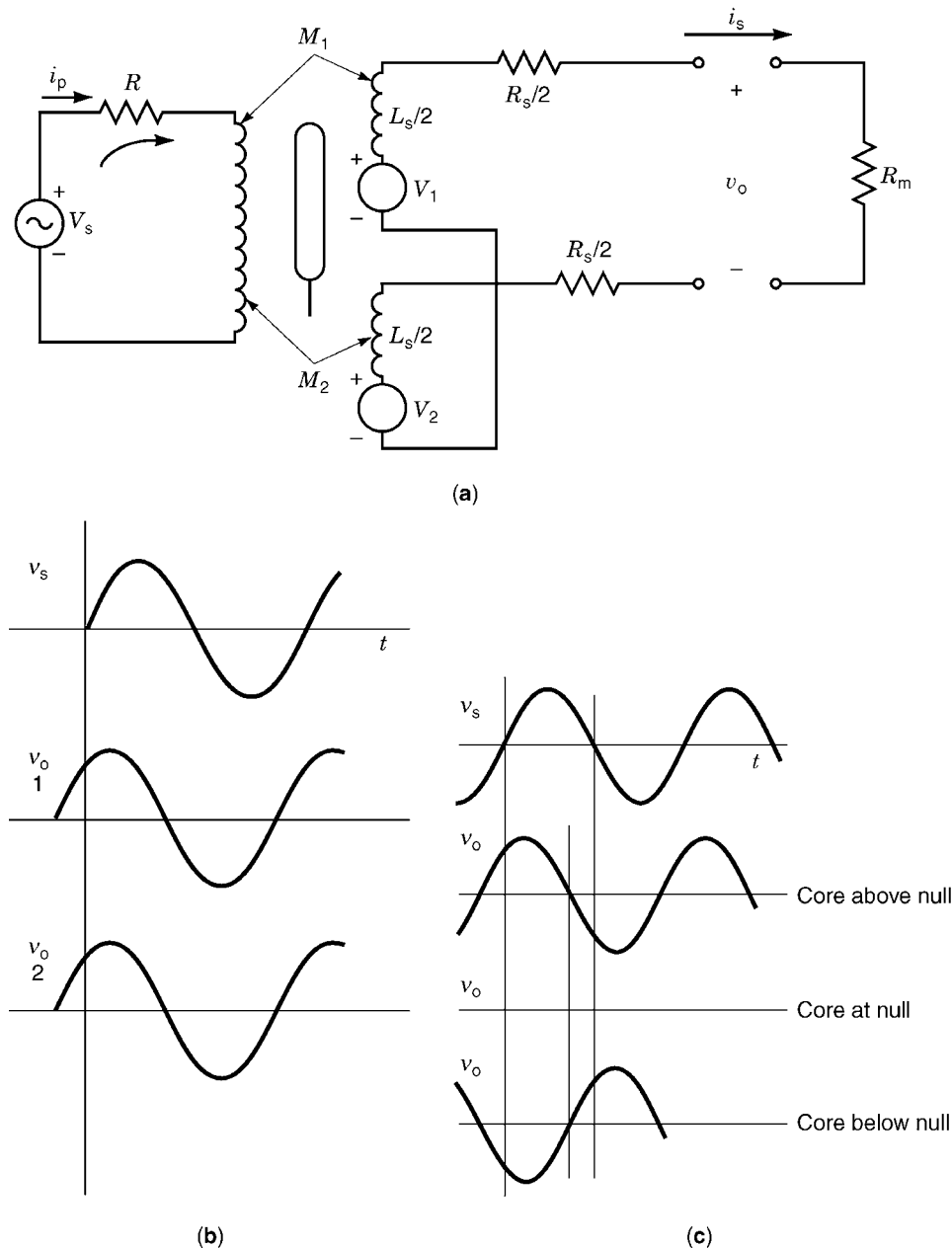


Figure 11. The voltages induced in the secondaries of a linear variable-differential transformer (a) may be processed in a number of ways. The output voltages of individual secondaries v_1 and v_2 at null position are illustrated in (b). In this case, the voltages of individual coils are equal and in phase with each other. Sometimes the outputs are connected opposing each other, and the output waveform v_o becomes a function of core position x and phase angle α as in (c). Note the phase shift of 180° as the core position changes above and below the null position.

TRANSFORMATIVE SENSORS

Transformative sensors make use of the principles of transformer action, that is magnetic flux created by one coil links with the other coil to induce voltages. There are many different types, such as linear variable transformers, rotary variable differential transformers, and flux-gate magnetometers.

Linear Variable-Differential Transformer

The linear variable-differential transformer (LVDT) is a passive inductive transducer that has found many applications. It consists of a single primary winding positioned between two

identical secondary windings wound on a tubular ferromagnetic former, as shown in Fig. 10. The primary winding is energized by a high-frequency 50 Hz to 20 kHz ac voltage. The two secondaries are made identical by having an equal number of turns. They are connected in series opposition so that the induced output voltages oppose each other.

In many applications, the outputs are connected in opposing form, as shown in Fig. 11(a). The output voltages of individual secondaries v_1 and v_2 at null position are illustrated in Fig. 11(b). However, in opposing connection, any displacement in the core position x from the null point causes amplitude of the voltage output v_o and the phase difference α to change. The output waveform v_o in relation to core position is shown in

Fig. 11(c). When the core is positioned in the middle, there is an equal coupling between primary and secondaries, thus giving a null point or reference point of the sensor. As long as the core remains near the center of the coil arrangement, output is very linear. The linear ranges of commercial differential transformers are clearly specified, and the devices are seldom used outside this linear range.

The ferromagnetic core or plunger moves freely inside the former; thus altering the mutual inductance between the primary and secondaries. With the core in the center, or at the reference position, the induced emfs in the secondaries are equal, and because they oppose each other, the output voltage is zero. When the core moves, say to the left, from the center, more magnetic flux links with the left-hand coil than with the right-hand coil. The voltage induced in the left-hand coil is therefore larger than the induced emf on the right-hand coil. The magnitude of the output voltage is then larger than at the null position and is equal to the difference between the two secondary voltages. The net output voltage is in phase with the voltage of the left-hand coil. The output of the device is then an indication of displacement of the core. Similarly, movement in the opposite direction to the right from the center reverses this effect, and the output voltage is now in phase with the emf of the right-hand coil.

For mathematical analysis of the operation of LVDTs Fig. 11(a) may be used. The voltages induced in the secondary coils are dependent on the mutual inductance between the primary and individual secondary coils. Assuming that there is no cross coupling between the secondaries, the induced voltages may be written as

$$v_1 = M_1 s i_p \quad \text{and} \quad v_2 = M_2 s i_p \quad (15)$$

where M_1 and M_2 are the mutual inductances between primary and secondary coils for a fixed core position, s is the Laplace operator, and i_p is the primary current.

In the case of opposing connection, no load output voltage v_o without any secondary current may be written as

$$v_o = v_1 - v_2 = (M_1 - M_2) s i_p \quad (16)$$

writing

$$v_s = i_p (R + s L_p) \quad (17)$$

Substituting i_p in Eq. (16) gives the transfer function of the transducer as

$$\frac{v_o}{v_s} = \frac{(M_1 - M_2)s}{R + s L_p} \quad (18)$$

However, If there is a current resulting from output signal processing, then describing equations may be modified as

$$v_o = R_m i_s \quad (19)$$

where $i_s = (M_1 - M_2) s i_p / (R_s + R_m + s L_s)$ and

$$v_s = i_p (R + s L_p) - (M_1 - M_2) s i_s \quad (20)$$

Eliminating i_p and i_s from Eqs. (19) and (20) results in a transfer

function

$$\frac{v_o}{v_s} = \frac{R_m (M_1 - M_2) s}{[(M_1 - M_2)^2 + L_s L_p] s^2 + [L_p (R + R_m) + R L_s] s + (R_s + R_m) + R} \quad (21)$$

This is a second-order system, which indicates that with the effect of the numerator the frequency of the system changes from $+90^\circ$ at low frequencies to -90° at high frequencies. In practical applications, the supply frequency is selected such that at null position of the core the phase angle of the system is 0° .

The amplitudes of the output voltages of secondary coils are dependent on the position of the core. These outputs may directly be processed from each individual secondary coils for slow movements of the core, if the direction of the movement of the core does not bear any importance. However, for fast movements of the core, the signals may be converted to dc, and the direction of the movement from the null position may be detected. There are many options to do this; however, a *phase-sensitive demodulator* and filter are commonly used as shown in Fig. 12(a). A typical output of the phase-sensitive demodulator is illustrated in Fig. 12(b), for core positions as in Fig. 12(c), in relation to output voltage v_o , displacement x , and phase angle α .

The phase-sensitive demodulators are extensively used in differential-type inductive sensors. They basically convert the ac outputs to dc values and also indicate the direction of movement of the core from the null position. A typical phase-sensitive demodulation circuit may be constructed, based on diodes shown in Fig. 13(a). This arrangement is useful for very slow displacements, usually less than 1 or 2 Hz. In Fig. 13(a), bridge 1 acts as a rectification circuit for secondary 1, and bridge 2 acts as a rectifier for secondary 2. The net output voltage is the difference between the outputs of two bridges as in Fig. 13(b). The position of the core can be worked out from the amplitude of the dc output and the direction of the movement of the core can be determined from the polarity of the dc voltage. For rapid movements of the core, the output of the diode bridges need to be filtered, and this passes only the frequencies of the movement of the core and filters all the other frequencies produced by the modulation process. For this purpose, a suitably designed simple *RC* filter may be sufficient.

In the marketplace, there are phase-sensitive demodulator chips available, such as AD598 offered by Analog Devices, Inc. These chips are highly versatile and flexible to suit particular application requirements. They offer many advantages over conventional phase-sensitive demodulation devices; for example, frequency of excitation may be adjusted to any value between 20 Hz and 20 kHz by connecting an external capacitor between two pins. The amplitude of the excitation voltage can be set up to 24 V. The internal filters may be set to required values by external capacitors. Connections to analog-to-digital converters are made easy by converting the bipolar output to a unipolar scale.

The frequency response of LVDTs is primarily limited by the inertia characteristics of the device. In general, the frequency of the applied voltage should be ten times the desired frequency response. Commercial LVDTs are available in a broad range of sizes, and they are widely used for displacement measurements in a variety of applications. The displacement sensors are available to cover ranges from ± 0.25 mm to ± 7.5 cm. They are sensitive enough to be used to respond to displacements

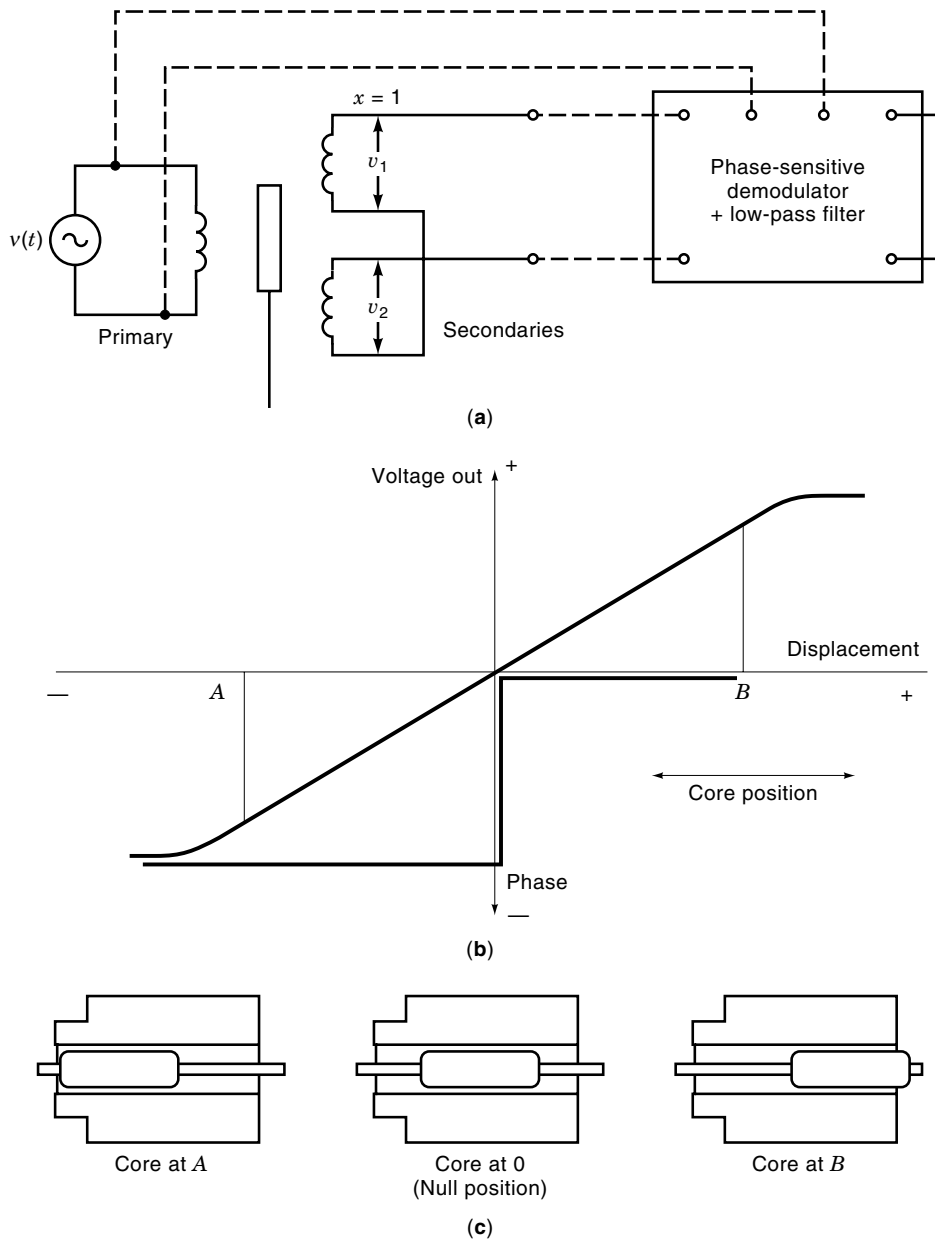


Figure 12. Phase-sensitive demodulator and filter (a) are commonly used to obtain displacement-proportional signals from LVDTs and other differential-type inductive sensors. They convert the ac outputs from the sensors into dc values and also indicate the direction of movement of the core from the null position. A typical output of the phase-sensitive demodulator is shown in (b). The relationship between output voltage v_o and phase angle α is also shown against core position x as sketched in (c).

well below 0.0005 mm. They can have operational temperature range from -265° to 600°C . They are also available in radiation-resistant designs for operation in nuclear reactors. For a typical sensor of range ± 25 mm, the recommended supply voltage is 4 V to 6 V, with a nominal frequency of 5 kHz and a maximum nonlinearity of 1% full scale. Several commercial models, which can produce a voltage output of 300 mV for 1 mm displacement of the core, are available.

One important advantage of the LVDTs is that there is no physical contact between the core and the coil form; hence there is no friction or wear. Nevertheless, there are radial and longitudinal magnetic forces on the core at all times. These magnetic forces may be regarded as magnetic springs that try to displace the core from its null position. This may be a critical factor in some applications.

One problem with LVDTs is that it may not be easy to make the two halves of the secondary identical; their inductance,

resistance, and capacitance may be different, causing a large unwanted quadrature output in the balance position. Precision coil-winding equipment may be required to reduce this problem to an acceptable value.

Another problem is associated with null position adjustments. The harmonics in the supply voltage and stray capacitances result in small null voltages. The null voltage may be reduced by proper grounding, which reduces the capacitive effects and the center-tapped voltage source arrangements. In center-tapped supplies, a potentiometer may be used to obtain a minimum null reading.

The LVDTs find a variety of applications, which include jet engines controls that are in close proximity to exhaust gases, and controls that measure roll positions in the thickness of materials in hot-slab steel mills. After some mechanical conversions, LVDTs may also make force and pressure measurements.

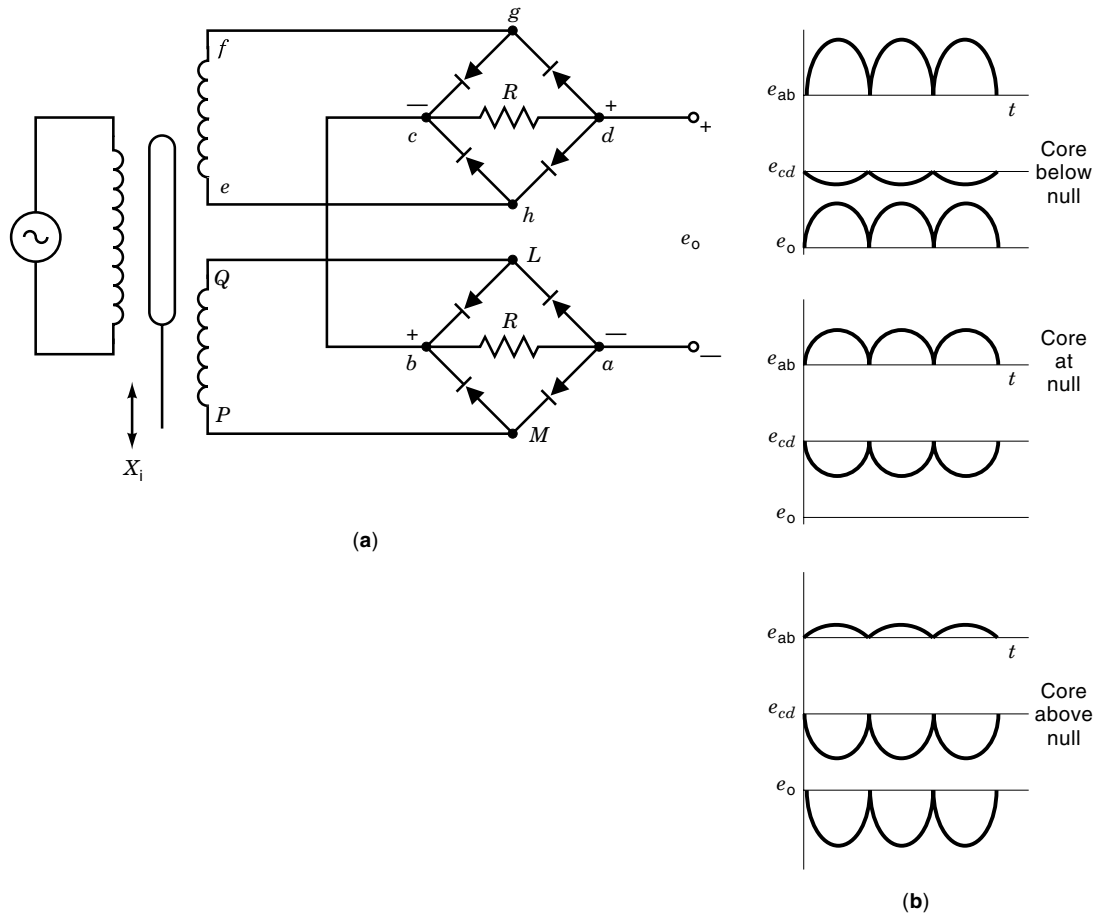


Figure 13. A typical phase-sensitive demodulation circuit based on diode bridges as in (a). The bridge 1 acts as a rectification circuit for secondary 1, and bridge 2 acts as a rectifier for secondary 2 where the net output voltage is the difference between the two bridges as in (b). The position of the core can be worked out from the amplitude of the dc output, and the direction of the movement of the core can be determined from the polarity of the voltage. For rapid movements of the core, the outputs of the diode bridges need to be filtered. For filters, a suitably designed simple RC filter may be sufficient.

Rotary Variable-Differential Transformer

A variation from the linear variable-differential transformer is the rotary core differential transformer, as shown in Fig. 14. Here the primary winding is wound on the center leg of an E core; the secondary windings are wound on the outer legs of the E core. The armature is rotated by an externally applied force about a pivot point above the center leg of the core. When the armature is displaced from its reference or balance position, the reluctance of the magnetic circuit through one secondary coil is decreased; simultaneously the reluctance through the other coil is increased. The induced emfs in the secondary windings, which are equal in the reference position of the armature, are now different in magnitude and phase as a result of the applied displacement. The induced emfs in the secondary coils are made to oppose each other, and the transformer operates in the same manner as LVDTs. The rotating variable transformers may be sensitive to vibrations. If a dc output is required, a demodulator network may be used, as in the case of LVDTs.

In most rotary linear-variable differential transformers, the rotor mass is very small, usually less than 5 g. The nonlinearity

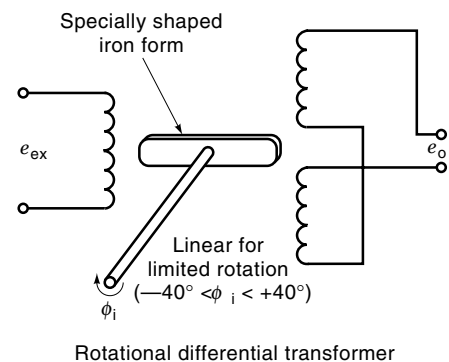


Figure 14. A rotary core differential transformer has an armature rotated by an externally applied force about a pivot point above the center leg of the core. When the armature is displaced from its reference or balance position, the reluctance of the magnetic circuit through one secondary coil is decreased; simultaneously the reluctance through the other coil is increased. The induced emfs in the secondary windings are different in magnitude and phase as a result of the applied displacement.

in the output ranges between $\pm 1\%$ and $\pm 3\%$, depending on the angle of rotation. The motion in the radial direction produces a small output signal that can affect the overall sensitivity. But this transverse sensitivity is usually kept less than 1% of the longitudinal sensitivity.

MOS MAGNETIC FIELD SENSORS

The technology of integrated magnetic field sensors, also called semiconductor magnetic microsensors, is well developed. This technology uses either high-permeability (e.g., ferromagnetic) or low-permeability (e.g., paramagnetic) materials. The integrated circuit magnetic techniques support many sensors, such as magnetometers, optoelectronics, Hall-effect sensors, magnetic semiconductor sensors, and superconductive sensors.

At the present, silicon offers an advantage of inexpensive batch fabrication for magnetic sensors. Most of the integrated circuit magnetic sensors are manufactured by following the design rules of standard chip manufacturing. For example, MOS and complementary metal oxide semiconductor (CMOS) technologies are used to manufacture highly sensitive Hall-effect sensors, magnetotransistors, and other semiconductor sensors. Some of the magnetic semiconductor sensors are discussed in detail in the following sections.

MAGNETOMETERS

Magnetometers are devices that are produced to sense external magnetic fields mainly based on Faraday's Law of Induction. There are many different types of magnetometers including search coil magnetometers, SQUIDS, flux-gates, nuclear, and optical magnetometers. The two most commonly used types are the flux-gate and the search coil magnetometers.

Flux-Gate Magnetometers

Flux-gate magnetometers are made from two coils wound on a ferromagnetic material, as illustrated in Fig. 15. One of the coils is excited with a sinusoidal current, which drives the core into a saturation state. At this stage, the reluctance of the core to external magnetic field increases, thus repelling the external flux and so reducing the effect of the external field on the second coil. As the core becomes unsaturated, the effect of the external field increases. The increase and reduction of the effect of the

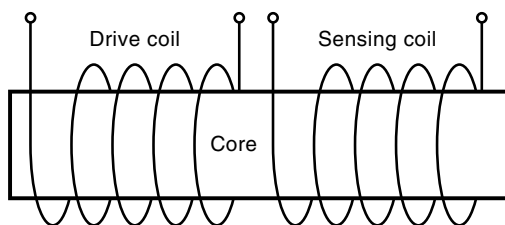


Figure 15. A flux-gate magnetometer consists of two coils wound on a ferromagnetic core. The driving coil is excited with a sinusoidal current, which drives the core into a saturation state. When saturated, the reluctance of the core to external magnetic field increases, thus repelling the external flux and hence reducing the effect of the external field on the second coil. The harmonics of the induced voltage in the sensing coil is an indication of the magnitude of the external magnetic field.

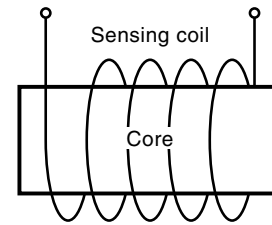


Figure 16. A search coil magnetometer uses Faraday's Law of Induction. The changing magnetic flux induces an emf in the coil. The sensitivity of the magnetometer depends on the properties of the core material, dimensions of the coil, number of turns, and rate of change of flux through the coil. They can sense weak fields as low as 100 pT within the frequency range of 1 Hz to 1 MHz.

external field on the second coil is sensed as the harmonics of the induced voltage. These harmonics are then directly related to the strength and variations in the external field. The sensitivity of these devices is dependent on the magnetic properties and the shape of the saturation curve of the core material. The measurement range can vary from 100 pT to 10 mT operating in the 0 kHz to 10 kHz frequency range.

Search Coil Magnetometers

Search coil magnetometers operate on the principle of Faraday's Law of Induction. A typical search coil magnetometer is shown in Fig. 16. The flux through the coil changes if the magnetic field varies in time or the coil is moved through the field. The sensitivity depends on the properties of the core material, dimensions of the coil, number of turns, and rate of change of flux through the coil.

Search coil magnetometers are manufactured from 4 cm in dimensions to 100 cm. They can sense weak fields as low as 100 pT within the frequency range of 1 Hz to 1 MHz. The upper limit is dependent on the relative magnitudes of resistance and inductance of the coil. During the signal processing, they can be used as part of a bridge or resonant circuits.

SQUID Magnetometers

Superconducting quantum interference device, SQUID, sensors are used in many diverse applications from medicine, geophysics, and nuclear magnetic resonance to nondestructive testing of solid materials. The SQUID offers high sensitivity for the detection of weak magnetic fields and field gradients. They are made from conventional superconductors such as niobium operating at liquid helium temperatures. They are manufactured by using integrated circuit technology, as illustrated in Fig. 17.

The principle of operation is the Josephson effect. If a magnetic flux links to a ring-shaped superconducting material, a current is induced in the ring. This current flows forever due to the lack of any resistance. The intensity of current is proportional to the intensity of the field. The current in the ring is measured by using the Josephson effect wherein a weak link in the superconducting ring causes the superconducting current to oscillate as a function of magnetic field intensity. These oscillations can be sensed by many techniques such as coupling to a radio-frequency circuit and other resonance techniques.

The SQUIDS are extremely sensitive devices, with sensitivities ranging from 10 fT to 10 nT. The SQUID sensors can be

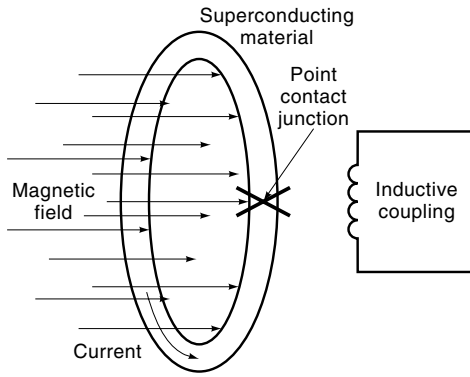


Figure 17. A SQUID consists of a superconducting ring. When subjected to an external field, the current is induced in the ring, which flows forever. The current in the ring is measured by using the Josephson effect and by creating a weak link in the ring. This weak link makes the superconducting current oscillate as a function of the external magnetic field intensity.

arranged to measure magnetic fields in three dimensions in x , y , and z directions of Cartesian coordinates, as in the case of high-sensitivity gradiometers. One disadvantage of SQUIDs is that they need supercooling at very low temperatures. Much research is concentrated on materials exhibiting superconductivity properties at high temperatures. SQUIDs are used in gradiometers, voltmeters and amplifiers, displacement sensors, geophysics, gravity wave detection, and nondestructive testing, to name a few.

MAGNETORESISTIVE SENSORS

In magnetoresistive sensors, the magnetic field causes a change in resistance of some materials such as permalloys. In these materials, current passing through the material magnetizes the material in a particular magnetic orientation. An external field perpendicular to the current, as illustrated in Fig. 18, causes the magnetic orientation to be disturbed. The resistance is highest when the magnetization is parallel to the current and lowest when it is perpendicular to the current. Hence, depending on the intensity of the external magnetic field, the resistance of the permalloy changes in proportion.

Magnetoresistive sensors are manufactured as thin films and usually integrated to be part of an appropriate bridge

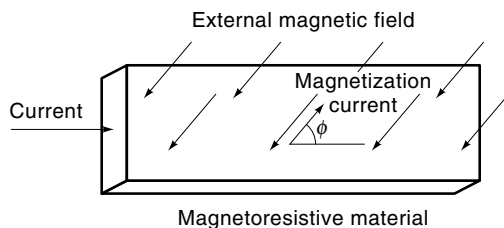


Figure 18. A magnetoresistive sensor's resistance changes in response to an external magnetic field. Magnetic domain orientation of the material is a function of external field and current flowing through it. The resistance is highest when the magnetization is parallel to the current and lowest when it is perpendicular to the current. These sensors are manufactured as thin films, and they have good linearity.

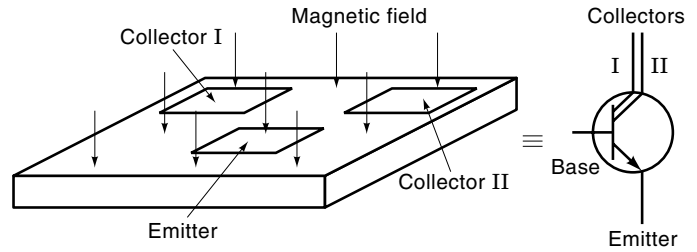


Figure 19. A magnetotransistor is an integrated silicon device that contains nnp or pnp junctions. The electron flow between emitter and collector is influenced by the external magnetic field in favor of one of the collectors. The two collector voltages are sensed and related to the applied magnetic field.

circuit. They have good linearity and low temperature coefficients. These devices have a sensitivity ranging from $1 \mu\text{T}$ to 50 mT . By improved electronics with suitable feedback circuits, the sensitivity can be as low as 100 pT . They can operate from dc to several gigahertz.

Magnetotransistor and Magnetodiode Sensors

Magnetotransistor and magnetodiode sensors are integrated silicon devices. They contain n -doped and p -doped regions forming pn , nnp , or pnp junctions. In the case of magnetotransistors, there are two collectors, as shown in Fig. 19. Depending on the direction, an external magnetic field deflects electron flow between emitter and collector in favor of one of the collectors. The two collector voltages are sensed and related to the applied magnetic field. These devices are more sensitive than Hall-effect sensors.

In the case of magnetodiodes, p and n regions are separated by an area of undoped silicon containing the sensor. An external magnetic field perpendicular to the flow of charges deflects the holes and electrons in the opposite directions, resulting effectively in a change in the resistance of the undoped silicon layer.

HALL-EFFECT SENSORS

In Hall-effect sensors, the voltage difference across a thin conductor carrying current depends on the intensity of the magnetic field applied perpendicular to the direction of current flow, as shown in Fig. 20. An electron moving through a magnetic field experiences Lorentz force perpendicular to the direction of motion and to the direction of the field. The response of electrons to Lorentz force creates a voltage known as the Hall voltage. If a current I flows through the sensor, the Hall voltage can mathematically be found by

$$V = R_H IB/t \quad (22)$$

where R_H is the Hall coefficient (cubic meters per degree Celsius), B is the flux density (tesla), and t is the thickness of the sensor (meters).

Therefore, for a specified current and temperature, the voltage is proportional to B . If the current is dc, the voltage has the same frequency as magnetic flux.

Hall-effect sensors can be made by using metals or silicon, but they are generally made from semiconductors with high

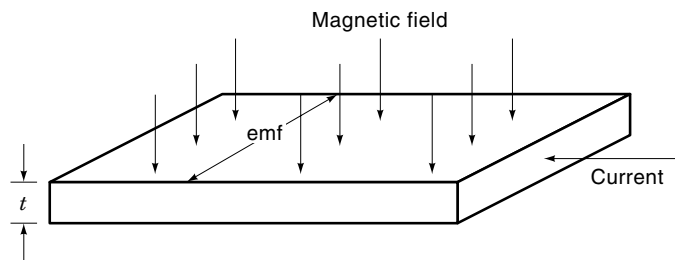


Figure 20. A Hall-effect sensor makes use of Lorentz force. The response of electrons to Lorentz force creates the Hall voltage, which is perpendicular to both the external magnetic field and the direction of current flow. If the current is dc, the voltage has the same frequency as magnetic flux.

electron mobility such as indium antimonide. They are usually manufactured in the form of probes with a sensitivity down to $100 \mu\text{T}$. Silicon Hall-effect sensors can measure constant or varying magnetic flux having an operational frequency from dc to 1 MHz, within the range of 1 mT to 100 mT. They have good temperature characteristics from 200°C to near absolute zero.

MAGNETO-OPTICAL SENSORS

In recent years, highly sensitive magneto-optical sensors have been developed. These sensors are based on fiber-optics, polarization of light, Moire effect, and Zeeman effect, among others. This type of sensors leads to highly sensitive devices and is used in applications requiring high resolution such as human brain function mapping and magnetic anomaly detection. Here, because of the availability of space, only the polarization effect will be discussed briefly. Interested readers can find further information in the Refs. 1–6.

In the polarization effect, a plane of polarized light in a strong magnetic field rotates its plane of vibration. The per-unit angular rotation is related to the per-unit magnetic field in a given length of material by the Verdet constant. For example, terbium gallium demonstrates a Verdet constant of $50 \text{ min}/\mu\text{T}\cdot\text{cm}$, whereas bismuth-substituted iron garnet can have up to $0.04 \text{ min}/\text{T}\cdot\text{cm}$. This polarization effect was first noticed by Faraday in 1845; hence, it is generally known as the Faraday effect. Recently, this principle was applied to semiconductors and crystals, which have different physical properties such as interband effects, intraband free carrier effects, and absorption of magnetism by impurities.

MAGNETIC THIN FILMS

Magnetic thin films are an important part of superconducting instrumentation, sensors and electronics in which active devices are made from deposited films. The thin films are usually made from amorphous alloys, amorphous gallium, and the like. As an example of this use of thin-film technology, a thin-film Josephson junction is given in Fig. 21. The deposition of thin films can be done by thermal evaporation, electroplating, sputter deposition, or chemical methods. The choice of technology depends on the characteristics of the sensors. For example, thin-film superconductors require low-temperature operations, whereas common semiconductors operate at room temperature.

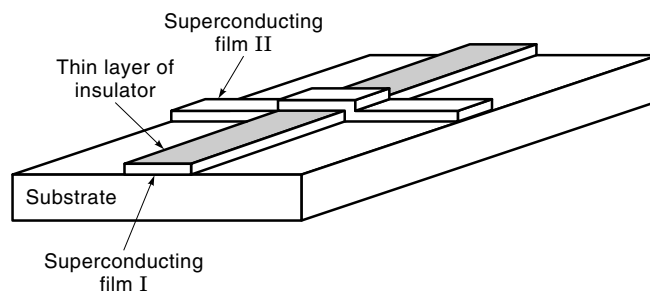


Figure 21. The magnetic thin films are made from deposited thin films from amorphous alloys such as gallium. This figure illustrates thin-film technology to form a Josephson. The deposition of thin films can be done by thermal evaporation, electroplating, sputter deposition, or chemical methods.

The magnetic thin films find extensive applications in memory devices where high density and good sensitivities are required. In such applications, the magnetic properties of the coating are determined by the magnetic properties of the particles that can be controlled before coating. The choice of available materials for this purpose is extremely large. Thin-film technology is also developed in magneto-optics applications where erasable optical media for high-density magnetic storage is possible. The miniature magnetoresistive sensors for magnetic recording and pick-up heads, the Hall-effect sensors, and other magnetic semiconductors make use of thin-film technology extensively.

AMORPHOUS MAGNETIC MATERIALS

The amorphous magnetic materials can be classified as amorphous alloys (Fe, Co, Ni), amorphous rare earths, and amorphous superconductors. Amorphous alloys have good soft magnetic materials and are extensively used in magnetic heads. They have high-saturation magnetization, high permeability, and flat high-frequency dependence. Amorphous alloys are produced in the form of $20 \mu\text{m}$ to $50 \mu\text{m}$ thick ribbons using rapid solidification methods. Nevertheless, thermal stability of these alloys is the major drawback preventing the wider applications. Many of the amorphous metallic films are based on rare earth 3D transition metal alloys, such as Gd_xCo_x . They are used in bubble domain devices.

On the other hand, the amorphous superconductors are a class of superconducting materials such as bismuth and gallium. They are manufactured in the form of powders, ribbons, or flakes by using evaporation methods, ion mixing or ion implantation, or liquid quenching. The vapor deposition technique is used in semiconductor-type sensors. These materials are used in high field magnets, memory devices, and other computer applications.

SHIELDING AND SENSITIVITY TO ELECTROMAGNETIC INTERFERENCE

Magnetic fields are produced by currents in wires and more strongly by the coils. The fields produced by coils are important as a result of magnetic coupling, particularly when there are two or more coils in the circuit. The magnetic coupling between coils may be controlled by large spacing between

coils, the orientation of coils, the shape of the coils, and shielding.

Inductive sensors come in different shapes and sizes. Even though some sensors have closed cores such as toroidal shapes, others have open cores and air gaps between cores and coils. Closed cores may have practically zero external fields, except small leakage fluxes. Even if the sensors do not have closed cores, most variable-inductor sensors have a rather limited external field, as a result of two neighboring sets of coils connected in opposite directions, thus minimizing the external fields. Because the inductive sensors are made from closed conductors, a current will flow, if the conductor moves in a magnetic field. Alternatively, a magnetic change produces current in stationary closed conductor. Unless adequate measures are taken, there may be external magnetic fields linking (interference) with the sensor coils, thus producing currents and unwanted responses.

Because of inherent operations, inductive sensors are designed to have a high sensitivity to magnetic flux changes. External electromagnetic interference and external fields can affect the performance of the sensors seriously. It is known that moderate magnetic fields are found near power transformers, electrical motors, and power lines. These small fields produce current in the inductive sensors elements. One way of eliminating external effects is accomplished by magnetic shielding of the sensors and by grounding appropriately. In magnetic shielding, one or more shells of high-permeability magnetic materials surround the part to be shielded. Multiple shells may be used to obtain a very complete shielding. Ends of each individual shell are separated by insulation so that the shell does not act as a single shorted turn, thus accommodating high current flows. Similarly, in the case of multiple shielding, shells are isolated from each other by proper insulation.

Alternating magnetic fields are also screened by interposing highly conductive metal sheets such as copper or aluminium on the path of the magnetic flux. The eddy currents induced in the shield give a counter mmf that tends to cancel the interfering magnetic field. This type of shielding is particularly effective at high frequencies. Nevertheless, appropriate grounding must be observed.

In many inductive sensors, stray capacitances may be a problem, especially at null position of the moving core. If the capacitive effect is greater than a certain value, say 1% of the full-scale output, this effect may be reduced by the use of a center-tapped supply and appropriate grounding.

BIBLIOGRAPHY

1. J. P. Bentley, *Principles of Measurement Systems*, 2nd ed., Burnt Mills, UK: Longman Scientific and Technical, 1988.
2. E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
3. J. P. Holman, *Experimental Methods for Engineers*, 5th ed., New York: McGraw-Hill, 1989.
4. J. E. Lenz, A review of magnetic sensors, *Proc. IEE*, **78** (6): 973–989, 1990.
5. W. Gopel, J. Hesse, and J. N. Zemel, *Sensors—A Comprehensive Survey*, Weinheim, Germany: WCH, 1989.
6. J. Evetts, *Concise Encyclopedia of Magnetic and Superconducting Materials*, New York: Pergamon, 1992.

HALIT EREN
Curtin University of Technology

HALL EFFECT TRANSDUCERS

Hall effect transducers belong to a family of semiconductors that produce an output voltage proportional to a magnetic field or magnetic induction. Initially, the Hall effect was a curiosity comparable to the Seebeck or Peltier effects, but soon it was realized that it could be used for the study of conduction mechanisms in semiconductors. From 1960 on, the first technical Hall effect applications were introduced—for example, the magnetic field measurement probe, magnetic multiplier, and current-to-voltage transducer. At present, a range of powerful Hall integrated circuits are available for magnetic field measurement purposes, the position control of objects, and angle position of wheels. The latest development was the discovery of the quantum Hall effect. This currently has no technical applications, but it has permitted electrical metrology to make considerable advances.

In 1879, E. H. Hall made the remarkable discovery (1) that a magnetic field perpendicular to a conducting foil deviates the charge carriers in a conductor in such a way that a voltage is generated transversally to the foil. The use and exploitation of this effect was, however, severely restricted to research purposes due to the shortage of good materials. Favorite materials during the exploration period were bismuth and even germanium, but the stability obtained with these elements was very disappointing due to their temperature characteristics. With the expansion of research on semiconductors after World War II, Hall measurements were refined with success for the determination of resistivity, carrier concentration, carrier type, and mobility in semiconductors. This research led to still better semiconductor materials; soon, the first industrial instruments for magnetic field measurements with sensitive calibrated Hall effect sensors were introduced on the market.

With the development of integrated circuits (IC) the applications domain of Hall devices was further extended because it became possible to integrate an amplifier with the Hall sensor in one unit. Such an IC can be used for several purposes: position detection, measurement of current, and electronic compass. In 1980, the German scientist Klaus von Klitzing (2) discovered a new effect that was strongly related to the classical Hall effect: the quantum Hall effect. By studying a two-dimensional electron gas at low temperatures in the presence of a very strong magnetic field (14.6 T), it was found that the ratio of the Hall voltage to the device current depends only on the Planck constant h (6.626×10^{-34} Js) and the elementary charge e of the electron (1.6021×10^{-19} C). The von Klitzing effect was very important to electrical metrology because it became possible to replace the material ohm standard in the same way as the volt standard was replaced by the Josephson junction.

HALL EFFECT

For the derivation of the classical Hall effect (3), a rectangularly conducting sample (Fig. 1) with width w , length L , and thickness t is considered. At the four side planes, the contacts 1, 2, . . . 6 have been added for the connection of a current source I (1, 4) and for potential measurements (2, 3, 5, 6). It is supposed that the width of the contacts is small in comparison with the distance l between (2, 3) and (5, 6). Also, the distances l_1 and l_2 are much larger than the width w . The sample is placed in a magnetic field with induction B perpendicular to its largest

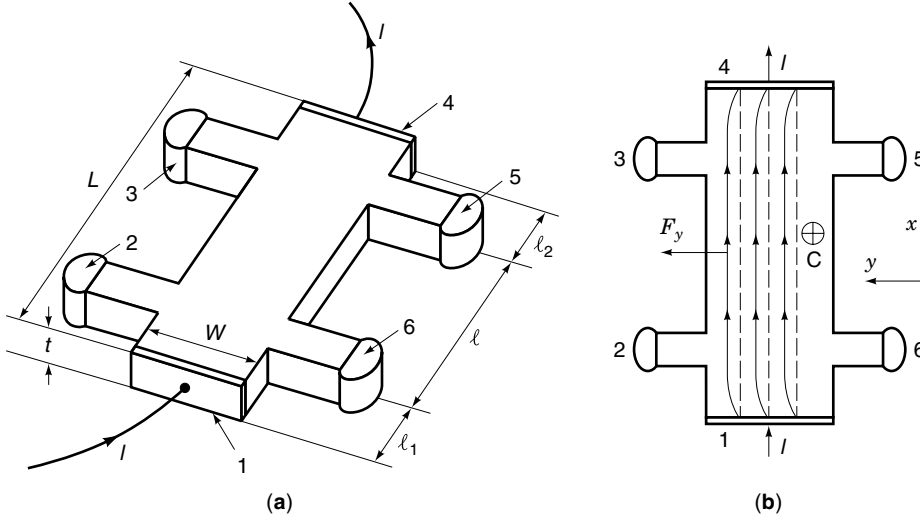


Figure 1. (a) Hall bridge-type sample with four voltage contacts (2, 3, 5, 6) and current contacts (1, 4); (b) deviation of current lines due to an applied B -field. The contacts 1 and 4 are supposed ideal, and ℓ_1 , ℓ_2 , and ℓ are much larger than the width W of the sample. The carrier density at the left and right side will differ due to the Lorentz force F_y .

plane. Figure 1(b) shows that this plane is associated with a rectangular coordinate system with the x -axis in parallel with the length axis of the sample. As long as $B = 0$ in this system, the charge carriers in the sample move along the dotted straight lines between contacts 1 and 4.

The particular shape has the advantage that it can also be used for the determination of the classic resistivity of the sample material. The resistivity ρ_{xx} in the x -direction of the sample is determined with $B = 0$, by measuring the potential difference V_{23} (or V_{56}) between the contacts (2, 3) or (5, 6). It follows that

$$\rho_{xx} = \frac{V_{23}}{I} (wt/l)$$

If the contacts are symmetrically placed around the length axis of the sample, then we expect the voltage between contacts 2 and 6 (or 5 and 6) to be zero and in this case we also have $V_{23} = V_{56}$.

In the presence of a B -field, the moving charge carriers deviate from the straight lines due to the Lorentz force that the field exercises on the carriers. If the current flows in the x -direction, the carriers with an elementary charge q have a velocity v_x , and the corresponding force will be in the y -direction. The absolute value F_y of the Lorentz force is:

$$F_y = Bv_x q$$

As a result, the carrier concentration will be somewhat higher at the left-hand side than at the right-hand side of the sample [Fig. 1(b)]. It means that internally an electric field E_y counteracting the effect of the Lorentz force is generated. This field exercises a repelling Coulomb force qE_y on the carriers and in a steady-state situation both forces are equal and thus

$$E_y = Bv_x$$

The carrier concentration n (number of carriers per m^3) and the velocity v_x are related to the device current I , because the amount of charge displaced in 1 s is equal to the product $wt n q v_x$.

Therefore we have

$$v_x = I / wt n q$$

Because the field E_y is constant for all points between contacts (2, 6) or (3, 5), it follows immediately that the voltage V_H between (3, 5) or (2, 6) is equal to $w \times E_y$

$$V_H = BI / t n q = R_H BI / t, \text{ with } R_H = 1 / n q \quad (1)$$

The constant R_H is the Hall coefficient (independent of the width w), and it has the dimension $\text{m}^3 \text{C}^{-1}$. For the sensitivity of the device to be large, it is thus required that the sample is thin and the carrier concentration not too large. If these conditions are fulfilled, the mobility of the carriers (proportional to the carrier velocity) has to be large in order to give a reasonable conduction in the x -direction. This explains why in ordinary conductors the Hall effect is extremely small. From the foregoing derivation we can deduce some important facts.

1. Measurement of the Hall voltage permits determination of the average polarity of the charge carriers. It was indeed only supposed that the direction of the current was known, but this current could as well be established with holes or electrons depending on the type of semiconductor used. In the first case the left-hand side of the sample will have an excess positive charge and point 2 will become positive with respect to 6 (or 3 to 5). In the second case, 2 will be negative with respect to 6. For extrinsic p -type materials we find from R_H the carrier charge density (in C/m^3) $nq = 1/R_H$ and for the n -type this becomes $-1/R_H$.

2. From Eq. (1) of the Hall voltage we obtain the Hall resistance $R_{xy} = V_H / I$

$$R_{xy} = R_H B / t$$

This quantity increases linearly with the applied field B . The Hall effect device is also called a Hall transducer because it converts the value of the magnetic field into a voltage.

3. It can be seen that the current lines in the sample are slightly longer when the field is applied than when it is not. It can therefore be expected that the longitudinal resistivity

ρ_{xx} increases when the field increases; this is the magnetoresistance effect. In the Hall sample, however, the magnetoresistance effect is of minor importance because the ratio $L:w$ is large.

Until now it has been supposed that the contact pairs are symmetrically placed with respect to the sample. If this is true, no Hall voltage can be generated between the contact pairs when the field is zero. In practice there always remains a slight asymmetry and, because of the finite resistivity in the x -direction, a small offset voltage V_{os} will add to the Hall voltage. This offset voltage can only be eliminated by reversing the direction of B or by rotating the Hall device over 180° around its x -axis and combining the two resulting Hall voltages. The two voltages corresponding to this procedure are, respectively, $V_1 = V_{os} + V_H$ and $V_2 = V_{os} - V_H$. The Hall voltage thus is found from $(V_1 - V_2)/2$. For ac measurements, the offset can be eliminated by blocking the dc component of the output voltage of the Hall device.

QUANTUM HALL EFFECT

In order to observe the quantization of the Hall resistance, several conditions have to be fulfilled: the charge carriers have to behave like a two-dimensional electron gas (2DEG), temperature has to be very low (e.g., below 4.2 K) and a very strong magnetic induction has to be applied (4,5,6).

A two-dimensional electron gas is the special condition that occurs in strong inversion or accumulation layers at planar semiconductor-semiconductor or semiconductor-insulator interfaces. In the high electric field at the interface, the electrons are only allowed to move freely in a very thin layer in parallel with the interface. When a magnetic field is applied perpendicular to the interface, the electrons will tend to move along circular paths due to the interaction of the field and the electron charge e . The frequency ω_c of this rotation is called the cyclotron frequency and it can in a classic way be derived from e , B and the cyclotron effective mass m_e by equating the centrifugal force and the centripetal force due to the B -field. From this it follows that

$$\omega_c = eB/m_e$$

Under normal conditions the electrons are scattered before a full revolution occurs. However, if the electron mobility is very high, a number of revolutions can exist, and the effect is that an angular momentum quantization occurs that influences the density states of the 2DEG. The net effect is that the resistivity ρ_{xx} shows an oscillating behavior as a function of the magnetic induction (Shubnikov-Haas effect). A necessary condition is that the product of the cyclotron frequency and the scattering time τ of the electrons is much larger than one: $\omega_c \tau \gg 1$. Because the scattering increases with temperature and ω_c is proportional to B , the experiments have to be performed at low temperatures (<40 K) and in the presence of a high field (e.g., 15 T) in order to observe quantization. Figure 2 shows the resulting Hall resistance R_{xy} and the magnetoresistivity ρ_{xx} as a function of B under these conditions. As derived in the preceding, the Hall resistance increases linearly with B as long as the temperature is high, for example, above 40 K, and the magnetoresistance also increases with B . If temperature is low enough (about 1 K), ρ_{xx} becomes very small for certain

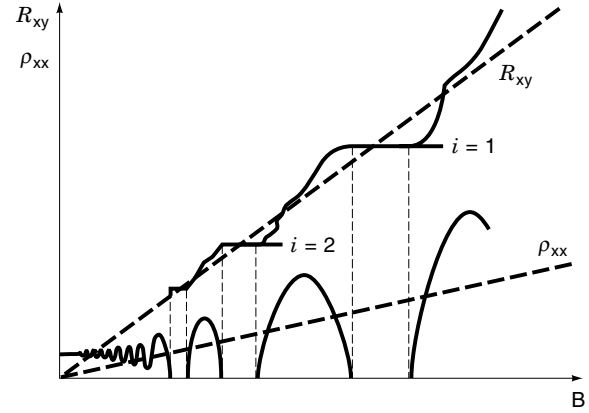


Figure 2. Hall resistance R_{xy} and longitudinal magnetoresistivity ρ_{xx} on an arbitrary scale and as a function of the induction B . The dotted lines correspond to temperatures above 44 K while the full lines are for less than 4.2 K. The Hall resistance shows quantized plateaus in this case.

ranges of B and the Hall resistance R_H becomes quantized following the relation:

$$R_H(i) = h/e^2, \quad i = 1, 2, 3, \dots \quad (2)$$

In the special case $i = 1$, the Hall resistance becomes the von Klitzing constant R_K

$$R_K = R_H(1) \quad (3)$$

For practical applications, the 2D gas can be implemented in Si-metal-oxide semiconductor field-effect transistors (MOSFET) or in a heterostructure with layers of GaAs and AlGaAs.

Because the quantified Hall effect is used in metrology, scientists have tried to detect deviations from the theoretical behavior by experiments or by extending the models for the samples in order to see whether or not $R_H(i)$ is universally valid.

Dependence on the Semiconductor Material

As the 2D gas can be implemented in MOSFET structures as well as in AlGaAs/GaAs structures, both materials have been carefully investigated (7,8). Measurements have confirmed that the value of R_H is with certainty independent of the material. The precision obtained with these measurements was at a relative error level 3×10^{-10} . A small number of researchers have found anomalies, but it has been shown that these can always be attributed to bad samples (9). It was possible to develop a test for the samples, in order to see in advance if the correct conditions for observing the quantized Hall effect are fulfilled.

Dependence on the Sample

Several heterostructure samples of the same material but with very different carrier mobilities have been studied (8) and it appears that the differences for R_H all fall within the uncertainty range of the measurement setup. On the other hand, the measurements also show that R_H varies as predicted with the integer i .

Width of the Sample

It is predicted by some theories (e.g., 10,11) that there could exist a relation between the width of the sample and R_H . However, studies performed on samples with widths of between 10 μm and 1 mm did not reveal deviations larger than the measurement precision (12).

Contact Resistance

For connection of the Hall sample to the outside world, metallic contacts are required. Good contact resistances are of the order of a few milliohms, but in some samples values in the kilohm range have been observed. If the latter is the case, relative errors of the order 1×10^{-7} can be observed. The effect of bad contacts can not be predicted due to their microscopic nature. The experiments (13,14) show that this behavior is strongly influenced by the device current or temperature. It is estimated that the error caused by a good contact lies in the range 1×10^{-11} . It has to be noted that the errors encountered for samples with bad contacts are not due to the loading of the device by the voltmeter circuit. If the device is bad, there is always a measurable longitudinal voltage V_{xx} associated with it. Finally, the most important electrical attributes of a good quantized Hall resistor can be found in Ref. 15.

APPLICATIONS OF THE CLASSICAL HALL EFFECT

The Hall effect can be used in semiconductor research as well as in technical applications.

Research Applications

The Hall effect is used intensively in semiconductor research for the determination of carrier concentration (n) and mobility (μ) in combination with the determination of conductivity (σ) or resistivity (ρ). In the first instance, it is assumed that the layers under study are uniform. If the magnitude of the carrier charge is q , then Eq. (1) for R_H gives the concentrations for p -type and n -type semiconductors, respectively

$$p = 1/qR_H \text{ and } n = -1/qR_H \quad (4)$$

For the case where two carrier types are involved, the general relation between the conductivity mobilities μ_p , μ_n and the conductance σ is given by

$$\sigma = q(n\mu_n + p\mu_p) \quad (5)$$

and, if $p \gg n$, this reduces to

$$\sigma = pq\mu_p = \mu_p/R_H \quad (6a)$$

For electron carriers the corresponding equation is:

$$\sigma = nq\mu_n = \mu_n/R_H \quad (6b)$$

In principle, it is possible to determine the mobility μ_p from this equation by measuring σ and R_H . Nevertheless, some problems that can arise require corrections.

A first problem is the occurrence of carrier scattering mechanisms that require a correction factor to be introduced when μ_p or μ_n are calculated from Eq. (6). A scattering factor r has

to be introduced in Eq. (6a) and its equivalent Eq. (6b) for the electrons as follows:

$$p = r/qR_H \text{ or } n = r/qR_H \quad (7)$$

Depending on the scattering mechanism, the scattering factor lies in the range 1 . . . 2. Therefore, the values of the conductivity mobilities derived from R_H have to be corrected with r

$$\mu_p = \gamma_H/r \text{ or } \mu_n = \mu_H/r \quad (8)$$

The mobility μ_H is derived from the Hall voltage and σ is the Hall mobility and μ_p and μ_n are the ordinary conductivity mobilities. Although the latter are more important, the Hall mobility is easier to determine and can still give comparative information about the material.

Second, R_H depends also on the B -field and, when both types of carriers are present as well as a B -field, the Hall coefficient becomes rather complex. In Ref. 16 the following equation is given:

$$R_H = r \frac{(p - b^2n) + (\mu_n B)^2(p - n)}{q[(p + bn)^2 + (\mu_n B)^2(p - n)^2]} \quad (9)$$

The coefficient b is equal to the ratio of the mobilities μ_n and μ_p or $b = \mu_n/\mu_p$.

For $B \rightarrow 0$ and $B \rightarrow \infty$, the general expression [Eq. (9)] can be simplified to a low-field and a high-field expression, respectively. Depending on the mobilities it is possible to have semiconductors which obey either the low- or high-field expression, or the general equation [Eq. (9)]. The low-field expression will be valid when $B\mu_n \ll 1$ if $p \ll n$ or $B\mu_p \ll 1$ if $p \ll n$. Likewise, the high-field limit is valid when $B\mu_n \gg 1$ if $p \gg n$ or $B\mu_p \gg 1$ if $p \ll n$.

For some semiconductors the conduction mechanism changes with temperature. For example, in HgCdTe one finds that for $T < 200 \text{ K}$. . . 300 K electrons are the main carriers. At the lower temperatures a mixed carrier conduction occurs and, as a result, R_H eventually can change its sign. Therefore, Hall measurements are always performed at different temperatures and induction in order to extract the necessary information from the measured data. An extra complication is the dependence of r on B .

In order to be able to determine the spatial resistivity and mobility profiles in nonuniform layers, it is necessary to remove small layers from the sample so that Hall coefficients can be measured as a function of thickness. This can be done by etching or, preferably, by anodic oxidation. In another method used for layers on an isolator, a pn -junction, a MOSFET or a moscapacitor is formed at its upper surface. If, for example, a metal gate layer is used, the penetration depth of the space-charge region can be controlled by means of the reverse bias of the junction and accurate measurements of the mobility and carrier concentration are the result.

In Ref. 3 one can find a review of the most common sample shapes used, and the precautions to be taken for the reduction of errors. Especially, if parameter estimation techniques are used the results can deliver much information about the material. Hall measurements on irregular sheets are also possible, provided some conditions are met. The theory for such samples was derived in 1958 by van der Pauw; see Ref. 17.

Technical Applications

Because it became possible to integrate the electronics and the Hall effect transducer on the same substrate, a number of important industrial applications have been demonstrated. The Hall transducer can be used for direct measurement of magnetic induction and power, switching applications, position control, and so on. A Hall device will not be destroyed by strong magnetic fields and when semiconductors of the III-V type (GaAs, InSb) are applied, it is possible to enhance the sensitivity and the allowable temperature range. With ion-implanted GaAs a temperature range of -40 to $+250$ °C is obtained. The advantage of using silicon lies in the fact that the integration of an amplifier with the Hall element is easy. However, sensitivity is lower due to the low mobility of the carriers in silicon.

In combination with magnets fixed on moving parts, the Hall transducer effectively becomes a position sensor. By monitoring the output, one can get an idea about the velocity or the acceleration of a moving object. Such a transducer will never wear out because there is no friction. Aging will also be very low because of the large mean time between failures (MTBF) of an integrated circuit. In contrast with optical devices, the operation of this transducer will not be disturbed by moisture, dust, or dirt except when the dust contains iron particles. The Hall device is also relatively fast: the rise and fall times are of the order $0.2 \mu\text{s}$ and magnetic ac fields with frequencies of up to 100 kHz can be detected. The substrate containing the transducer is either a ceramic material or a ferromagnetic substance. The latter is used if it is necessary to decrease as much as possible the permeability of the magnetic circuit in which the device is mounted.

Linear Hall Effect Transducers. The output of a linear Hall effect device is proportional to the product of the induction B and the device current I ; in this way, it performs an analog multiplication. It is also obvious that such an element can be dedicated to the measurement of magnetic fields if I has a constant value.

Measurement of Magnetic Fields. The instrument using Hall effect sensors for the determination of magnetic induction (18–22) is called a gaussmeter although it should be termed a teslameter, as the unit of induction is tesla (T) in SI units.

From the preceding it follows that the Hall voltage is due to the component of the field normal to the main sheet plane of the device. If the angle between the field B and the normal is α [Fig. 3(a)], then the normal component is $B \cos \alpha$. For maximum

sensitivity it is required that $\alpha = 0$ and if the field has an unknown direction it is necessary to search for maximum output by rotating the device in space. If three separate Hall devices are orthogonally arranged [Fig. 3(b)] so that their main planes coincide with the planes of a rectangular axis system, the outputs will be proportional to the spatial components B_x , B_y and B_z of B . The magnitude and direction of the vector B in space can then be derived from the three outputs. It is, however, required that the probes be carefully calibrated. This can be done by adjusting the individual device currents, or by measuring the three sensitivity factors and storing them in a memory for further use with the software calculations. If large spatial gradients exist in the field, then the device output represents a kind of average. This error can be reduced by using very small elements: mostly it is possible to adapt the magnetic circuit to minimize this effect.

As sensitivity depends on I it seems appropriate to take I as large as possible. However, this also increases both device dissipation and current consumption. Therefore the advantage gained by increasing I is rather marginal and it is better to amplify the Hall voltage electronically if a higher sensitivity is required. For spatial magnetic field measurements it is preferable that the Hall device has a ceramic substrate as this ensures that the field is not disturbed by the device. Due to the large field of applications of gaussmeters, several types of probes with a calibrated built-in Hall element are available on the market. For measurement of fields in air gaps, for example, in electrical machines or relays, thin probes are preferred but these probes are very sensitive to mechanical stress. The Hall device is mounted at the end of a long flat isolating stylus so that less accessible points can be probed as well. The transverse probe is sensitive for fields perpendicular to the stylus plane. The axial probe is cylindrical and the Hall device is mounted perpendicular to the axis of the cylinder; it measures fields parallel to the probe axis. Full-scale sensitivity can be as large as 15 T and rms readings of ac fields are also possible.

Measurement of Electrical Current. Although not strictly necessary, a magnetic circuit is commonly used with the Hall device for current measurement (19,23) purposes [Fig. 4(b)]. This has two advantages: the sensitivity of the measurement is increased, and the position of the current carrying conductor becomes unimportant. The magnetic field at a distance d of a straight conductor [Fig. 4(a)] carrying a current I , has a value $H = I/2\pi d$ and this is position sensitive. A Hall device at this position will give an output proportional to $\mu_0 H$. If the

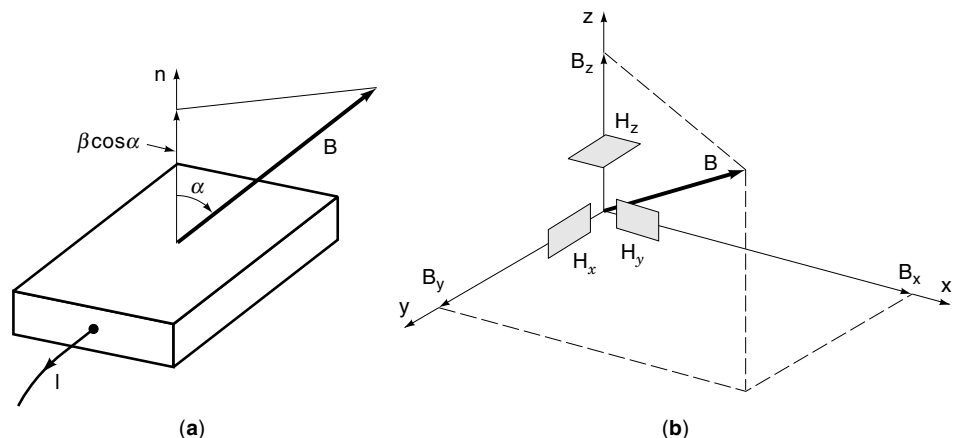


Figure 3. (a) Component normal to Hall device for an arbitrary direction of the field and (b) setup with three elements for the determination of magnitude and direction of field.

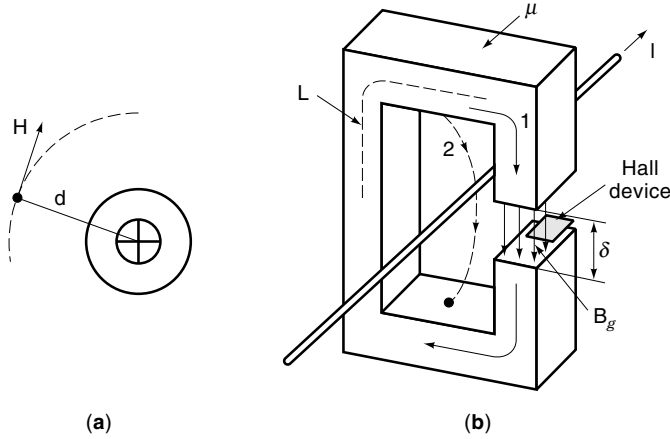


Figure 4. (a) The radial field around a current-carrying conductor depends on the position of the Hall element; (b) magnetic circuit with a Hall device in the air gap. At low saturation all field lines in the core follow the magnetic path (1). When saturation occurs, some of the field lines (2) are lost for the air gap.

current encircles a magnetic circuit, the force lines [e.g., 1 in Fig. 4(b)] will concentrate in the magnetic material (mean length L , magnetic permeability μ) and in the air gap with width δ . The value B_g of the magnetic induction in the air gap becomes, with Ampère's law and because of the constant flux in the core and air gap section,

$$B_g = \mu \mu_0 I / (L + \mu \delta) \quad (10)$$

For $\mu = 10^4$, $L = 0.1$ m, $\delta = 1$ mm (10^{-3} m) the value $\mu \delta = 10$ and L can thus be neglected. It means that in the range where $\mu \delta \gg L$, the gap induction B_g can be approximated by $\mu_0 I / \delta$ and the sensitivity becomes independent of μ and yet can be very high because δ is so small. If the wire encircles the magnetic circuit N times, then B_g also becomes N times larger. The material properties of the core material barely influence the precision of the measurement, but a very stable air-gap width δ is required. Also, when the current increases, the core induction increases and the material has a tendency to saturate. As a result, some of the magnetic field lines can leave the legs of the core and pass through the window area [e.g., line 2 in Fig. 4(b)]. Because a part of the flux is lost, the air-gap induction becomes too low and a nonlinearity is introduced in the transducer characteristic. The existence of an air gap is favorable in this respect because it reduces the core induction at the cost of a lower sensitivity. The design of such a circuit is therefore a trade-off between sensitivity and full-scale behavior. The nonlinearity can remain below $\pm 1\%$ with ordinary core materials such as ferrite. Besides the electrical offset of the Hall device, a magnetic offset is caused by the core material. This is due to the remanent magnetism that remains after a heavy magnetization of the core. The remanent field gives an additional electrical output offset. Reduction of the remanent magnetism is done by choosing a very soft magnetic material. It is not possible to correct for this remanent field using software because its value depends on the history of magnetization and, as precise models for the core behavior are lacking, a precise offset error correction is unpredictable. In Ref. 24, one can find how the problem of the intrinsic Hall plate offset voltage can be alleviated by combining the output of several elements. The

simple circuit can still be very useful if the highest precision is not required because it requires only a Hall IC and a magnetic circuit, and power consumption is very low.

Closed-Loop Hall Effect Current Transducer. The linearity and offset problems can be reduced considerably by applying feedback (25) to the simple circuit (Fig. 5). In this case, the Hall voltage amplifier drives a second winding N_2 coupled with the core and a resistor R . The voltage drop RI_2 over this resistor is taken as the new output voltage. By choosing the correct winding, the amplifier output current I_2 can counteract the effect of the unknown current I_1 . The number of ampere turns generating the field is now equal to the difference of the ampere turns of the current and feedback windings. Equation (10) for the magnetic circuit becomes, with H_1 and H_g , respectively, the magnetic field in the magnetic circuit and in the air gap

$$H_1 L + H_g \delta = H_1 (\delta + L/\gamma) = N_1 I_1 - N_2 I_2 \quad (11)$$

The amplifier output voltage E_0 is proportional to the air-gap induction and with an induction-to-voltage gain K , we obtain

$$E_0 = K \gamma_0 H_g / R$$

As the corresponding current I_2 is E_0/R , the air-gap field becomes

$$H_g = N_1 I_1 / (\delta + L/\gamma + N_2 K \gamma_0 / R)$$

If the amplifier gain K is made very large, it is seen that $H_g \rightarrow 0$ and the field in the air gap disappears due to the feedback. It follows that $N_1 I_1 \rightarrow N_2 I_2$ and

$$V_0 \approx (N_1/N_2) I_1 \quad (12)$$

As is the case in every feedback system, the static error can be neglected if K is high enough. The output V_0 is a very accurate image of I_1 because the winding resistances, leakage inductances, and losses in the core, have no effect on accuracy. If the material is initially sufficiently demagnetized, the permeability will be very high and the magnetic offset error is reduced. In

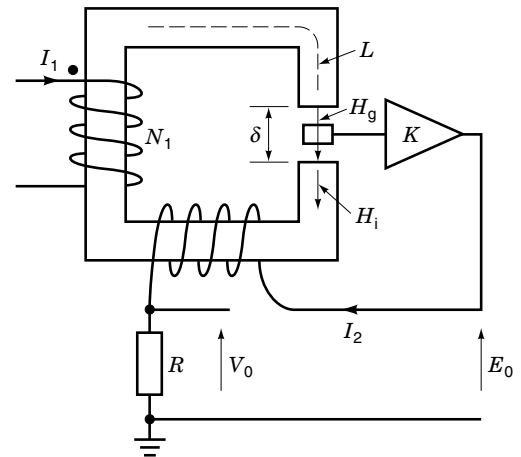


Figure 5. Closed-loop Hall effect device in magnetic circuit. If the loop-gain is very large the primary and secondary ampere turns will cancel and V_0 becomes proportional to I_1 .

contrast with the open-loop system, the length of the gap barely influences the measurement precision; as a result, the temperature behavior is also enhanced. When the current I_1 exceeds the output current capability of the amplifier, for example, under short-circuit conditions, the core is likely to be saturated and, therefore, a means for demagnetizing the core is standard in such an instrument. Demagnetization is obtained by driving temporarily the winding N_2 with an exponentially damped oscillating (e.g., sinusoidal) current. If the initial value of this current drives the core into strong saturation, eventually the core will be demagnetized. For current probe applications the core is divided into two parts, which can be separated in such a way that the conductor can be enclosed by the core window without breaking the electric circuit. Especially, in this case it can be appreciated that the possible instability caused by imperfect core halves contact is eliminated by the feedback. The required current output capability of the amplifier depends entirely on the winding ratio N_2/N_1 and the full-scale current I_1

$$I_2 = I_1 N_1 / N_2 \quad (13)$$

For practical reasons, the winding ratio has to be limited. This means that the amplifier in general will have to deliver a much larger current than in the open-loop system and its power consumption will be large. Another problem is that additional circuitry is required for wideband measurements because the frequency response of the Hall device is limited to about 100 kHz. A solution is to split the circuit into two parts. First, the frequency response of the Hall amplifier will deliberately be limited to a few kilohertz by first-order lowpass filtering of the output V_0 . The purpose is to reduce the effect of the poles in the open-loop transfer of the Hall device. Next, for the higher-frequency part of the spectrum, the two windings can be made to operate as a current transformer. The signal obtained from this current transformer is also first-order high-passed, with the same 3 dB cut-off frequency and transfer gain as for the Hall device. By combining the two filter outputs a flat frequency characteristic is then obtained. Typical probes show a flat frequency response from dc to 20 or 100 MHz and have full-scale ranges from 1–10 A with a full-scale output voltage of 50–200 mV.

Multiplier Applications of the Hall Effective Device. Although ac power can be measured by sampling the current and the voltage of a consumer, in general, the potential drop, and, hence, the dissipation in the current shunt, are relatively large. Further, it is difficult to protect the electronics unconditionally against a short-circuit at the consumer side. On the other hand, when the load current flows through the magnetizing coil of Fig. 4, a short will not harm the Hall device or the magnetizing coil because the coil is very rugged. As suggested by Eq. (1) the Hall device output is proportional to the product of B and I . In fact, one of the first applications of the device was as an analog multiplier (19,20). Although this application has been superseded completely by the use of the Gilbert cell analog multiplier ICs, the device still has attractive features for power applications. For example, for determination of the power an ac load consumes, the product of the momentary voltage $v(t)$ and current $i(t)$ has to be calculated. The former current-measuring circuit can easily be adapted for this purpose. It suffices to vary the Hall current I proportional to the voltage $v(t)$ so that the output voltage $v_0(t)$ of the device becomes proportional to the momentary power. By filtering this voltage, a measure for the average

power is obtained. If $v_0(t)$ is fed to a voltage-to-frequency converter followed by a counter, the counter output will represent the energy consumed by the load. For household applications the preferable counter is a mechanical one because it gives a clear indication and holds the current value, even during a power loss. The advantage of using the Hall device lies in the fact that the current scale for the instrument can be adapted through a change of the winding N_1 in the same way as a classical Watt-hour meter with a rotating disk induction motor.

Clinical Hall Effect Devices as Position Sensors. A number of industrial applications require knowledge of the position of a mechanical part; the Hall effect ICs have been developed especially for this task (22). For example, in brushless dc motors the angular position of the rotor has to be known in order to activate the drive electronics of the field coils. More recently, the Hall sensor has been applied in magnetic bearings (26) for magnetic levitation purposes.

If there is no magnetic field available, the Hall devices will be used in combination with permanent magnets (27). Suppose the rotation angle α of a permanent magnet has to be determined [Fig. 6(a)]. If a linear sensor H_1 is positioned in a zone between the pole shoes, where the field of a rotating magnet is homogeneous and has a value B the output V_{H1} will be proportional to the cosine of the angle of rotation α on the condition that the rotation axis coincides with the length-axis of the Hall

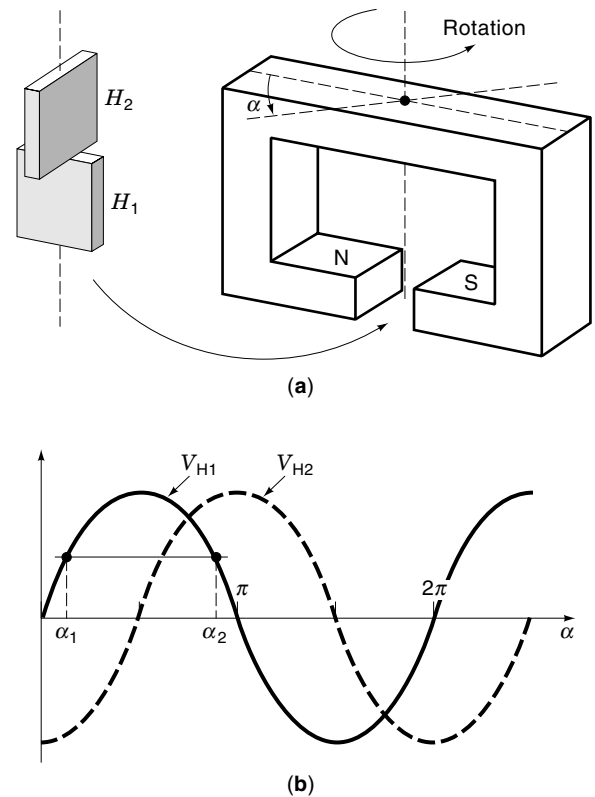


Figure 6. (a) Hall device used as an angle sensor: a magnet is fixed on a turning axis and the Hall device is mounted between the pole shoes of the magnet. The output of V_{H1} is proportional to $\sin \alpha$ because for $\alpha = 0$ its plane coincides with the plane of the magnet. (b) A second device is necessary in order to remove all possible ambiguity from determination of α from the Hall voltages.

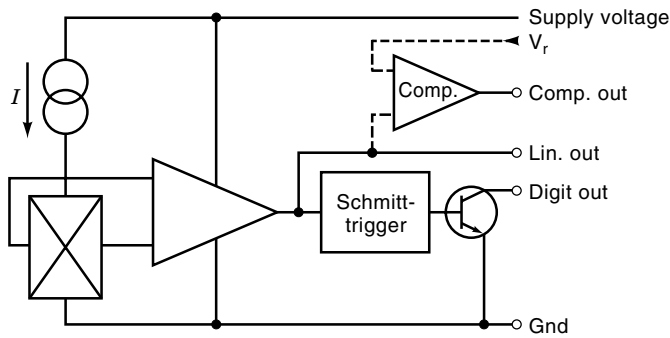


Figure 7. Internal circuit of a Hall IC with Schmitt-trigger and linear output. A comparator output is also useful for limit detection.

sensor plane. From output V_{H1} it is, however, impossible to obtain α , as there are two angle solutions for every voltage [α_1 and α_2 in Fig. 6(b)]. This ambiguity can be resolved by placing a second transducer H_2 (output V_{H2}) perpendicular to H_1 . From V_{H1} and V_{H2} , and taking into account the signs of both outputs, the position of the magnet can be determined in the whole α -range $(0, 2\pi)$.

A digital output can be obtained by including a Schmitt-trigger circuit (Fig. 7), driven by the analog output. Mostly, an open-collector transistor forms the output as this permits adaptation of the IC to the logic level of the system connected to it. Sometimes a small magnet is mounted close to the sensor in order to provide for a magnetic bias. The output can be made to react in two ways as shown in Fig. 8. In Fig. 8(a), the purpose is to discriminate between the presence or absence of a magnetic pole. Without field the output is at the logical level "1" (line 1). When the magnet approaches the device its analog output increases. At the trip point of the Schmitt-trigger corresponding to a field B_1 , the digital output falls to the logical level "0" (line 2). In a third phase, the magnet is removed from the device but output remains low (line 3) as long as B lies above the lower trip point B_2 . In Fig. 8(b), the lower trip point corresponds to a negative field, that is, the other pole of the magnet. Operated in these two ways, the Hall device acts like a switch. The hysteresis of the Schmitt-trigger reduces the effect of bouncing and noise. There exist three common ways for detecting the position of a moving magnet: Fig. 9(a–c). It is assumed that the magnet moves along the indicated x -axis in all three cases. In Fig. 9(a), the magnet moves in a direction perpendicular to the Hall device plane. The analog output falls continuously with increasing displacement x , and a digital action is possible if the circuit is followed by a comparator. In Fig. 9(a), the compara-

tor switches from low to high if the Hall voltage exceeds reference voltage V_d . The main drawback is the poor linearity of the analog output with respect to the movement of the magnet. In Fig. 9(b), the movement is along an axis in parallel with the device plane and the south pole is near the magnet. The output will be at maximum when the south pole is positioned at $x = 0$. The system can be used as a mechanical window detector: if the magnet is positioned within the range $(-x_1, +x_1)$, the comparator output is low and vice versa. The sensitivity is controlled by adapting the distance d between the magnet and the device. Linearity can be acceptable in a small part of the displacement curve around the point of the curve where the second derivative with respect to B vanishes. In the third case of Fig. 9(c), the magnet is turned over 90° and a linear device sensitive for negative and positive fields is used. Sensitivity can be controlled as in the former case and linearity can be extremely good but the sensitivity control will also influence the linearity. In the setup of Fig. 9(c), sometimes both magnet and Hall device are fixed but a high permeability screen moving between them can interrupt the field. Such a system can detect the teeth of a cog-wheel or it can replace the mechanical breaker of the ignition in a car.

Quantum Hall Effect in Electrical Metrology. In the International System of Units (SI) the four base units are the meter (length), kilogram (mass), second (time) and the ampere (current). The ampere is the current that generates a force of 1 N per length between a pair of very thin and infinitely long conductors. This definition is artificial and of little practical value, but it ensures that the mechanical and electrical units of force are the same. Apart from this, some of the mechanical units derived are introduced for practical purposes, for example, the Henry, coulomb. These secondary quantities can always be expressed in the base units. The magnitude of magnetic permeability in the SI system is fixed to $4\pi \times 10^{-7}$ H/m. The second is defined by means of the period of a well-defined transition in Cesium 133. The velocity of light (c) has a fixed value and the meter is the distance traveled in a vacuum by light during $1/299.792.458$ s. This makes the use of the material standard meter obsolete. All electrical units can be defined in this system if two of them, for example, the volt and the ampere, are related to the mechanical units meter, second, kilogram. This comparison can be implemented via a kelvin current or volt balance. In practice the ampere has been replaced by the ohm as the latter is related to the ampere by Ohm's law.

Because the realization of physical standards requires complex instruments, national metrology laboratories still possess

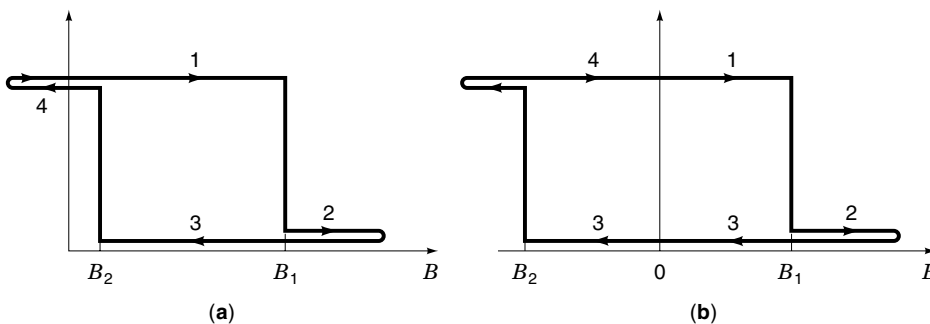


Figure 8. (a) Schmitt-trigger output for a unipolar sensitive Hall IC; (b) output for bipolar sensitive IC.

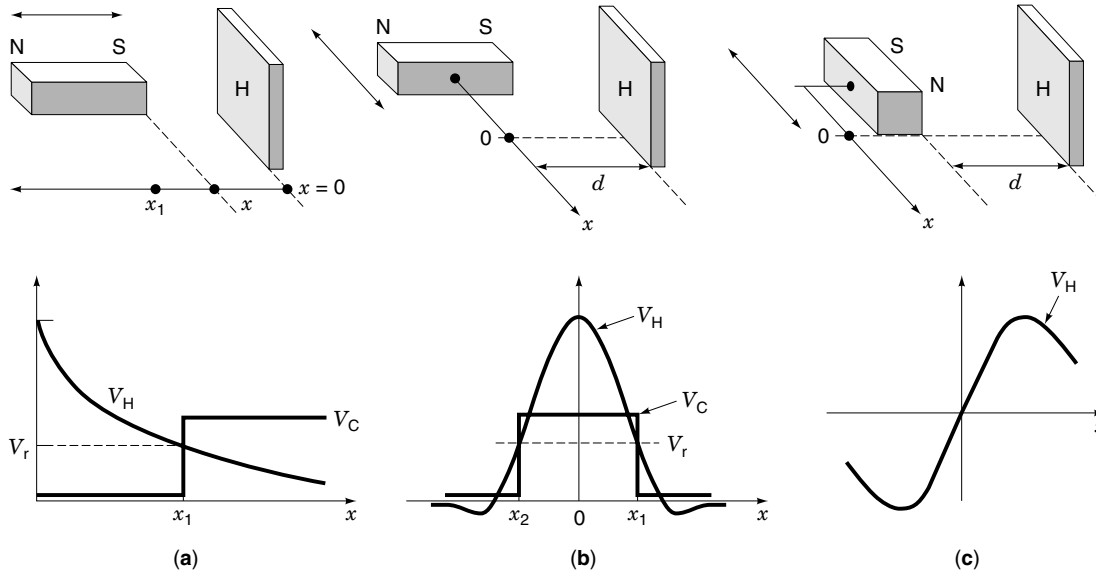


Figure 9. Three ways to detect a magnet movement: (a) magnet axis and movement normal to Hall plane and resulting linear (V_H) and comparator (V_C) output voltages; (b) movement axis parallel with Hall plane gives a window function for V_C ; (c) output is linear over a large displacement range.

material standards such as Weston cells and $1\ \Omega$ resistors for their electrical measurements. These standards are considered to be a kind of flywheel in which an image of the physical unit is conserved for a limited time. By regularly comparing the physical standards with the material standards, the correction for each of these is determined. The national volt (V_{LAB}) and ohm (Ω_{LAB}) of a nation is then defined from the average values of these material standards. Before the invention of the Josephson junction and the quantified Hall effect, it was imperative to perform regularly cumbersome international comparisons in order to determine the corrections between the national standards V_{LAB} and Ω_{LAB} of the different countries.

In 1956, Thomson and Lampard (28,29) derived a striking theorem concerning a special cylindrical four-electrode capacitor structure. It was proved that for the structure of Fig. 10(a), independently of the shape of the cross section of the electrodes A, B, C, and D, the capacities C_1 and C_2 per unit length (1 m) of opposite electrodes obey the expression

$$\exp(-\pi C_1 \geq 0) + \exp(-\pi C_2 \geq 0) = 1 \quad (14)$$

with ϵ_0 the permittivity of the absolute vacuum. The value of ϵ_0 is known because, in vacuum, the velocity of light is $c = (\mu_0 \epsilon_0)^{-0.5}$ and c as well as μ_0 are fixed. For a symmetrical capacitor Fig. 10(b), $C_1 = C_2 = C$ and from Eq. (14) we obtain $C = 1.953549043\ \text{pF/m}$. In practical applications, a capacitor with finite length is used and this length is varied by inserting a guard tube S in the space between the four electrodes (30–32). By careful construction it is possible to reduce the effect of the fringing fields at both ends of the capacitor. In this way, the physical length of the capacitor is changed by displacing the guard tube from l_1 to l_2 and the corresponding variation of capacity is $1.953549043 (l_2 - l_1)\ \text{pF}$. The capacity variation in this system is, therefore, only determined by one dimension; the displacement $l_2 - l_1$ of the guard tube can be measured very accurately with a laser interferometer. The very complicated setup for this system has been called the calculable capacitor because of its capacitance formula. By employing a chain of ac comparator bridges the material ohm standard can be related to the impedance of the calculable capacitor, and thus with the mechanical units.

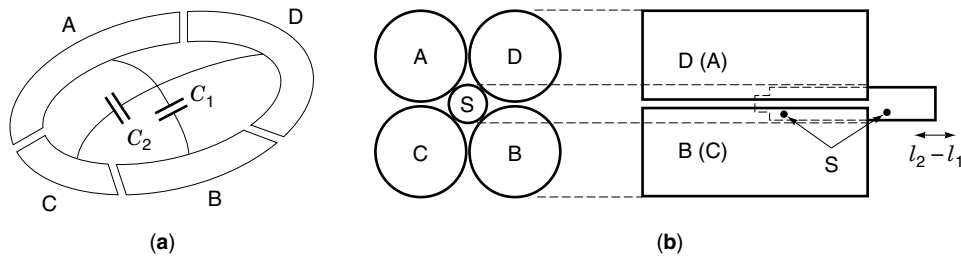


Figure 10. (a) Calculable capacitor structure with four electrodes; (b) simplified system with four cylindrical electrodes A, B, C and D. The effective length is varied by moving a screen S in the opening between the cylinders. Not shown are a similar but fixed screen at the left side, the screen surrounding the tubes, interferometer, and trimming screws. The smaller part of the screen is the “spike.” It can be shown that this spike reduces the effects of irregularities in the cross section of the capacitor (Ref. 32).

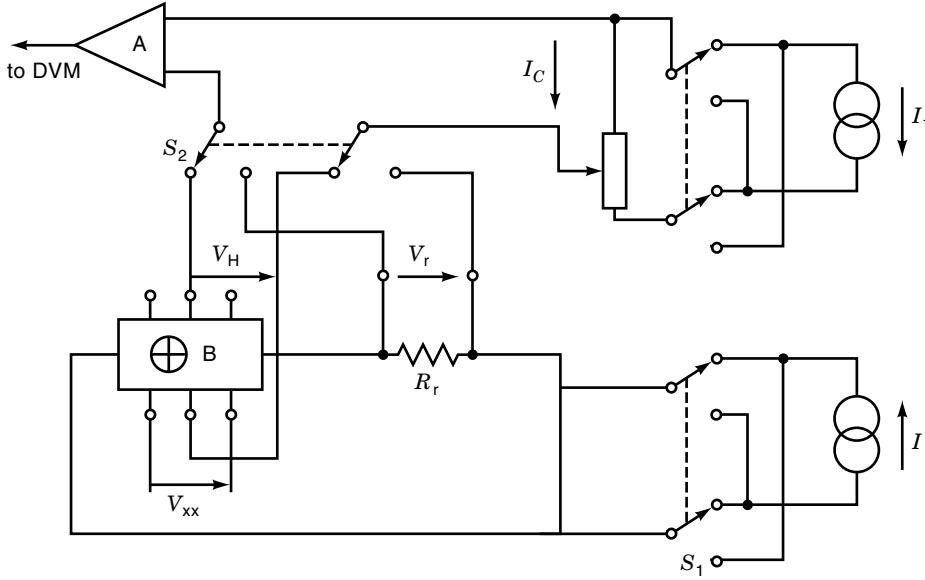


Figure 11. Potentiometric setup for comparing the quantified Hall resistance with the laboratory standard resistor.

The quantum Hall effect has opened the door to a new material ohm standard, which is solely dependent on the universal physical constants e and h . On the other hand, it is generally accepted (33) that the von Klitzing constant R_K , that is, the value of R_H for $i = 1$, can also be calculated from μ_0 , c and the dimensionless fine structure constant α . This constant can be derived via complex quantum electrodynamic calculations from measurements of the anomalous magnetic moment of the electron. The relation between μ_0 , c , α , h and e is given by

$$R_K = R_H(i = 1) = h/e^2 = (\gamma_0 c/2)\alpha^{-1} \quad (15)$$

with a value $\alpha = 1/137.0359898$. Since μ_0 and c have well-defined values, the accuracy of R_K and α is the same in Eq. (15).

As it became clear that the quantum Hall effect could be used for definition of the unit of resistance, metrologists succeeded in refining the measurement procedures so that R_K could be expressed in Ω_{LAB} . It also became possible to compare the existing material ohm standards with a relative error of 10^{-8} . At that moment the reproducibility of the quantum Hall measurements was better than the implementation of the material ohm in the International System of Units (SI).

In 1990, the Comité Consultatif d'Electricité (CCE) decided, based on an analysis of all the known international comparisons of the material resistance standards by the Bureau International des Poids et Mesures (BIPM), to attach a conventional value to the von Klitzing constant

$$R_{K-90} = 25812.807 \, \Omega$$

The relative uncertainty for this value is 2×10^{-7} and this definition makes it possible to replace the maintained standards with the quantum Hall effect device. At the same time, for the second quantum effect used in metrology, that is, the Josephson effect, a conventional value K_{J-90} has also been assigned to the Josephson constant. The material standards for the volt and ohm are now replaced by the Josephson junction and the quantum Hall device; the outputs of these devices are independent of the site on earth and the shape or materials used for

their construction. Intercomparisons are therefore no longer as critical.

Measurement Setup. Comparisons of the Hall resistance with standard resistors can be performed in different ways: potentiometric or with a cryogenic current comparator bridge (CCC). Figures 11 and 12 show generic and simplified diagrams representative for various measurement setups. In the potentiometric method (Fig. 11) a reference resistor R_r of the same nominal value of $R_H(i)$ is used. As the plateau $i = 1$ is difficult to obtain with reasonable fields, often the integers $i = 2$ or $i = 4$ are preferred. Superconducting magnets that generate an induction of 15 T between its poles are commonly used. The value I of the dc current source is in the $50 \, \mu\text{A}$ range and this alleviates the work of the temperature controller because of the low power dissipation in R_r . Measurement of the voltage V_{xx}

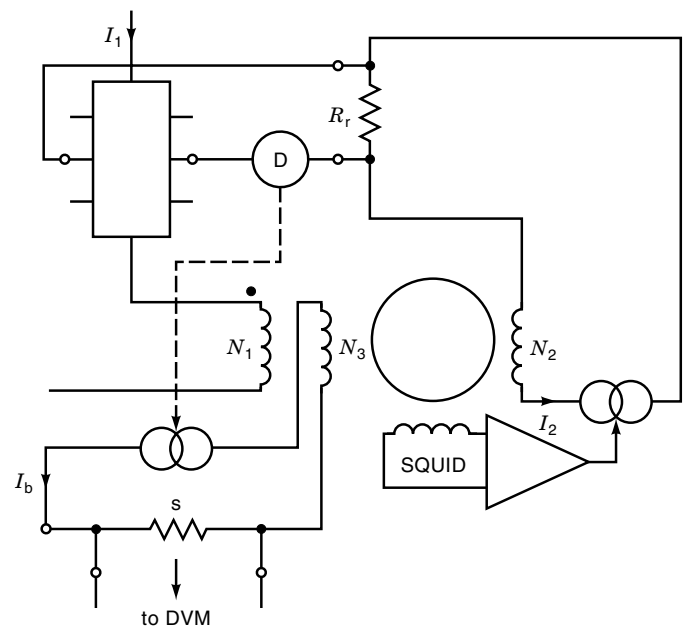


Figure 12. Cryogenic current comparator setup for comparing the quantified Hall resistance with the laboratory standard resistor.

permits inspection if the correct conditions for the occurrence of the quantum Hall effect are met. The Hall plateaus will appear when V_{xx} is very small. A zero detector is lacking and, instead, a high-gain amplifier followed by a digital voltmeter (DVM) is used. A compensating source V_C is derived from a potentiometer fed with a second current source I_1 . By changing the position of S_2 , two voltage differences are measured: $\Delta_1 = V_H - V_C$ and $\Delta_2 = V_r - V_C$. From the Δ -values measured it is possible to calculate the current I and V_C and thus V_H . To eliminate the drifts of the current sources and the thermal emfs it is necessary to perform measurements over a longer period, for example, an hour, and calculate averages. Because all current sources can change direction (by turning over S_1) the effect of thermal voltages can be eliminated. If the CCC is used as shown in the simplified circuit of Fig. 12, it is possible to compare R_H with smaller resistors, for example, 100 Ω because the turns ratio N_2/N_1 can be varied. In Fig. 11, there are two balance conditions to be met: $V_H = R_r I_2$ for the zero detector balance, and $N_1 I_1 = N_2 I_2$ for the zero flux condition in the core. The magnetic flux is measured by the very sensitive superconducting quantum interference device (SQUID), and the output of this device controls the current source I_2 in such a way that the core flux becomes almost zero. If the balance of D is obtained by varying N_1 , then we find with $V_H = R_H I_1$, the quantum Hall resistance $R_H = R_1 N_1/N_2$. In order to increase the precision to the level 10^{-9} or better, a fixed ratio N_1/N_2 is more appropriate. In this case a third current winding N_3 is added to the comparator and a balance current I_b derived from the detector output performs the automatic balancing. This balance current is measured by means of a shunt resistor s and a DVM. The balance current is much smaller than the main current and, therefore, the limited precision of the digital voltmeter does not impair the measurement accuracy.

If the two methods are compared (Ref. 5), it can be calculated for a specific case that the potentiometric method generates more system noise than the CCC. In a typical case the CCC method is therefore almost 45 times faster than the potentiometric method. As a result, the measurement time is reduced to a few minutes and the random uncertainty still is at the level 2×10^{-9} .

BIBLIOGRAPHY

1. E. H. Hall, On a new action of the magnet on electrical currents, *Amer. J. Math.*, **2**: 287–292, 1879.
2. K. von Klitzing, G. Dorda, and M. Pepper, *Physical Rev. Lett.*, **45**: 494–497, 1980.
3. D. K. Schroder, *Semiconductor Device Characterization*, New York: Wiley, 1990.
4. G. Landwehr, The discovery of the quantum Hall effect, *Metrologia*, **22**: 118–127, 1986.
5. A. Hartland, The quantum Hall effect and the resistance standards, *Metrologia*, **29**: 175–190, 1992.
6. B. Jeckelmann and B. Jeanneret, Application de l'effet Hall quantique à la métrologie, Swiss Federal Office of Metrology info (OFMET info), 4 (2): 8–14, 1997.
7. B. Jeanneret et al., Comparison of the quantized Hall resistance in different GaAs/Al_xGa_{1-x}As heterostructures, *IEEE Trans. Instrum. Meas.*, **40**: 231–233, 1991.
8. B. Jeckelmann, A. D. Inglis, and B. Jeanneret, Material, device and step independence of the quantized Hall resistance, *IEEE Trans. Instrum. Meas.*, **44**: 269–272, 1995.
9. B. Jeckelmann, A. D. Inglis, and B. Jeanneret, Are anomalous values of the quantized Hall resistance really anomalous? *Metrologia*, **33**: 499–502, 1996.
10. A. H. MacDonald and P. Streda, Quantized Hall effect and edge currents, *Phys. Rev. B*, **29**: 1616–1619, 1984.
11. B. Shapiro, Finite-size corrections in the quantum Hall effect, *J. Phys. C*, **19**: 4709–4721, 1986.
12. B. Jeanneret et al., Influence of the device-width on the accuracy of quantization in the integer quantum Hall effect, *IEEE Trans. Instrum. Meas.*, **44**: 254–257, 1995.
13. K. C. Lee, Bonding wires to quantized Hall resistors, *IEEE Trans. Instrum. Meas.*, **44**: 249–253, 1995.
14. B. Jeanneret and B. Jeckelmann, Influence of the voltage contacts on the four-terminal quantized Hall resistance in the nonlinear regime, *IEEE Trans. Instrum. Meas.*, **44**: 276–279, 1995.
15. A. D. Inglis and I. Minowa, Fabrication of precision quantized Hall devices, *IEEE Trans. Instrum. Meas.*, **46**: 205–207, 1997.
16. R. A. Smith, *Semiconductors*, Chapt. 5, Cambridge, UK: Cambridge University Press, 1959.
17. L. J. van der Pauw, A method of measuring specific resistivity and Hall effect of discs of arbitrary shape, *Phil. Res. Rep.*, **13**: 1–9, Feb. 1958.
18. C. L. Chien and C. R. Westgate, *The Hall Effect and Its Applications*, New York: Plenum, 1980.
19. E. Schwaibold, Der Halleffekt und seine technische Anwendung, *Archiv für Technisches Messen*: V943–2/3, 1956.
20. E. Schwaibold, *Halbleiter-Bauelemente in der Messtechnik*, Berlin: VDE Verlag, 1961, pp. 201–213.
21. J. Janicke, Putting the Hall effect Gaussmeter to work, *Int. Electron.*: 33–37, February 1965.
22. H. P. Baltes and R. S. Popovic, Integrated semiconductor magnetic field sensors, *Proc. IEEE*, **74**: 1107–1132, 1986.
23. F. J. Oliver, *Practical Instrumentation Transducers*, New York: Hayden Book Company, pp. 46, 106, 138, 249, 287–288, 1971.
24. A. Bilotti, G. Monreal, and R. Vig, Monolithic magnetic Hall sensor using dynamic quadrature offset cancellation, *IEEE J. Solid-State Circuits*, **32**: 829–836, 1997.
25. Z. L. Warsza, Hall effect feedback transducers and their application, *Acta Imeko 1976*, Elsevier, 7th IMEKO Congress, London: 551–559, 10–14 May 1976.
26. J. Boehm, R. Gerber, and N. R. C. Kiley, Sensors for magnetic bearings, *IEEE Trans. Magn.*, **29**: 2962–2964, 1993.
27. J. Spencer and M. Perkins, Hall-effect works in two-state switches, *Electron. Design*: 125–127, July 5, 1980.
28. D. G. Lampard and A. M. Thompson, A new theorem in electrostatics with application to calculable standards of capacitance, *Nature*, **177**: 888, 1956.
29. D. G. Lampard, A new theorem in electrostatics, *J. IEE*, **104C**: 271–280, 1957.
30. W. K. Clothier, A calculable standard of capacitance, *Metrologia*, **1**: 36–35, April 1965.
31. N. Elnékvé, An absolute determination of the ohm based on calculable standard capacitors, *EUROMEAS-77*, Precise Electrical Measurement Conference publication no. 152, *IEE* 1977: 53–57, 5–9 Sept. 1977.
32. G. H. Raynier, NPL calculable capacitor, *IEEE Trans. Instrum. Meas.*, **21**: 361–365, 1972.
33. B. W. Petley, *The fundamental physical constants and the frontier of measurement*, Bristol: Adam Hilger, 1985.

DANIËL H. J. BAERT
University of Ghent

PRESSURE SENSORS

Each physical quantity A can be measured by use of primary methods of measurement or by secondary methods, that is, by measurement of some other quantities W_i in physical relation with quantity A . Quantity A must be measured or must be controlled (regulated), but it is desirable to measure quantities W_i because of their advantages. There are many physical and technical possibilities for designing a specific device giving the relation between A and W_i . Such a device is called a transducer (measurement transducer, because the meaning of the word *transducer* is very wide and can mean any change-processing devices). A transducer is a device that provides a usable output signal in response to a specific measurand (a measurand is the physical quantity being measured). Transducers have their own internal structure. In the simplest case we can imagine some elementary transducers connected in linear order. Based on the physical background we can find a specific physical phenomenon, in our case, pressure that has a known, well-defined, influence on some quantities that characterize a specific body: volume, dimensions (linear and angular displacements, strain), resistance, sound velocity, density, dielectric constant, or other quantities whose variations allow determination of pressure. Let us call that body a *pressure-sensing element*, or *primary pressure transducer*. From a transformation theory point of view, it receives and transforms information by itself. A *pressure sensor* device, see Fig. 1, consists of the pressure-sensing element and at least one basic (secondary) transducer. This transducer converts a sensing-element signal into another quantity, one of the quantities W_i chosen as the most convenient, not only an electrical quantity. In this article only sensing elements (primary transducers) and basic (secondary) transducer systems are considered. Within the group of secondary transducers, capacitive, inductive, reluctance, strain gauge, resistive, potentiometric, and piezoelectric transductions are used for information processing in modern pressure sensors. Pressure sensors can have digital or analog output. In analog sensors output signals are voltage amplitude, voltage ratio, current, frequency, capacitance, and inductance.

FUNDAMENTAL DEFINITIONS

Pressure is one of the fundamental physical quantities that describes the thermodynamic state of a body. Pressure is an intensive type of quantity, that is, defined at points, it is a scalar parameter describing the “force side” of the body. Pressure characterizes the intensity of the internal effective repelling forces tending to increase the volume of a body. The intensity

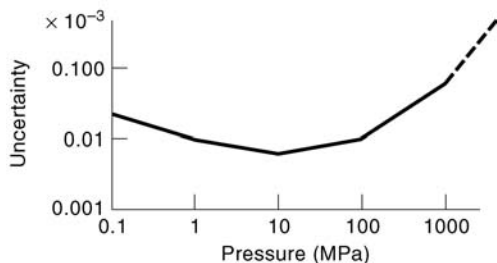


Figure 1. Schematic representation of a pressure sensor.

of this force in every case is demonstrated as a continuously distributed force acting on a unit area. By this we mean the surface area of a vessel enclosing a defined body (for example, gas) or any virtual surface inside the body (gas). Pressure is not a basic physical quantity and therefore it is defined by other basic quantities resulting from the defining equation, that is, newton/m² (N/m²) or pascal (Pa). Let us interpret the pressure using the stress tensor σ_{ij} defined at each point in the volume of a concrete body. In the ideal gas no tangential component of the stress tensor exists. Normal and negative components are equal to each other so we can write $\sigma_{ij} = -p\delta_{ij}$, where δ_{ij} is a Kronecker delta and p is a parameter of state called pressure. In this case pressure can be expressed by the classical formula $p = 2n\langle E \rangle / 3 = \rho \langle v^2 \rangle / 3$, where n is the number of particles in a unit volume, $\langle E \rangle$ is the mean kinetic energy of the particles, ρ is the density of gas, and $\langle v^2 \rangle$ is the mean value of the square velocity of particles. Pressure in an ideal gas is associated only with the thermal motion of particles and therefore can be considered as fully “hot.” In the case of real gas the hot pressure is partially reduced by the so-called *internal pressure* caused by the some intermolecular attractive bond interactions. In the case of liquids the internal pressure makes the hot pressure practically “frozen.” “Cold pressure” can be produced in liquids by mass or in a closed volume by surface forces. In real liquids (e.g., natural water) in thermodynamic equilibrium at some place close to the Earth’s surface, the forces acting inside the liquid on a directed surface element dA are given by the formula $dF = -n p dA$, where n is an unitary vector of element dA and $p = \rho gh$, where g is the Earth’s acceleration, ρ is the density of a given liquid, and h is the distance between the free surface of the liquid and a specific point. This formula was given by Pascal, who defined p as the hydrostatic pressure. In the case of solids the correct interpretation of pressure and its effects is more difficult. The stress tensor σ_{ij} can be formally separated into two components: the axial stress tensor $A_{ij, i=j}$, with identical values of diagonal elements equal to $(\text{Tr } \sigma_{ij})/3$, and the deviatoric stress tensor D_{ij} , whose elements can be expressed as $D_{ij} = \sigma_{ij} - \delta_{ij} A_{ij}$. If all modules of deviatoric elements are small relative to the modules of negative-sign axiatoric elements, we define such a stress state as a quasihydrostatic pressure, which is the real situation. When the deviatoric stress tensor disappears and axiatoric elements are negative, hydrostatic pressure is acting in the solid. In the field of pressure measurement one should become familiar with the following definitions: *Absolute pressure* is pressure measured relative to zero pressure (MPa abs). *Zero pressure* is a state of vacuum. *Vacuum* is a state of space (a closed volume) in which pressure is reduced to a minimum that can be attained through practical means. *Perfect vacuum* is zero absolute pressure, resulting in the complete absence of any particles of matter. *Standard pressure* is the pressure of one normal atmosphere. *Gauge pressure* is the pressure measured relative to ambient pressure, usually relative to actual atmospheric pressure (MPa gag). *Differential pressure* is the difference (usually small) between pressures acting in two specific volumes (points) measured using one sensing element (MPa dif). *Stationary pressure* is pressure that is held constant long enough relative to the longest relaxation time of specific processes that take place in the liquids or solids under study. *Dynamic pressure* is a pressure generated by an impact wave causing a “jump” in the density (also internal energy) on a front of the impact wave. This phenomenon is fully nonreversible. We can speak also about *quick variable pressures*, which are gener-

ated continuously but the length of the pressure wave exceeds the dimensions of the container or the pressure sensor. With a pressure medium flowing in tubes, there is *static pressure* of the flowing fluid (the fluid is a liquid or gaseous substance), which can be measured normal to the direction of flow at the inner surface of a tube. *Impact pressure* is the pressure that can be generated in a moving fluid due to flow velocity. The *stagnation pressure* is the sum of the static pressure and impact pressure. *Partial pressure* is a pressure exerted by one constituent of a mixture of gases.

PRIMARY PRESSURE MEASUREMENTS

There are two primary methods for generation and measurement of pressure: using liquid manometers (1), which are the most accurate range from a few pascals to about 0.5 MPa, and using gas or liquid-operated dead-weight manometers (pressure balances or piston gauges) (2). For the first type of manometer the difference of pressure Δp between two levels h_2 and h_1 is given by

$$\Delta p = \int_{h_1}^{h_2} \rho(h, T) g(h) dh \quad (1)$$

where ρ is density of liquid (mercury in the most cases), known most accurately through the primary measurements of mass and length with temperature and pressure corrections and $g(h)$ is a local Earth acceleration known also on the basis of the most accurate primary measurements of time and length. In the second method, based on metrological properties of a quasisealed piston-cylinder system, the generated pressure is given by

$$p = Q_{\text{eff}} / A_{\text{eff}} \quad (2)$$

where p is a pressure generated just under the bottom surface of the piston and Q_{eff} is the effective weight of a standard mass used as a force generator. The acceleration of the Earth must be known at the specific location of the experiment with highest accuracy; air buoyancy, the surface tension effect, and other corrections should also be taken into consideration. The quantity A_{eff} is the effective area of the piston-cylinder system. At first approximation for low pressures it is equal to the mean values of the piston area and the inner-hole area of the cylinder. A_{eff} depends slightly on temperature and more strongly on pressure, and therefore its pressure dependence is a fundamental problem for designers of this type of pressure etalon (3). There are three fundamental types of piston-cylinder systems: simple-piston, reentrant-cylinder, and controlled-clearance systems. The last one is schematically shown in Fig. 2. The first and third systems have found wide applications in practice; the first for pressures below 1 GPa and the third for pressures up to 1.4 GPa are commercially available. The accuracy of pressure balance depends on its pressure range and is illustrated in Fig. 3 (2).

GENERAL CHARACTERISTICS OF PRESSURE SENSORS

A pressure sensor is normally designed to sense (measure) a specific measurand, pressure, and to respond only to this measurand according to the theoretical formula or to the pressure

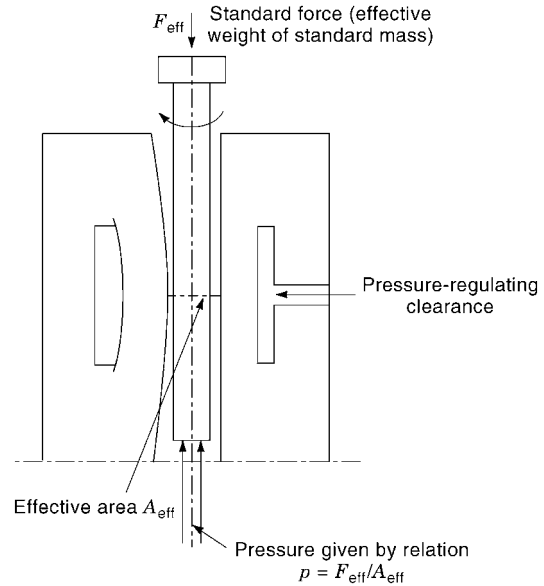


Figure 2. Schematic of a piston gauge (pressure balance) with a clearance-controlled piston-cylinder system. On the right an initial clearance and on the left a reduced clearance at the measurement moment are shown.

scale. The pressure scale is given by several pressure values well established from primary methods of pressure measurements. Real metrological properties of sensors are described by the following characteristics [see, e.g., Ref 4]:

1. Measurand characteristics
2. Electrical design characteristics
3. Mechanical design characteristics
4. Static performance characteristics
5. Dynamic performance characteristics
6. Environmental characteristics
7. Installation (reinstallation) effects and mounting errors
8. Reliability characteristics (overload, life, and the number of operating cycles)

Measurand characteristics for pressure sensors is the *character* of pressure, which can be static, dynamic, or variable (quick-variable), and *range* of pressure defined by lower and upper limits. The range of pressure sensors is in principle unidirectional (e.g., from 0 MPa abs to 10 MPa abs) and asymmetrically bidirectional in the case of gauge pressures when the reference pressure is atmospheric pressure (e.g., from

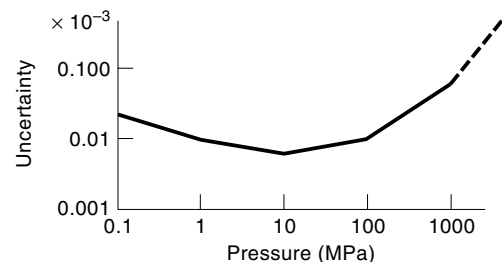


Figure 3. Accuracy of pressure balances as a function of the maximum pressure range (2).

–0.1 MPa gag to 1.0 MPa gag) or expanded and suppressed-zero scale (e.g., from 0.3 MPa abs to 0.5 MPa abs). The algebraic differences between the limits of the range in the span of the sensor (for the preceding examples) are equal to 10 MPa, 1.1 MPa, and 0.2 MPa. Some electrical design characteristics are as follows:

1. *An excitation*, usually stated in terms of dc or ac voltage or of a current applied to the sensor (only piezoelectric pressure-sensing elements do not require external electrical excitation for operation)
2. *Output*, an electrical quantity produced by the sensor as a function of the measured pressure, in analog form as a continuous function of current or voltage magnitudes and also a continuous variation of capacitance, inductance, frequency. In the case of digital output, the system gives information in the form of discrete quantities coded in the specific notation, preserving the accuracy of the sensor throughout the measuring system.
3. *End points*, which are the output values at the lower and upper limits of the sensor range.
4. *Impedances* across the output terminals and across the excitation terminals, and load impedance presented to the sensor output by the associated external circuitry and transmission line (mismatching of output impedances can cause loading error).
5. *Demodulator and amplifier* characteristics and also *grounding*, which can be isolated or common for input and output, or floating when two or more portions of sensor are electrically isolated.

The main static performance characteristics of a sensor is usually determined by the test during which known values of pressure are applied to the sensor in an ascending mode and descending mode (a calibrating cycle), and corresponding output readings are recorded and correlated. The output of the actual sensor is affected by its real behavior, which causes the indicated pressure value to deviate from the ideal characteristic (true values). The algebraic difference between the indicated value and true value is a sensor error or sensor uncertainty (usually expressed as a percentage of full-scale output). The ratio of the error to the full-scale output is defined as the relative uncertainty or accuracy class. Pressure sensor errors can be considered in terms of maximum deviation over the range from known reference lines, but interpretation of individual errors such as nonrepeatability, hysteresis, zero and sensitivity drift, possible friction errors, and nonlinearity (when linear functions of output pressure can be accepted) are required. For high-pressure sensors hysteresis induced by sensing elements is the main source of measurement errors. Dynamic performance characteristics are characterized by a *frequency response*, that is, frequency dependence of the output to pressure amplitude ratio when the pressure varies sinusoidally. We can speak of $\nu_{\max}$ below which the response is correct within some percent of error and $\nu_{\min}$ above which the sensor response is correct, in the cases when a sensor is designed for quick variable pressure measurements. Another useful dynamic parameter is the *response time*, when a stepped change of pressure is applied to the sensor and its output reaches some arbitrary fixed level, say 90% of the final value. In most cases, when $p_{\text{step}} > 0$, the pressure change satisfactorily describes the relation

$p_{\text{read}} = p_{\text{step}} (1 - e^{-t/\tau})$ where τ is the time constant of the sensor. After that time the response of the sensor reaches about 63% of its final value or decreases by 63% of the p_{step} value when it has a negative sign. The upper limit of the frequency response and response time characteristics of the sensor are controlled by its energy-dissipating properties, that is, damping processes. Each oscillation process can be underdamped, critically damped, and overdamped. In an underdamped state of the sensor, an oscillation of the output is observed. In overdamped and critically damped states, the steady value is reaching monotonically with different values of τ . In the last situation (at critical damping) the time constant is smallest. The specific external conditions (temperature, shock, vibration, and humidity) to which a pressure sensor may be exposed during storage and operation constitute *environmental conditions*. The influence of temperature in the *operating temperature range* should be defined within specific error limits—a temperature error or temperature error band. The effect of temperature causes a thermal zero drift, thermal sensitivity shift, and temperature gradient error (a transient error). For the pressure sensors the knowledge of the influence of temperature on the sensor properties is very important because any change of pressure is connected with a change in temperature, especially for dynamic pressure. There are many scientific and industrial demands related to the pressure measurement at elevated or high temperatures (or cryogenic low temperature). For such demands a specific pressure sensor has been developed (5,6). Acceleration effects may be especially large when mechanical pressure-sensing elements are used. The error connected with this effect is usually more severe along one axis of the sensor than along the other axis, which depends on the sensing element's construction. Too large an acceleration (10g or higher) may lead to the failure of the sensor. For rocket technology, special pressure sensors are constructed. In some pressure-sensing elements such as the resistance pressure gauges, the sensitivity to acceleration is negligible. Vibration effects (vibratory acceleration) can affect pressure sensors in the same manner as constant acceleration does. The vibration effects may reduce static error connected with friction error and may induce resonance effects at some characteristic frequencies, with the largest effect connected with the main resonance.

Let us define *low pressures* as pressures for which the standard pressure is not negligible. Practically, we can accept it as 100 MPa, which makes the standard pressure 10^{-3} of the maximum value of the low-pressure value. For this range of pressure the differences among *absolute*, *gauge*, and *differential* pressures are essential. In Fig. 4 there are shown reference configurations adequate to low-pressure using diaphragms as pressure-sensing elements. The most commonly used classical, in principle mechanical, elastic sensing elements and the nature of their deflections are illustrated in Figs. 5 and 6.

DIAPHRAGM PRESSURE-SENSING ELEMENTS

Flat diaphragms are essentially circular, rectangular, or square-edge-clamped elastic thin plates (noncircular plates play an important role in silicon pressure sensors). Let us consider here the simplest situation, that is, a deformation state of a thin circular plate loaded on one side by measured pressure, which is a classical problem in the elastic theory of plates (7). Figure 7, in enlarged scale, shows a plate deformed by

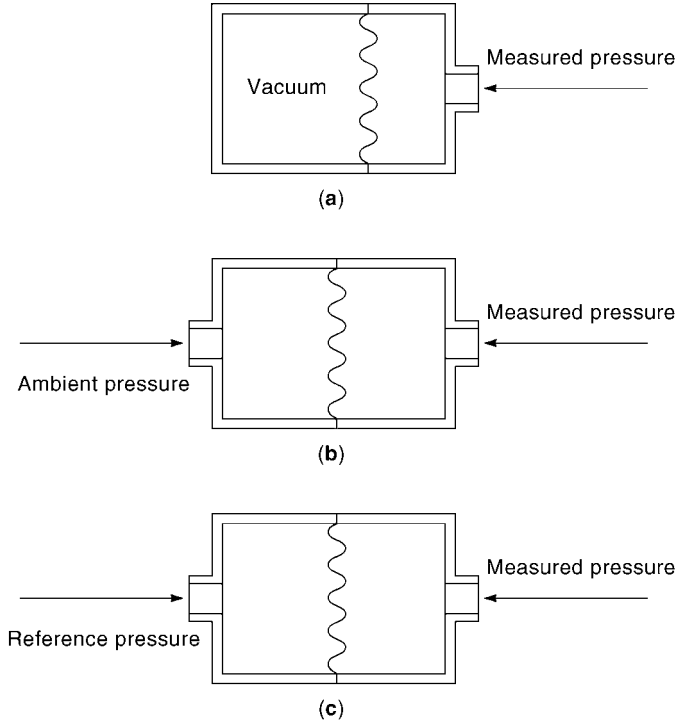


Figure 4. Reference pressures: vacuum (for absolute pressure measurements), ambient pressure or normal pressure (for gauge pressure measurements), and basic, variable pressure (for differential pressure measurements).

pressure. The radial ϵ_r and tangential strains ϵ_t of the bottom free surface are given by

$$\begin{aligned}\epsilon_r &= 3(1 - \nu^2)(a^2 - 3r^2)p/8Eh^2 \\ \epsilon_t &= 3(1 - \nu^2)(a^2 - r^2)p/8Eh^2\end{aligned}\quad (3)$$

where a is the radius of the diaphragm, h a thickness, ν Poisson's ratio, E Young's modulus of the plate material and

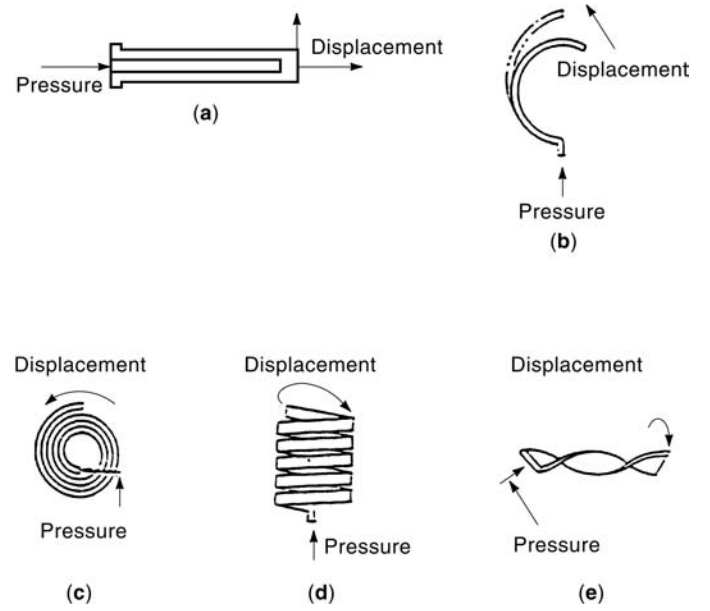


Figure 6. Elastic pressure-sensing elements: (a) straight tube, (b) curvilinear (Bourdon tube) basic type, (c) spiral type, (d) helical type, and (e) twisted type.

r the local (optional) radius. At the center of diaphragm, that is, for $r = 0$, we have

$$\epsilon_r = \epsilon_t = 3(1 - \nu^2)a^2p/8Eh^2 \quad (\text{tensile, positive strain}) \quad (4)$$

and deflection

$$f = 3(1 - \nu^2)a^4p/16Eh^3 \quad (5)$$

On the edge of diaphragm, assuming perfect clamping, for $r = a$, we have

$$\begin{aligned}\epsilon_r &= 3(1 - \nu^2)a^2p/4Eh^2 \quad (\text{compressive, negative strain}) \\ \epsilon_t &= 0\end{aligned}\quad (6)$$

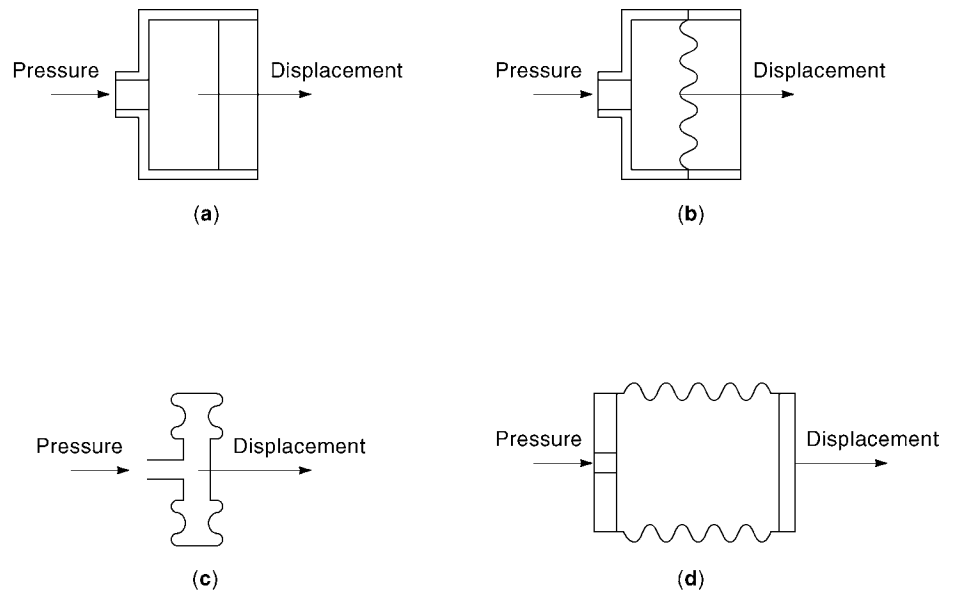


Figure 5. Elastic pressure-sensing elements: (a) flat diaphragms, (b) corrugated diaphragms, (c) capsules, and (d) bellows.

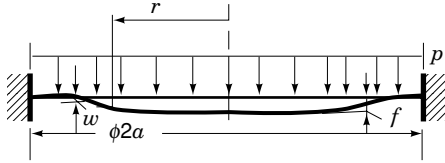


Figure 7. Pressure-loaded edge-clamped thin elastic plate (illustration of deformation in enlarged scale).

A good approximation of the fundamental resonant frequency of the diaphragm is given by

$$f_R = 0.47hE^{0.5}/a^2\rho^{0.5}(1-\nu^2)^{0.5} \quad (7)$$

So, one can see that the deflection at the center is proportional to the applied pressure (differential pressure across the diaphragm) and varies with the fourth power of its diameter. The preceding expressions are valid for small deflections only. With increasing pressure in the deformed diaphragm, tension stress starts to play an increasingly important role. The characteristic pressure or deflection for large deflections is therefore strongly nonlinear. The deflection at the center of a diaphragm for $f < 4h$ can be to a good approximation calculated from following expression (8):

$$pa^4/Eh^4 = 16f/3(1-\nu^2)h + 6f^3/7h \quad (8)$$

For the higher deflections, but not higher than $10h$, the following, simpler, formula can be used:

$$pa^4/Eh^4 = 3.45(f/h)^3 \quad (9)$$

Flat diaphragm pressure-sensing elements as primary transducers of the pressure-to-displacement or pressure-to-strain type are widely used in practice within the pressure range from some kilopascals to about 100 MPa and in special designs up to 1 GPa (high-pressure performance). Their typical accuracy varies from 0.5% up to 2%. *Corrugated diaphragms* contain a number of concentric corrugations of different shapes and dimensions. These increase the stiffness as well as the effective area of the diaphragm, the measurement of which is very important in the case of the pressure-to-force trans-

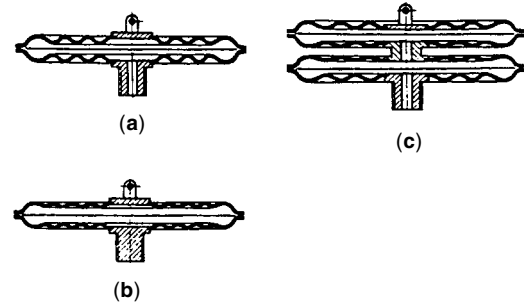


Figure 9. Different type of the capsule systems: (a) open forms for gauge pressure measurements, (b) closed form (aneroid) for measurements with a given reference pressure, if vacuum is used for absolute pressure measurements (typically for atmospheric pressure, and (c) double system of capsules.

ducer. The useful deflection is therefore larger than that of the flat diaphragms. The deflection can be a linear function of applied pressure with proper selection of corrugation and thickness of diaphragms. The coefficient of proportionality between pressure and deflection depends on the shape, location, and number of corrugations. In Fig. 8 examples are given illustrating this problem (8). Corrugated diaphragm sensing elements are typical pressure-displacement primary transducers, having a typical accuracy of 1.5% to 2.5%, and they can measure the pressure from some kilopascals up to some megapascals. Diaphragms made of very thin metal or nonmetallic materials (soft diaphragms) are called *membranes*. Membranes used as pressure-sensing elements are backed by a spring that provides necessary elastic properties. Typically membranes are used to isolate the pressure-sensing element from the pressurized chemically active liquid. As a pressure-transmitting medium a silicone oil with optimal properties from a metrological point of view is used. Pressure-sensing elements called *capsules* can have open or closed form [see Figs. 9(a) and 9(b)]. The closed-form element with vacuum as a reference pressure is frequently referred as an *aneroid* and is used in atmospheric pressure manometers. It consists of two corrugated metal diaphragms formed into shells of opposite curvature and fastened together with hermetically sealed joints around their peripheries, by (most preferred) inert-gas or electron-beam welding.

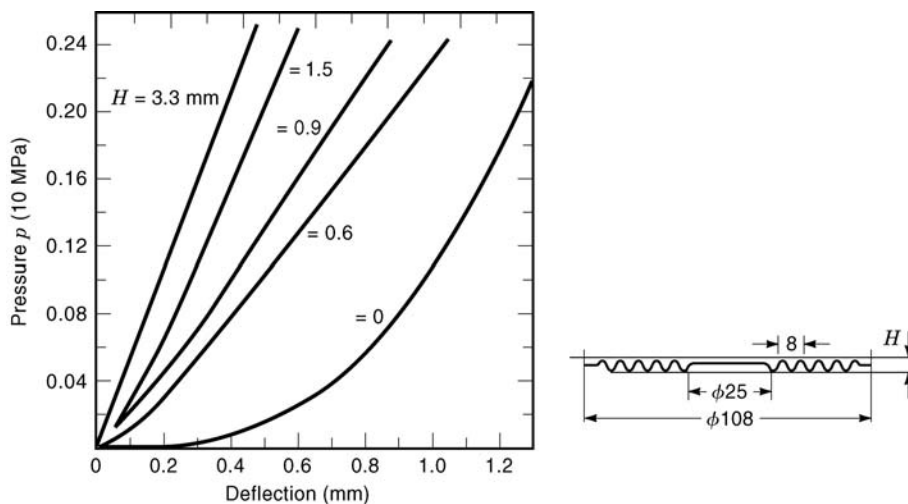


Figure 8. Illustration of the problem of obtaining of the linear dependence of deflection for corrugated diaphragm under pressure (8).

Factors influencing the deflection of capsules are the same as those that apply to corrugated diaphragms. The use of two corrugated diaphragms in one capsule nearly doubles the deflection obtained from each diaphragm. Capsules are often joined in double or triple systems [see Fig. 9(c)]. The *bellows* sensing element [see Fig. 5(d)] is usually made of a thin-walled tube formed in a large number of deep convolutions sealed at one end, which can translate axially when pressure is applied to a port in its opposite anchored end. Due to large numbers of their convolutions, bellows are frequently used when a large stroke is required. The number of convolutions vary from five to over twenty depending on pressure and displacement requirements and on outside diameter. Bellows are sensing elements that can work as a pressure-to-displacement transducer or pressure-to-force transducer. In pressure-to-force transducing the determination of the effective area of the bellows is very important. The effective area of the bellows is the effective area of the piston gauge in equilibrium with the force resulting from the pressure acting inside the bellows, which is equal to the pressure in the piston gauge. The effective area of the bellows is constant to a high degree. The pressure range covered by bellows sensors is typically from 0.1 MPa up to 1.0 MPa and the accuracy is about 1%. In the case of bellows, a large hysteresis up to 2% can be observed. Diaphragm and bellow sensing elements are made of high-elastic metal alloys such as brass, bronze, phosphor bronze, beryllium copper alloy, stainless steel, and alloys called Monel, Inconel, and NI-Span C. An important consideration in the selection of diaphragm materials is the chemical nature of the fluids that will come in contact with the diaphragm. Proper manufacturing methods include heat-treating and pressure-cycling to reduce elastic aftereffects like drift, creeping, and hysteresis in the diaphragms.

BOURDON-TUBE SENSING ELEMENTS

The Bourdon tube is named after the French inventor, Eugenie Bourdon, who patented it in June 1849, and has found very wide applications. The Bourdon tube is a curved or twisted tube, typically oval or elliptical in cross section, which is sealed on one end (tip) and rigidly mounted on the other. When pressure is applied internally to the tube, the tube tends to straighten. This phenomena can be easily but quantitatively explained as follows. When the Bourdon tube is subjected to an internal pressure, the cross section approaches a circle and the length of the tube increases slightly with pressure; hence the effect is to increase the radius of the tube. This results in a curvilinear tip deflection (tip travel) in curved tubes and an angular tip deflection in twisted tubes. The principle of action of this simple device is very difficult to describe in terms of elastic theory. For *C-shaped Bourdon* tubes (see Fig. 10), the initial total angle of curvature ψ_0 (typically of 180° to 270°) decreases by $\Delta\psi$ up to ψ as a result of the effective pressure (the differ-

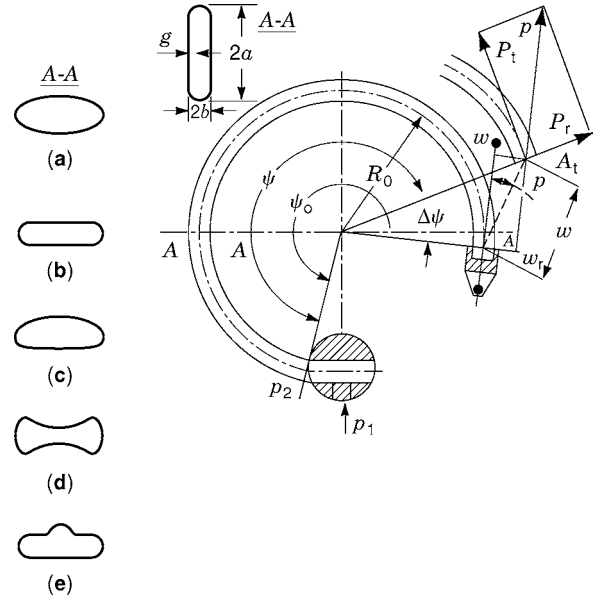


Figure 10. C-shaped Bourdon tube main dimensions; various cross sections of tubes for low-pressure measurements (8).

ence between internal and ambient pressures). In the case of low pressure (thin-walled Bourdon tube, i.e., when the ratio of the thickness g of the tube to the minor dimension of its cross section b is smaller than 0.6) we have the following approximate expression defining the main characteristic of a Bourdon tube (8):

$$\Delta\psi/\psi = p(1 - \nu^2)R^2(1 - b^2/a^2)\alpha/Ebg(\beta + R^2g^2/\alpha^4) \quad (10)$$

where R is the mean curvature radius of the tube and α and β are parameters depending on the shape of the tube cross section the values of which are given in the Table 1. In the case of a high-pressure Bourdon tube (thick-walled tubes, $g/b > 0.7$) we can use following expression:

$$\Delta\psi/\psi = p(1 - \nu^2)R^2(1 - X)/Ebg(g^2/12b^2 + X) \quad (11)$$

where $X = (\sinh^2 c + \sin^2 c)/c(\cosh c \sinh c + \cos c \sin c)$ and constant $c = (3^{0.5}\alpha^2/Rg)^{0.5}$. The tip travel w can be described by two components, radial, w_r , and tangential, w_t as follows:

$$w_r = \Delta\psi R(1 - \cos \psi_0)/\psi_0, \quad w_t = \Delta\psi R(1 - \sin \psi_0)/\psi_0 \quad (12)$$

In practice the deflection of the free end of the tube is a linear function of the pressure. From Eqs. (10) and (11) we can see that the wall thickness, the a/b ratio Young's modulus, the radius of curvature and other parameters in the case of more complicated cross sections of the Bourdon tube determine the

Table 1. Data for Deformation Calculation of C-Shaped Bourdon Tube Expressions (10-11)

Cross Section of the Tube		a/b										
		1	1.5	2	3	4	5	6	7	8	9	10
(a)	α	0.750	0.636	0.566	0.493	0.452	0.430	0.416	0.406	0.400	0.395	0.390
	β	0.083	0.062	0.053	0.045	0.044	0.043	0.042	0.042	0.042	0.042	0.042
	α	0.637	0.549	0.548	0.480	0.437	0.408	0.388	0.372	0.360	0.350	0.348
(b)	β	0.096	0.110	0.115	0.121	0.121	0.121	0.121	0.120	0.119	0.119	0.118

sensitivity of the tube to pressure changes. For laboratory use, the precision of about 0.1% of full scale is attainable for pressures as high as 1 GPa. For pressures up to 2 GPa, the Bourdon tubes have been developed with an accuracy on the order of 1% to 2%. In many commercially available versions of Bourdon manometers the servo systems have been incorporated into the gauge mechanism, enabling automatic recording and regulation. Typical pressure ranges for this type of Bourdon manometers are from 0.01 MPa up to 0.6 MPa with typical accuracy of 0.2% to 2%. As an important extension of the simple Bourdon gauge, pressure-sensing elements in the form of spirally and helically wound tubes have been developed and also widely used. The *spiral Bourdon tube* amplifies tip travel due to its multiturns (usually 3 to 6 turns). Of course, its curvature radius increases with each turn from port to tip but the tube centerline is in one plane. In the case of the *helical Bourdon tube*, since it is coiled into multiturns, the total angle of curvature is typically between 1800° and 3600°. Its tip travel is proportionally greater than that of the tube with less than 360° curvature. For low-pressure applications (up to about 5 MPa) the pressure-sensing element is manufactured from quartz, and for higher pressures only high-strength elastic metallic materials are used. The *twisted Bourdon tube* is a strong flattened tube, twisted along its length. The total number of twists is usually between 2 and 6, but it can be less than one full turn. The twist rate (twists per unit of length) also varies within wide limits. The centerline of the tube is straight throughout its length under normal and pressurized states. The metrological deflection is an angle created by the rotating free end of tube in relation to its anchored second end (pressure port). In special cases prestressed twisted tube gauges do not show any deflection until a pressure close to the definite lower limit of its measuring range is reached. Then deflection occurs only over the narrow part of the pressure range, so zero suppression can be obtained for that kind of transducer. The deflection of this kind of Bourdon tube varies also with the ratio of its major to minor cross-sectional axes, the difference between internal and external pressure, the total tube length, and the rate of twist of the twisted tube. It also strongly depends on the wall thickness of the tube and the modulus of elasticity of the tube material. Materials used for fabrication of Bourdon tubes are practically the same as for diaphragm fabrication. More complicated is the process of its fabrication. The tubing can be machined from one bar stock or can be drawn. Tube fabrication begins with a regular shape and then flattening and forming into the desired configuration takes place. Machined tubes can have one or two ends integrally machined from the same bar stock. Drawn tubes are welded or brazed to the fittings or to the machined base port section of the transducer. Then the tip is sealed, welded, or brazed with great care to ensure leak-proof joints. After these operations the pressure and temperature cycling is performed.

STRAIGHT-TUBE SENSING ELEMENTS

Straight-tube sensing elements are used in an increasing number of pressure sensors. Based on the Bourdon idea it was suggested (9) to design a straight tube with an eccentric bore. When pressurized with liquid the tube will bend because there is a net force due to pressure in the bore acting on one side of the neutral axis of the tube (on the hole axis side). The resultant deflection can be recorded by means of any displacement transducer, for

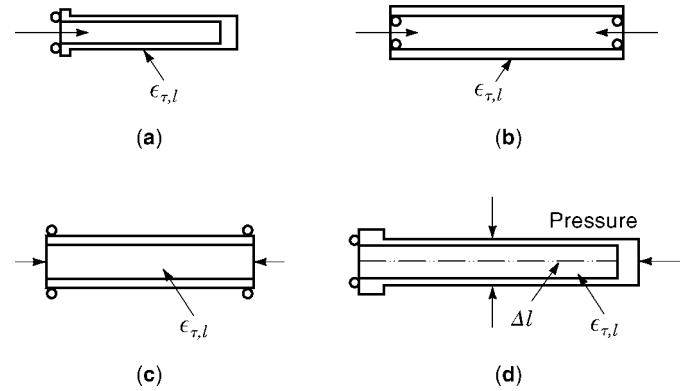


Figure 11. Elastic pressure-sensing elements, short-tube types: (a) closed end, (b) open ends, (c) loaded open ends, and (d) short-tube type bulk modulus.

example a dial micron gauge. Not too high an accuracy from 1% up to 2% of this transducer has been reported. The design is rather easy to fabricate and is well suited to high-pressure applications. The second type of straight-tube element widely used in practice has a circular cross section. A differential pressure acting across its wall causes a very slight expansion or change of the tube diameter and length. These changes of tube dimensions can be easily detected by means of the proper displacement transducer or can be calculated based on knowledge of the longitudinal and circumferential strains of the open surfaces of the tubes. These tube elements being short tubes or cylinders, are divided into two groups: thin-walled cylinders and thick-walled cylinders. Each group can have closed-end or free-tube versions; see Fig. 11. Thin-walled cylinders are sensing elements for small pressures, and thick-walled cylinders are sensing elements for high-pressure measurements. With a strain gauge transduction system $\epsilon \rightarrow \Delta V$ many commercial types of low and high-pressure sensors were developed. In the case of thin-walled cylinders one obtains the longitudinal strains ϵ_l and circumferential strains ϵ_r from Laplace sheet theory (7) taking into account also the existing longitudinal stress due to the pressure acting on the closed end:

$$\epsilon_r = pD(2 - \nu)/4gE, \quad \epsilon_l = pD(1 - 2\nu)/4gE \quad (13)$$

where D is the inside diameter of a cylinder, g the thickness of a cylinder, E Young's modulus and ν Poisson's ratio of cylinder material. The classical tube sensing elements have some disadvantages: the influence of a closed end and the mounting stress in the open end of the tube on strains where the strain gauges are bonded, so in any case such pressure sensors must be tested. In the case of the free cylinder pressure-sensing element (10), that is, with a tube having two free ends, the longitudinal stress does not exist. In this case, for a thin-walled cylinder, from Laplace sheet theory, one can obtain

$$\epsilon_r = pD/2gE, \quad \epsilon_l = -\nu pD/2gE \quad (14)$$

The longitudinal strains are negative here and the results from Poisson's ratio contrast the classical case, where they are positive. This fact enables us to increase the bridge sensitivity by thermostating the strain gauge bonded along the cylinder. Further advantages of this solution are an exact expression for

strains and the lack of an influence of the mounting stress, except the small friction force due to o-ring sealing applied in this design. For high pressures the thick-walled cylinder must be used. Thus the calculations of circumferential and longitudinal strains can be based on Lamé's theory (7), which gives following expressions: For closed-end tubes

$$\epsilon_r = p(2 - \nu)/(W^2 - 1), \quad \epsilon_l = p(1 - 2\nu)/E(W^2 - 1) \quad (15)$$

and for free tubes

$$\epsilon_r = 2p/E(W^2 - 1), \quad \epsilon_l = -2\nu p/E(W^2 - 1) \quad (16)$$

where W is the wall ratio (the ratio of the outer to inner diameter of cylinder). When pressure is applied not only inside the tube but also on the end surfaces (10) one can obtain the following expressions:

$$\epsilon_r = p[2/E(W^2 - 1) + \nu/E], \quad \epsilon_l = -p[2\nu/E(W^2 - 1) + 1/E] \quad (17)$$

when $W \rightarrow \infty$, that is, a cylinder system with pressure acting on the open ends also becomes a free-rod system. When the pressure is applied to the outside of the measuring cylinder that system is called a bulk-modulus pressure sensor (11), and the circumferential and longitudinal strains measured on the inside diameter of the cylinder are

$$\begin{aligned} \epsilon_r &= -p[W^2(2 - \nu)/E(W^2 - 1)] \\ \epsilon_l &= -p[W^2(1 - 2\nu)/E(W^2 - 1)] \end{aligned} \quad (18)$$

The change of the length of this type of cylinder is a measure of pressure,

$$\Delta l = -p[l(1 - 2\nu)W^2/E(W^2 - 1)] \quad (19)$$

where l is the length of the straight measuring tube. From the data presented in Ref. 11, $l = 100 \text{ mm} \pm 0.05 \text{ mm}$, $\nu = 0.296 \pm 0.002$, $E = 186 \text{ GPa} \pm 2 \text{ GPa}$, $W = 2 \pm 0.005$, we obtain the following formula: $\Delta l = p(0.294 \pm 0.007) \mu\text{m} \cdot \text{MPa}^{-1}$. For the strain-gauge free active element pressure sensors the pressure characteristics can be calculated based on its dimensions and material data. With this estimate we obtain an accuracy of 2% to 4%, which is high enough for many industrial purposes.

STRAIN-GAUGE PRESSURE SENSORS

The conversion of pressure-induced strains in definitive elastic pressure-sensing elements (usually flat diaphragms or straight-tube elements) into changes of resistance using a different type of strain gauge has been used in commercial pressure sensors for many years. Two or more commonly used sensors with four active arms of a Wheatstone bridge with ac and dc excitations in special form that gauge ϵ (or $\Delta R/R$), appropriate for pressure measurements, have been developed. For straight-tube sensing elements described previously, using four active element arms of a Wheatstone bridge we have the following expressions for different types of *strain-gauge dilating cylinder* pressure sensors:

Thin-walled closed end cylinder	$\Delta V/V \cong kpD(1 + \nu)/2gE$
Thin-walled free cylinder	$\Delta V/V = kpD(1 + \nu)/gE$
Thick-walled closed-end cylinder	$\Delta V/V \cong kp(1 + \nu)/2E(W^2 - 1)$
Thick-walled free cylinder	$\Delta V/V = kp(1 + \nu)/E(W^2 - 1)$
Thick-walled open-end free cylinder	$\Delta V/V = kp(1 + \nu)[2/(W^2 - 1) + 1]/2E$
Free-rod approximation	$\Delta V/V = kp(1 + \nu)/2E$
Thick-walled bulk-modulus type	$\Delta V/V = kpW^2(1 + \nu)/E(W^2 - 1)$

where ΔV is a output voltage signal, V the excitation voltage of the bridge, and $k = \Delta R/R\epsilon$ the constant of the strain gauge equal to about 2 for metallic gauges and even more than 50 for semiconductor gauges. In the case of a clamped diaphragm, the opposite-sign strains in the center of diaphragm and on its periphery (on the same side) are induced by pressure. For the four active arms of a Wheatstone bridge the following expression for output voltage can be easily derived: $\Delta V/V = \frac{9}{8}k(1 + \nu^2)(a/h)^2(p/E)$. This expression is only an approximation of the real characteristic of the sensor because strain gauges cannot be fixed exactly in the center nor at the periphery of the diaphragm due to its finite dimensions.

RESISTANCE PRESSURE SENSORS

Resistance pressure sensors utilizing the change of the electrical resistance of a metallic wire (12) a semiconductor element (13) with applied pressure are widely used especially in scientific and industrial laboratories for elevated and high-pressure measurements. Pressure-sensing elements, usually in the form of small-coil, double helical coil, or hairpin elements, should be strain-free mounted in a pressure chamber having direct contact with a pressure-transmitting medium, or they can be separated from the chamber using soft bellow, diaphragm, or membrane systems. The pressure-sensing element located inside the high-pressure chamber is connected with the measurement system by a special leadthrough in order to isolate it from the body of the pressure chamber (see Fig. 12). In most cases, manganin wire is used. The relative change of resistance of the manganin sensing element with pressure (typical resistance of 100 Ω) is almost linear up to 1 GPa and can be characterized by the pressure coefficient $\Delta R/Rp$ equal to $(2.2 \text{ to } 2.5) \times 10^{-5} \text{ MPa}^{-1}$. Specialized resistance bridges, modern multimeters, and strain-gauge amplifiers are successfully

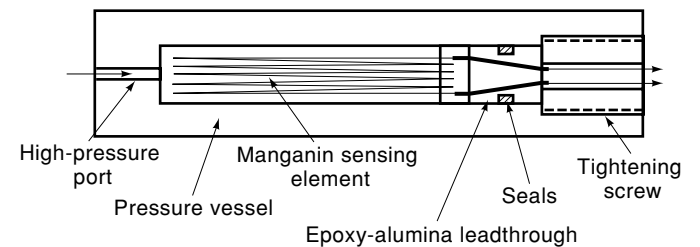


Figure 12. Resistance pressure sensor with a manganin hairpin form sensing element.

used. A resistance change corresponding to pressure changes of about 50 kPa can be attained. The heavily doped bulk n -type InSb single crystal as a semiconductor pressure-sensing element is widely used too (14). This type of gauge is characterized by high (but slightly nonlinear) sensitivity to pressure and low sensitivity to temperature and to the nonhydrostatic state of the pressure-transmitting medium. Typically the resistance of the gauge is below 0.2Ω , thus it is necessary to use a four-probe measurement system for its precise determination. The mean pressure coefficient of the relative resistance of this gauge is equal to about $0.7 \times 10^{-3} \text{ MPa}^{-1}$. Carbon resistors could be also used effectively as pressure sensors at pressures below 1 GPa (15). The variation of resistance is nonlinear, but thermal properties are sufficiently appropriate. Also, thin- and thick-film resistors can be successfully applied in some cases (15). The accuracy of this type of pressure sensor is typically 0.1% to 0.2%. In general, resistance pressure sensors have a number of advantages such as small volume, long-term stability, ease of manufacture, low cost, and need for conventional methods of measurements.

CAPACITIVE PRESSURE SENSOR

Classical capacitive pressure sensors (nonsilicon devices) can be divided into two groups. In the first group, for low-pressure measurements, the elastic diaphragm pressure-sensing elements and the capacitive method of transduction with standard methods of capacitance measurement are used. In the second group, for high-pressure measurements, pressure-sensing elements are the measuring capacitors with substances whose permittivity strongly depends on pressure and weakly on temperature. Here the sensing element is an axial capacitor, filled with a specific liquid dielectric under measured pressure, or has the form of a small, parallel-plate capacitor mounted directly in the high-pressure chamber using a special low-capacitance high-pressure leadthrough. In the first group there are two possibilities. In the first, applied pressure to the diaphragm deforms its center, which moves to and from a stationary electrode and changes the capacitance of the system; or a measured pressure and a reference pressure are applied to a diaphragm mounted between two stationary electrodes and results in changes in the capacitances of two capacitors connected in series in different directions. In the second possibility (see Fig. 13), dual stationary electrodes and a diaphragm create two arms of a four-arm

ac bridge circuit. The pressure-sensing element in most cases is used as a separate element. It is prestressed and clamped during the mounting of the sensor housing. The tension obtained in this way increases the diaphragms resonant frequency and reduces the vibration error of the transducer, which is important for vehicle applications. Prestressing also tends to shorten the transducer's time constant and lower its hysteresis. In the second group of sensors a pressure-measuring liquid polar dielectric or solid dielectric (16) can be used. There are some promising materials for wider application in pressure measurements using differential methods of measurement: First, using a combination of As_2O_3 and $\text{Bi}_{12}\text{GeO}_{20}$ sensing elements with pressure coefficients $110.6 \times 10^{-6} \text{ MPa}^{-1}$ and $-102.8 \times 10^{-6} \text{ MPa}^{-1}$ and temperature coefficients $76.7 \times 10^{-6} \text{ K}^{-1}$ and $73.7 \times 10^{-6} \text{ K}^{-1}$, or using a second combination of $\text{CaCO}_{3\perp}$ and $\text{CaCO}_{3\parallel}$ with pressure coefficients equal to $213.4 \times 10^{-6} \text{ MPa}^{-1}$ and $59.0 \times 10^{-6} \text{ MPa}^{-1}$ and temperature coefficients $3.6 \times 10^{-6} \text{ K}^{-1}$ and $3.6 \times 10^{-6} \text{ K}^{-1}$. The first combination gives a quite good temperature compensation effect and pressure sensitivity of $21.3 \times 10^{-5} \text{ MPa}^{-1}$, about 10 times higher than those of the manganin pressure sensor.

INDUCTIVE PRESSURE SENSORS

In inductive pressure sensors the self-inductance of a single coil is varied by pressure-induced displacement of a ferromagnetic diaphragm located close to the coil. So the motion of the diaphragm with respect to the iron core coil changes its inductance. In another solution of the inductive pressure sensor the inductance changes are produced by the large displacement of the ferromagnetic core within a long coil caused by the motion of bellows or the displacement of a Bourdon tube. In this type of sensor there are big difficulties in compensating for the environmental effects, especially the temperature effect. Therefore the reluctance pressure sensors are used much more frequently than the inductive types.

RELUCTANCE PRESSURE SENSORS

The reluctance changes in the magnetic circuit necessary to produce our output voltage signal in reluctant sensors are in most cases obtained by one of the following: (1) Utilize the pressure-induced deflection of a ferromagnetic diaphragm located between two ferromagnetic-core coils, connected as a two-arm inductance bridge. Then, the diaphragm displacement decreases the gap in the magnetic flux path of one coil while simultaneously increasing the gap in the second coil flux path (see Fig. 14). The output voltage is proportional to the ratio of the two inductances and is almost a linear function of pressure. (2) Use a twisted Bourdon tube with a ferromagnetic plate fastened to its end. The plate together with a two-coil magnetic E-shaped core element creates a magnetic circuit with two small gaps (see Fig. 15). The rotating plate decreases the flux gap of the one coil and decreases the gap of the other coil. Coil connection and production of output voltage signal are the same as those for the solution presented in item (1). (3) Use a differential transformer system with a ferromagnetic core whose displacement depends on pressure and is generated by pressure-sensing elements such as capsules, bellows, corrugated diaphragms, Bourdon tubes, or straight-tube bulk-modulus elements (4). Utilize the changes in permeability due

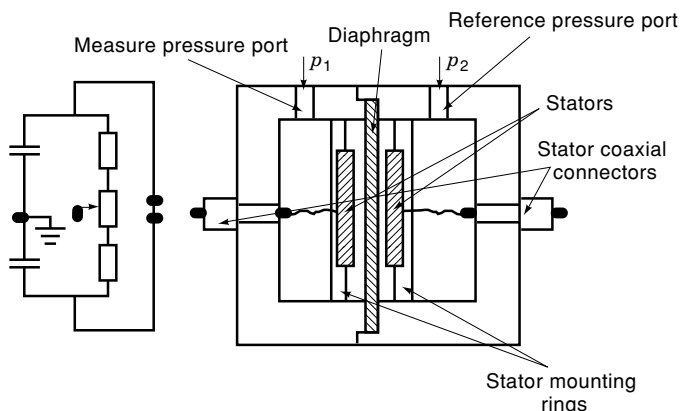


Figure 13. Dual-stator capacitive pressure sensor.

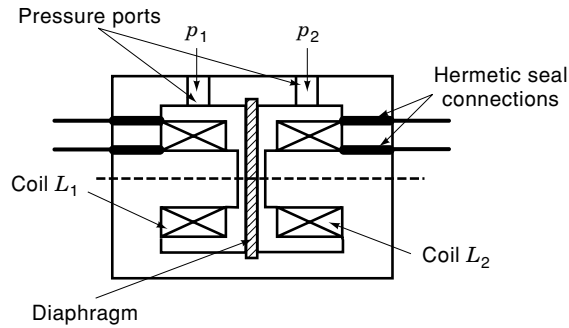


Figure 14. Relative differential pressure sensor with a diaphragm as the pressure-sensing element.

to a change in stress of a specially designed straight-tube sensing element to be a fixed core in the differential transformer circuit. The circuit operation is similar to that described previously but the decreases (or increases) in reluctance path occur in the fixed core proportional to the strain deformations due to the pressure. This method is used for elevated pressure measurements.

PIEZOELECTRIC PRESSURE SENSORS

Piezoelectric pressure sensors employing natural or synthetic quartz as a piezoelectrically active material have been used since 1900. The mechanical and electrical design has now been successfully developed such that they are capable of measuring pressures as small as acoustic pressure and as high as 1 GPa. The response time of a typical sensor is on the order of a few microseconds and resonant frequencies up to about 1 MHz. Piezoelectric sensors are extremely useful in dynamic and quickly-variable pressure measurements. Quartz is very stable, with a high mechanical strength (above 2 GPa) and high modulus of elasticity. The damping is very small and shows practically no hysteresis effect. Its piezoelectric and deformation response up to 2.5 GPa in the uniaxial compressive stress state is linear and can be used at temperatures up to 500°C. The quartz sensing element is a single plate, cut from a quartz crystal in the X-cut plane. Pressure is transmitted directly to the piezoelectric element by a flat diaphragm, producing in the crystal in the x direction a strain that gives rise to polariza-

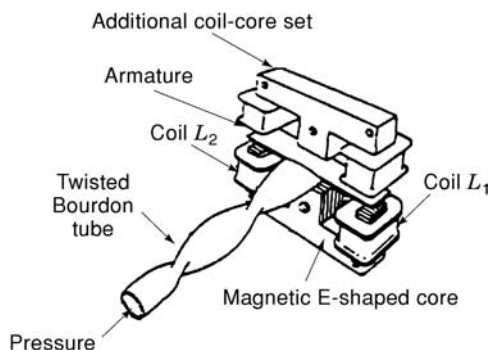


Figure 15. Relative pressure sensor with E-shaped core and a Bourdon twisted tube as the pressure-sensing element (4).

tion of the crystal with the surface charge density proportional to the strain. The total charge on the surface is $Q = d_{11}Ap$, where d_{11} is the appropriate piezoelectric coefficient of quartz equal to 2.25×10^{-12} C/N and A the area of the crystal plate surface (17). For a plate of diameter 5 mm, the charge generated by a pressure of 10 MPa is equal to 0.44×10^{-9} C. For a typical value of thickness, $h = 1$ mm, the capacitance of the crystal capacitor is about 25 pF. Connecting the leads in the form of a coaxial cable would increase this value additionally by 75 pF, so the voltage generated by this pressure would be equal to 0.44 V. As a piezoelectric sensor is designed for dynamic pressure measurement, its frequency characteristics play an important role. The resonant frequency ν_{res} of a thin vibrating plate is given by the expression $\nu_{\text{res}} = c_1/2h$, where $c_1 = (E/\rho)^{0.5}$ is the velocity of longitudinal waves in the plate and ρ the density. So for typical values of h , for $c_1 = 5.5 \times 10^3$ m/s and $\rho = 2.655$ g/cm³ the resonant frequency is approximately 2.8 MHz. The matched diaphragm and quartz crystal form a single mass-spring system. This fact and the bonding will reduce this value to less than one half. While the lower-frequency response of the quartz sensor is governed by electrical characteristics, the upper-frequency response is the result of its mechanical characteristics. The best commercially available sensors cover a wide frequency range, from about 1 Hz up to about 1 MHz. Two types of specialized amplifiers for this kind of sensor are used: the voltage amplifier and the charge amplifier (voltage-mode and charge-mode operated sensor). The construction of the sensing element makes quartz sensors sensitive to acceleration. If a sensor is mounted on an accelerating surface (this is a real situation) additional, unwanted electric signals will be generated. This can be compensated for by using within the same sensor housing an acceleration-sensitive crystal, whose signal reduces the unwanted signal generated in the pressure-sensing element (see Fig. 16). Other piezoelectric materials are used such as tourmaline (for low pressure because of its low mechanical properties), piezoelectric ceramics [barium titanate (BaTiO_3)] and ferroelectric ceramics [(Pb)(Zr)(Ti) O_3 , known commercially as PZT]. These materials have piezoelectric coefficient 100 times as great as quartz but a much smaller maximal working temperature, which is determined by Curie points (300°C for PZT and 200°C for BaTiO_3).

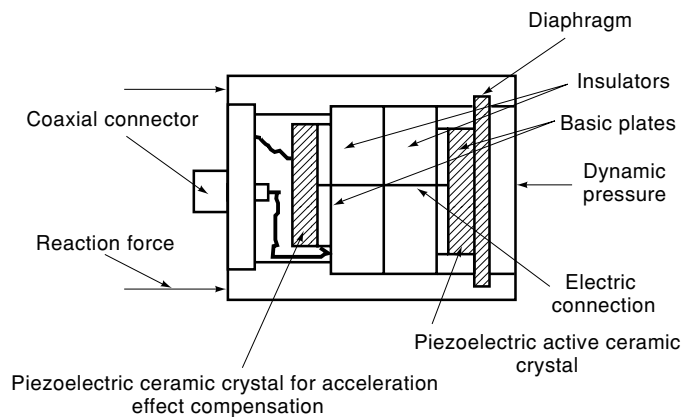


Figure 16. Acceleration-compensated quartz dynamic pressure sensor.

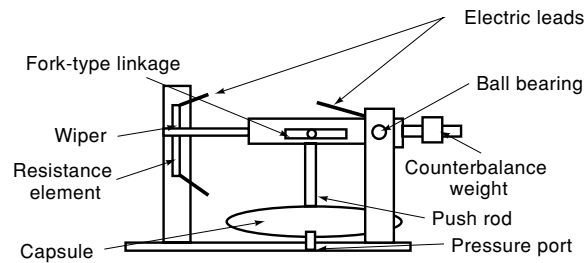


Figure 17. Potentiometric low-pressure sensor with a capsule sensing element and lever system.

POTENTIOMETRIC PRESSURE SENSORS

Pressure sensors with potentiometric transduction elements are widely used in industry because of their simple operation principle, high-level output voltage signal, low cost, availability, and sufficient accuracy. A large number of designs and configurations by many producers have been developed for different measurement requirements. The most commonly used sensing elements are corrugated diaphragms, capsules (single or multiple), and all types of Bourdon tubes. A wide pressure range is covered by these sensors, for low pressure with capsules (see Fig. 17) and for high pressure with C-shaped Bourdon tubes.

OPTICAL PRESSURE SENSORS

An increasing number of optical devices are being used for pressure measurements. Most of them are still in a semicommercial form. They can be divided into two groups: with and without optical fibers used as transducing or pressure-sensing elements (light transmission by optical-fiber systems can be used in both groups but discussion of such systems is not in this article's scope). In the first group, such phenomena as a pressure-dependent refractive index, pressure translation of characteristic spectral lines (ruby high-pressure scale), or pressure-dependent light intensity in some pressure-metric media, as well as pressure deformation followed by detection by interferometric methods, are used.

In the second group, some pressure-dependent properties of multimode or single-mode high-refractive-index optical fibers and different devices for fiber-optic systems, where pressure-dependent deformation changes the light intensity, are widely used (18). The measured pressure range for this group is from acoustic pressure (10 kPa) up to pressures as high as 1 GPa to 2 GPa. For example, Fig. 18 shows the general construction of a pressure sensor based on a straight-tube sensing element

(primary transducer) [compare Fig. 11(d)], with a differential fiber-optic linear displacement transducer (the first secondary transducer in the sensor system) mounted inside the tube to measure pressure-induced changes in its inner diameter. Two receiving fibers, either single-mode or multimode (50 μm and 125 μm) were located on the moving ends of cantilever arm of special carriage. The ends of fibers were slightly prefused to create microlenses. Exactly on opposite sides of those fibers, in the same plane, on the stationary part of the carriage, was located a third fiber that illuminated the cores of the receiving fibers. So we have an input signal of intensity I_0 and two output signals of intensities I_1 and I_2 . In the null state of the transducer the illuminations of the receiving fibers are the same. In order to obtain a response to pressure, a special electronic photosensitive device, giving a voltage signal proportional to $I_1 - I_2$ [or to $(I_1 - I_2)/(I_1 + I_2)$], was used. In the case of 10 mm inner diameter of the tube, with the distance between emitter input and receiving output fibers was approximately 0.3 mm, and with a superluminescent diode emitting light at $\lambda_c = 810$ nm with spectral halfwidth 4 nm, the sensitivity of the displacement transducer was about 1 mV/ μm , so that the sensitivity of the device was 45 $\mu\text{V}/\text{MPa}$. The overall error was estimated as no larger than 1.5% (19).

The main advantage of optic pressure measurement systems is the elimination of any electromagnetic energy transfer through hazardous spaces, for example in methane-containing coal mines or in explosive-materials storage.

SEMICONDUCTOR DEVICES AS SILICON PRESSURE GAUGES

Silicon pressure gauges based on silicon chips etched to form a thin (2 μm to 5 μm for low pressure) or thick (20 μm to 50 μm for elevated pressure) diaphragms with a diffused, ion-implanted piezoresistor or with a capacitor or other diaphragm deformation detectors are widely used in a large number of applications such as altimeters, blood-pressure gauges, tire-leak detectors, pneumatics, hydraulics, and oceanography investigations (20). The main features of silicon pressure sensors are that they can be mass-produced in a variety of ranges, which lowers the cost. They could be smart sensors, are very small, and have good linearity (1% or better), stability, and sensitivity. The sensitivity can be varied from 10 mV/kPa to 1 $\mu\text{V}/\text{kPa}$ for high pressures. Moreover, their interfacing is convenient for signal processing. It is expected that in the next decade silicon pressure sensors within the pressure range up to 200 MPa will continuously replace conventional pressure-measurement devices and within the pressure range up to 1 GPa their industrial application will be much wider. In Fig. 19 a short presentation of

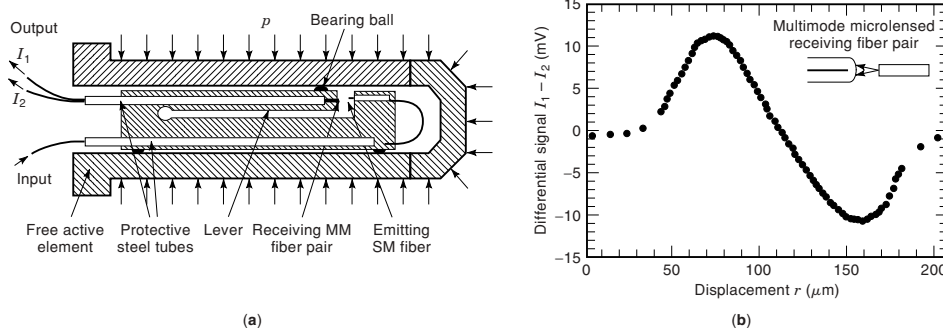


Figure 18. Schematic drawing of straight-tube pressure-sensing element with differential fiber-optic linear displacement transducer and its voltage-displacement characteristics (19).

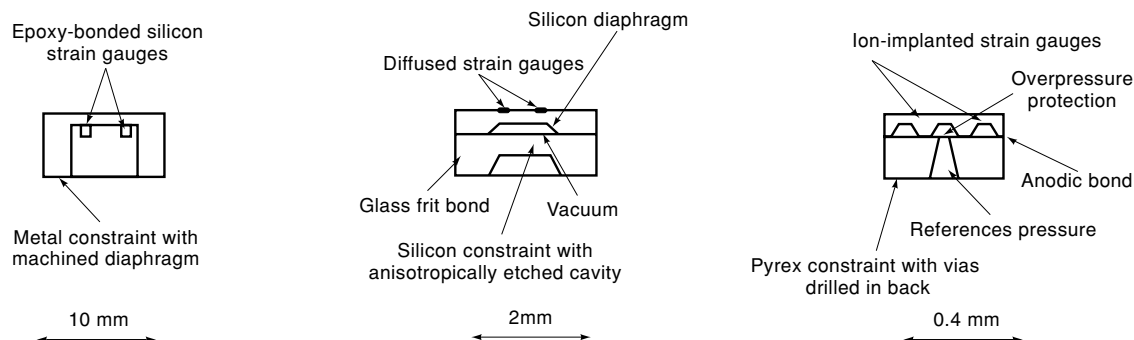


Figure 19. Illustration of the evolution of silicon pressure sensors. On the left a preprototype of sensor is shown.

the development of this kind of sensor is given. Figure 19(a) shows a schematic of the preprototype of a silicon sensor, with a metal constraint and diaphragm but with a silicon strain gauge. Its overall dimensions of about 10 mm are macroscopic. Figures 19(b) and 19(c) present the initial state of silicon sensor technology with a diffused and ion-implanted strain gauge having typical overall dimensions of about 2 mm. The typical absolute pressure sensor with a silicon diaphragm sensing element for low pressure with ion-implanted strain gauges and overpressure protection is shown in Fig. 19(c). In Fig. 20 is shown the dependence of output voltage of the transducer versus pressure, with the upper limit of measuring pressure well represented and good linearity. In the case of the output voltage graphed against elevated pressure, a large nonlinearity can be observed. The large nonlinearity and much smaller sensitivity are disadvantages of this type of sensor for high-pressure measurements. Currently produced silicon sensors have sophisticated silicon technology and can have overall dimensions of about 1 mm. Different types of deflection diaphragm detection methods are used, such as the widely used capacitive method (21) or field-effect transistor method (22), the advantage of which is the direct displacement to voltage (or current) transduction (see Fig. 21), (23). In the case of capacitance pressure sensors we have to note that stray capacitances are of the same order as the sensor capacitance. In all silicon pressure sensors, temperature compensation devices must be added to circuits.

Absolute Pressure Sensors

When the reference pressure in silicon sensors is a vacuum pressure [see Fig. 18(b)] a specific sensing element will be

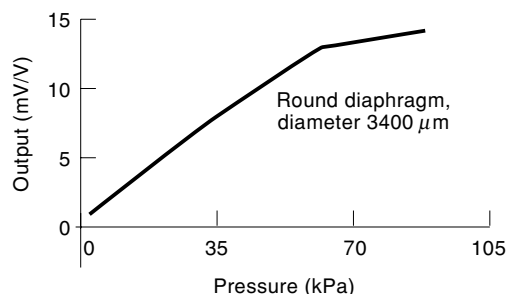


Figure 20. Low-pressure response of a pressure silicon sensor with overpressure protection (22).

under the influence of absolute pressure, but the operation principle of the sensing element does not change. Sensors can easily measure absolute pressures and are widely applied in the vacuum technology industry and space investigations. There are well-established technological procedures for fabrication of such sensors (24). For absolute pressure measurements, it is important to join a silicon diaphragm perfectly for a sufficiently long time to the sensor constrained under high vacuum with no leakage through the bonding surface.

Differential Pressure Sensors

The measurement of differential pressure plays a very important role for flow-parameter measurement, that is, the mass-rate or volume-rate measurement. It is especially important in the case of natural gas transportation, for example, in the case of intercontinental and international ventures. When the pressure difference between two pressurized volumes is small and the maximum pressure does not exceed some megabars the silicon sensor pressure is applied. Differential pressure sensors have the same configuration as gauge sensors, that is, sensors for which the reference pressure is an ambient pressure at the moment the reference cavity is hermetically sealed. In differential sensors the reference cavity is connected with a volume of smaller pressures. Zero deflection of a diaphragm is achieved at equal input pressures (from both sides of the diaphragm) but equilibrium of the measurement system may occur at a some value of differential pressure. In such cases the pressure sensor output is proportional not only to a differential pressure, but to a certain degree to a static pressure as well. This last effect is the static pressure sensitivity of the sensor, which depends on the design of the sensor and material used. The shape of

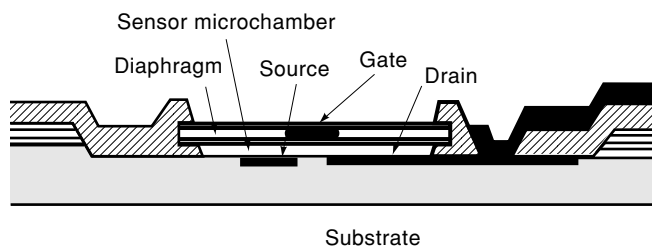


Figure 21. Schematic cross section of a pressure-sensitive field-effect transistor sensor (21).

Table 2. Commercially Available Pressure Sensors

Sensing Element and First Transduction Element	Pressure Range ^a	Calibration Accuracy	Producer
Diaphragm, strain resistance gauge	0–50 MPa (G,D)	0.5%–1.0% FS	Dimisco, West Wood, MN
Silicon, stainless steel diaphragms	0–50 MPa (G)	2.0% FS	Entran Sensors & Electronics, NJ
Diaphragm, thin resistance film	0–50 MPa (A,G,D)	0.5% FS	Statham, CT
Resistance, semiconductor <i>n</i> -type InSb single crystal	0–1 GPa (G)	0.5%–2.0% FS	High Pressure Research Center, Warsaw, Poland
Resistance, manganin wire	0–2.5 GPa (G)	0.5–2.0% FS	
Metal foil, semiconductor strain gauge	0–10 MPa (A,D,G)	0.5%–2.0% FS	Sensotec, Inc., OH
Bourdon tube	0–100 MPa (G)	0.1%–1.5% FS	Cole-Parmer Instrument Co.
Diaphragm	0–50 MPa (A,D,G)	0.1%–1.5% FS	Illinois
Bourdon tube, diaphragm	0–100 MPa (D,G)	0.5%–2.0%	Dwyer Instrum. Inc., Chicago, IL
Bourdon tube	0–0.4 GPa (G)	0.5%–1.0% FS	Nova Swiss, Nova Werke AG, Effreticon
Straight short tube, strain gauge	0–1.4 GPa (G)	0.5%–1.5% FS	Switzerland
Straight short tube bulk-modulus potentiometric	0–1.5 GPa (G)	1.0% FS	Harwood Engineering, MA

^a A, absolute pressure; G, gauge pressure; D, differential pressure; % FS, % of full scale.

diaphragms, especially their deflection, can be determined using conventional methods of calculation because of the small influence of static pressure on the elastic properties of silicon. The larger influence of pressure on a differential pressure sensor is connected with the pressure effect on the electronic properties of strain-sensing elements (for example, for the diffused strain gauge on the diaphragm surface). When the pressure is large and attains some hundreds of megapascals the effect of static pressure can be so large that differential pressure is more accurately measured as the difference between two absolute or gauge pressures.

ACTUAL MARKET INFORMATION

The designing and technological processes of modern pressure sensors are now in a strong development stage. Table 2 gives actual pressure-sensor producers, types, and main metrological properties of commercially available pressure sensors.

APPENDIX. REFERENCE TABLES

Tables 3 and 4 present respectively some common units of pressure and some values of the acceleration of gravity.

Table 3. Popular Earlier-Used Pressure Units

Bar	0.100000 MPa
Technical atmosphere, ata	0.0980665 MPa
Standard (normal) atmosphere, atm	0.101325 MPa
Pound force per square inch, psi	6.894757 kPa
mm Hg (torr)	0.133322 kPa
mm H ₂ O	9.80665 Pa

Table 4. Acceleration of Gravity in Some Capitals

Place	<i>g</i> (m · s ⁻²)
Washington	9.80082
Paris	9.80928
Moscow	9.81523
Warsaw	9.81223
Standard	9.80665 ± 0.00005

BIBLIOGRAPHY

1. C. R. Tilford, Three and half centuries later—The modern art of liquid-column manometry, *Metrologia*, **30** (6): 545–552, 1994.
2. G. F. Molinar, An old instrument in the new technological scenery—The piston gauge in liquid media up to 1 GPa, *Metrologia*, **30** (6): 615–623, 1994.
3. R. S. Dadson, The precise measurement of steady pressures, in G. N. Peggs (ed.), *High Pressure Measurement Technique*, New York: Applied Science, 1983, pp. 3–67.
4. H. N. Norton, *Handbook of Transducers for Electronic Measuring Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1969.
5. C. Rein, The pressure-temperature dependence of the electrical resistance of the Zeranin in the temperature range 160–370 K and at pressure up to 1 GPa, *Meas. Sci. Technol.*, **4**: 1194–1200, 1993.
6. S. Porowski et al., The SPG20 Semiconductor Pressure Gauge for temperature 77–300 K and pressure range 0–1.2 GPa, *UNIPRESS*, Warsaw: High Press. Res. Cent., Pol. Acad. Sci.
7. S. Timoschenko, *Theory of Plates and Shells*, New York: McGraw-Hill, 1987.
8. W. Trylinski, Pressure controlled elastic elements, in *Fine Mechanisms and Precision Instruments: Principle of Design*, Oxford, UK: Pergamon, 1971, pp. 202–223.
9. A. G. Nagatkin, Novel type of spring element for pressure gauges, *Priborostroenie (Moscow)*, **5**: 13–16, 1956.
10. R. Wisniewski and G. F. Molinar, Strain gauge manometer with pressure up to 0.5 GPa acting inside and on the open ends of a free cylinder, *Rev. Sci. Instrum.*, **63**: 3983–3984, 1989.
11. R. Wisniewski and G. F. Molinar, Free active element bulk-modulus high pressure transducer, *Rev. Sci. Instrum.*, **67**: 2020–2021, 1996.
12. G. N. Peggs and R. Wisniewski, Electrical resistance gauges, in G. N. Peggs (ed.), *High Pressure Measurement Technique*, New York: Applied Science, 1983, pp. 215–248.
13. M. K. R. Vyas, Gallium arsenide as pressure sensor, *High Temp. High Press.*, **6**: 237–240, 1974.
14. M. Konczykowski et al., Narrow gap semiconductors as low temperature pressure gauges, in C. V. Chu and J. A. Woolam (eds.), *High Pressure Low Temperature Physics*, New York: Plenum, 1978, p. 532.
15. W. G. Scaife and G. N. Peggs, Pressure transducer based on various physical effects, in G. N. Peggs (ed.), *High Pressure Measurement Technique*, New York: Applied Science, 1983, pp. 179–213.

16. J. H. Colwell, A solid dielectric capacitance pressure transducer, in K. G. Timmerhaus and M. S. Baber (eds.), *High Pressure Science and Technology*, New York: Pergamon, 1997, pp. 798–804.
17. L. J. Griffiths, Measurement techniques for dynamic pressure, in G. N. Peggs (ed.), *High Pressure Measurement Technique*, New York: Applied Science, 1983, pp. 249–303.
18. B. Culshaw and J. Dakin, (eds.), *Optical Fiber Sensors: Systems and Application*, Vol. 2, Norwood, MA: Artech House, Chapter 12.
19. W. J. Bock, T. Eftimov, and R. Wisniewski, A differential fiber-optic transducer for hydrostatic pressure measurement, *Sensors and Actuators A*, **70**: 238–242, 1998.
20. E. M. Blasser, H. Ko, and E. T. Yon, A miniature digital pressure sensor transducer, *Proc. 24th Annu. Conf. Eng. Med. Biol.*, Las Vegas, NV, 1971, p. 221.
21. H. Ko. Wen, Solid-state capacitive pressure transducers, *Sensors Actuators*, **10**: 303–320, 1986.
22. J. T. Suminto and W. H. Ko, Pressure-sensitive insulated gate field-effect transistor (PSIGFET), *Sensors Actuators*, **A21–A23**: 126–132, 1990.
23. J. M. Lysko and R. S. Jachowicz, Modification of PS-FET construction and technology, in R. Wisniewski and A. Magiera (eds.), *New Tendencies in Measurements of High Pressures*, Poland: Jadwisin, 1997.
24. J. Bryzek et al., *Silicon Sensor and Microstructures*, Silicon Valley, CA: Nova Sensor, 1990.

ROLAND WIŚNIEWSKI
Warsaw University of Technology

PIEZORESISTIVE DEVICES

Piezoresistive devices are devices whose functioning principle is based on the piezoresistive effect experienced by certain materials upon elastic deformation (discovered by Lord Kelvin, 1856). The word “piezoresistive” derives from the Greek “piezein,” which means “to press.” The piezoresistive effect involves a change in the resistance R of an electrical conductor [$R = \rho l/A$, where ρ is the material (Ω/cm), resistivity, l the conductor length, and A the transverse section area] as a consequence of a change in their geometrical parameters (at the macroscopic level) originated by an external stimulus. In certain materials, however, an equally important part of the piezoresistive effect is due to the change in resistivity ρ .

Piezoresistive materials are not themselves able to generate an electrical signal when strained; they need to be supplied with an external voltage source. If a voltage is applied between the extremities of a piezoresistive bar whose geometry is changing, the current also changes according to the change in resistance, in accordance with Ohm’s law. In this very elementary example, the current that flows through the piezoresistive bar is the output electrical signal directly related (by means of resistance variation) to the change of an external stimulus. Since piezoresistive materials upon deformation transform mechanical energy into electrical energy, they can be employed in the fabrication of transducers to make physical sensors. Piezoresistive sensors are called *active* sensors, in contrast with *passive* sensors, which themselves generate an electrical signal when stimulated and therefore do not need to be supplied with an external voltage source (1). Active sensors need specific supply circuits depending on the transducer technology.

Piezoresistive sensors are used to perform direct measurements of dynamic and geometric parameters (2) such as acceleration, cracks, creep, deformation, displacement, fatigue, flow, force, height, level, mass, pressure, and torque.

The fabrication of piezoresistive sensors involves the use of very different materials, such as conductive elastomers, carbon fiber, pure metals, or alloys (commonly with nickel and copper). More sophisticated processes are based on thin-film, thick-film, and solid-state technology (2).

BASIC THEORY

Let R be the resistance of an electrically conducting bar in the quiescent state at room temperature. By differentiating the expression

$$R = \rho \frac{l}{A}$$

we have

$$\begin{aligned} dR &= \frac{\partial R}{\partial \rho} d\rho + \frac{\partial R}{\partial l} dl + \frac{\partial R}{\partial A} dA \\ &= \frac{l}{A} d\rho + \frac{\rho}{A} dl - \frac{\rho l}{A^2} dA \end{aligned}$$

and dividing by R we finally obtain

$$\frac{dR}{R} = \frac{l}{A} \cdot \frac{d\rho}{R} + \frac{\rho}{A} \cdot \frac{dl}{R} - \frac{\rho l}{A^2} \cdot \frac{dA}{R}$$

which reduces to

$$= \frac{d\rho}{\rho} + \frac{dl}{l} - \frac{dA}{A} \quad (1)$$

Consider now a piezoresistive bar with cylindrical geometry (see Fig. 1), to which an electric field E_l is applied along the longitudinal direction, and suppose that, upon elongation, the initial value of the piezoresistor changes to the final value (the initial values of the length, the diameter, and the resistivity change too). As a consequence of the length increase a reduction in the cross sectional area occurs too, according to Poisson’s ratio $\nu = \epsilon_d/\epsilon_l$ (diametral strain per unit longitudinal strain).

It can be easily shown that the longitudinal gauge factor is given by

$$F_l = 1 + 2\nu + \frac{1}{\epsilon_l} \cdot \frac{d\rho}{\rho} \quad (2)$$

In the same manner also a transverse gauge factor can be defined.

Values of the longitudinal gauge factor for some commonly used materials are given in Table 1.

CONDUCTIVE RUBBERS

Conductive rubbers are commonly made by loading silicone-based polymers with conducting powder (preferably graphite)

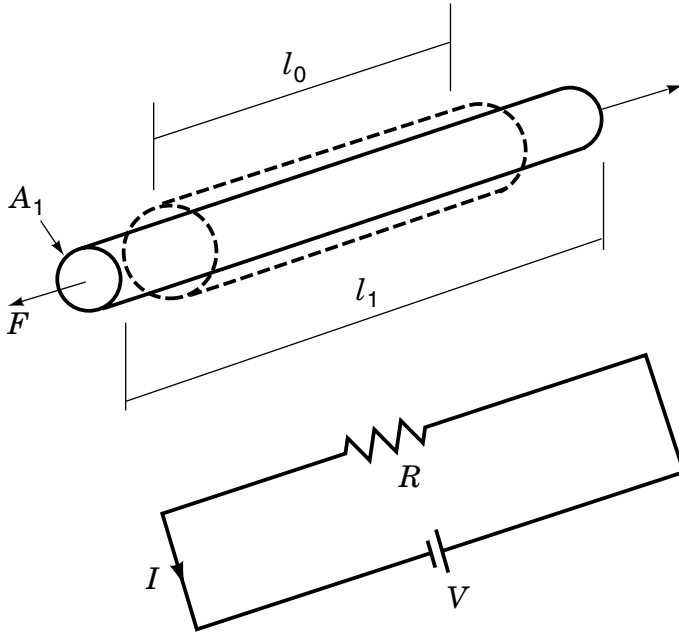


Figure 1. Piezoresistive bar to which traction is applied.

or grains of small dimensions, depending on the applications. Polymer strain gauges are largely used in robotic applications for making highly robust and compliant prototype tactile sensors that resist damage and accommodate inaccuracy during the end-effector positioning and the grasp. Different arrangements have been proposed, including row-by-column matrices protected by Mylar films (Fig. 2) or silicone rubber skin. They use polyurethane foam to increase compliance, as well as conductive rubber wires for electrical connection with the acquisition electronic unit, preferably realized in hybrid technology (5).

On applying a force to the rubber matrix (5), the contact area between the rubber increases and the resistance R decreases. The maximum useful load reaches about 454 kg. At low loads, the sensor response follows the law $R = k/F$, where $k = 4 \text{ M}\Omega\text{kg}$ and F is the applied force.

METAL STRAIN GAUGES

The most popular and reliable technology, at very low cost compared with other technologies available today, is that of metal strain gauges. The transducers are usually made with a metal

stripe a few micrometers thick (6). The length l can be much larger than the width (see Fig. 3); a typical thickness is about $5 \mu\text{m}$, and typical values of the resistance range from 120Ω to 750Ω . The metal foil grid is backed by a plastic carrier with a thickness of about $25 \mu\text{m}$. The longitudinal and transverse gauge factors are about 2 and 0 respectively. Conventional metal foils of piezoresistive strain gauges are particularly suitable for mounting on bodies on which the strain must be measured. They are also called *bonded* strain gauges (3). (*Unbonded* strain gauges are made of several conducting wires suspended between two insulating pins; they are excellent force transducers, but they are cumbersome and also difficult to fabricate.)

Bonded gauges are placed directly on the surface of the body of which the strain is to be measured. The initial dimension can be carefully reduced by etching away very small parts of the backing polymer carrier. The surface must be clean and dry to ensure good adhesion of the special glues and catalysts used (3,7). Because of its geometry, the piezoresistor is mounted with the grid parallel to the direction of the dominant strain (*longitudinal strain*), although the body can also experience a secondary strain (*transverse strain*) in the orthogonal direction.

Table 1. Resistivity and Longitudinal Gauge Factor of the Most Common Materials for Piezoresistive Sensors

Material	Resistivity ρ ($\times 10^{-6} \Omega \cdot \text{cm}$)	F_1
Silver (Ag)	1.628	3.5
Aluminum (Al)	2.828	3.3
Gold (Au)	2.4	3.7
Copper (Cu)	1.724	2.9
Iron (Fe)	10	3.2
Nichrome V ($\text{Ni}_{80}\text{Cr}_{20}$)	7.1	101
Nickel (Ni)	6.84	-11
Platinum (Pt)	10	4.4

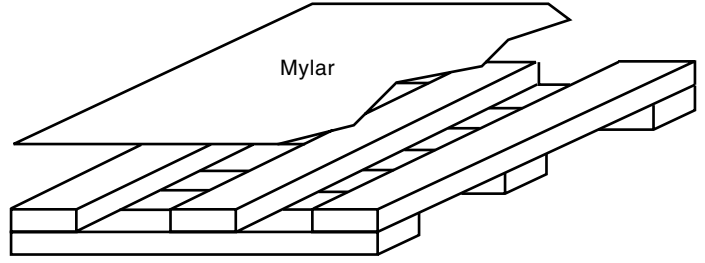


Figure 2. Conductive rubber tactile sensor.

A particular geometric disposition of more than one strain gauge, a *rosette* [see Fig. 1(c)] is used to sense nonuniform force fields involving multicomponent strains.

Metal strain gauges are not suitable for miniaturization.

THICK-FILM STRAIN GAUGES

Thick-film strain gauges consist of special resistive inks, loaded with submicrometer conducting particles and insulating glass parts suspended in viscous organic fluids (8). They are screened and fired on ceramic substrates. The strain sensitivity is strongly influenced by the resistivity, composition, structure, and nature of the conductive grains, and by the direction of the strain with respect to the electric field. The conduction mechanism and the strain sensitivity are dominated by the tunneling effect. A longitudinal gauge factor $GF_L = 13.8$ and a transverse gauge factor $GF_T = 11.6$ have been measured. Thick-film strain gauges are preferred for high-temperature operation (350°C) and in hostile environments (2). They are not satisfactory for miniaturization.

THIN-FILM STRAIN GAUGES

Thin-film strain gauges, usually made of SnO or SiO-Cr , are deposited on ceramic substrates by vacuum processes such as evaporation and sputtering. Device miniaturization is possible, and the maximum temperature of operation reaches about $+150^\circ\text{C}$.

SEMICONDUCTOR STRAIN GAUGES

In contrast with the previous technologies, which are largely employed for the fabrication of universal strain gauges in force field measurement, semiconductor monolithic sensors are limited to pressure and acceleration measurement.

Micromachining and integrated circuit (IC) technologies allow the fabrication of microsensors and microactuators for a temperature of operation not more than $+120^\circ\text{C}$. Semiconductors experience an additional piezoelectric effect, which leads to small differences in the gauge factor $GF = 2 - K$, where 2 is a pure geometrical factor and

$$K = -\frac{1}{\epsilon} \cdot \frac{d(n\mu)}{n\mu} \bigg|_{\epsilon}$$

takes account of the semiconductor nature of the material. In this expression ϵ can be the longitudinal or transverse strain, n is the number of the charge carriers, and μ is the mobility (2).

TYPES OF PIEZORESISTIVE DEVICES

Pressure Transducers

Pressure transducers convert pressure variations into resistance variations. Both unbonded and bonded metal strain gauges, and also sputtered thin-film gauges are used for this purpose. The transducers are usually located on the back side of a membrane, and the pressure is applied to the opposite side (9).

Force Transducers

Force transducers are realized by using metal, both unbonded and bonded, and semiconductor strain gauges. Force measurements are usually performed by combining shear stresses with cantilever beams of different configurations (9).

Accelerometers

Accelerometers, based on metal bonded strain gauges and semiconductor strain gauges, convert the displacement of a mass-spring system into a variation of resistance. In the most

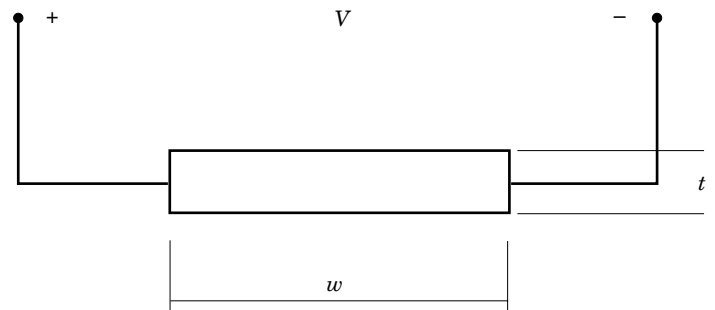


Figure 3. Cross section of a planar strain gauge to which an electrical potential V is applied along the width w , while a traction is applied along the length l (orthogonal to the page).

common devices, the gauges are bonded to a cantilever beam terminated with a seismic mass (9).

Torque Transducers

Torque transducers sense the shear stresses in a torsion bar as a consequence of the applied torque. Usually semiconductor and bonded metal foil strain gauges, in a cruciform arrangement, are used for this application (9).

Displacement Transducers

Displacements are measured through the deflection of a bending beam. The strain gauges are mounted to sense both the tension and the compression of the beam. The beam is deflected by a sensing shaft (9).

BIBLIOGRAPHY

1. J. J. Carr, *Sensors and Circuits*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
2. K. Bethe, The scope of the strain gage principle, *Proc. Int. Conf. VLSI Comput. Peripherals. VLSI and Microelectron. Applications Intell. Peripherals Interconnection Networks*, Washington, DC, May 1989, Vol. 3, pp. 31–38.
3. P. Dario, M. Bergamasco, and A. Fiorillo, Force and tactile sensing for robots, in P. Dario (ed.), *Sensors and Sensory Systems for Advanced Robots*, Berlin: Springer-Verlag, 1988.
4. R. A. Russel, An imaging force sensor for robotics applications, *National Conference on Robotics*, Melbourne, 1984, pp. 123–127.
5. Interlink Electronics, The force sensing resistor, a new tool in sensor technology, Technical Paper, Santa Barbara, CA, June 1986.
6. HBM, Darmstadt, Germany.
7. D. Hagner, Metal strain gages, in J. Webster (ed.), *Tactile Sensors for Robotics and Medicine*, New York: Wiley, 1988.
8. C. Canali et al., Piezoresistive effects in thick-film resistors, *J. Appl. Phys.*, **51** (6): 3282–3287, 1980.
9. H. N. Norton, *Handbook of Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

ANTONINO S. FIORILLO
University of Salerno

STRAIN SENSORS

In recent years the area of sensors has become increasingly important because of their varied applications in many areas. The term sensor is a broad terminology which encompasses a wide variety of devices. The present article deals with one such type of sensor, namely the strain sensor. The alternative term commonly used for strain sensor is strain gauge. Basically, a strain gauge is a device used to measure the linear deformation (mechanical surface strain) occurring in a material during loading. In addition to their fundamental use for measuring strains as such, strain gauges are also used for measuring other physical quantities such as pressure, load, displacement, torque, and so on by employing them as sensors in other measuring systems.

Historically (1), the development of strain gauges has followed different paths, and gauges have been developed based on electrical, mechanical, optical, acoustical, and pneumatic prin-

ciples. Among these, the electrical strain gauges have become so widely accepted that they now dominate the entire strain gauge field except for a few special applications. In its widest sense, the electrical strain gauge includes many varieties, utilizing the full range of electrical quantities, that is, resistance, reluctance, capacitance, inductance, and others. However, over the years the electrical resistance type strain gauge has become the most widely used device, and this is what is usually meant when the term strain gauge is used. In this article, the term strain gauge refers to the electrical resistance type strain gauge.

ORIGIN OF STRAIN GAUGES

The origin (1) of the strain gauge goes back to Robert Hooke (1635–1703) whose famous law states that, within certain limits, stress is proportional to strain. Later, Robert Young (1773–1829) provided a quantitative relation between stress and strain in a bar under simple tension (or compression) by his modulus of elasticity equation,

$$\sigma = E \times \epsilon \quad (1)$$

where E is the modulus of elasticity σ the stress and ϵ is the strain.

Poisson (1781–1840), a French mathematician, extended the laws of elasticity from uniaxial to two- and three-dimensional aspects which involved another well known material constant, now named Poisson's ratio. Although mathematicians of the last two centuries worked out a great deal of theory, it is comparatively only at a later stage that strain measurement has been done on a large scale. This situation undoubtedly is because of the difficulty of making precise quantitative measurements on metals whose elastic strains are extremely small.

CONCEPT OF STRESS AND STRAIN

All bodies can more or less be deformed by suitably applied forces. As a result of this, forces of reaction come into play internally. This is due to the relative displacement of its molecules. This tends to balance the load and restore the body to its original condition. The restoring or recovering force per unit area set up inside the body is called stress. The deformation or the change produced in the dimension of a body under the action of external forces is called strain. It is measured by the change per unit length (linear strain), per unit volume (volume strain), or the angular deformation (shear strain) depending on whether the change is along its length, volume, or the shape of the body.

If a bar of length L is subjected to a direct load W (Fig. 1), and the resulting change in length of the bar is ΔL , then the strain produced is given by Ref. 2

$$\text{Strain } (\epsilon) = \frac{\text{Change in length}}{\text{Original length}} = \frac{\Delta L}{L} \quad (2)$$

Strain is thus a measure of the deformation of the material and is nondimensional; it is simply a ratio of two quantities with the same unit. In practice, the extension of materials under

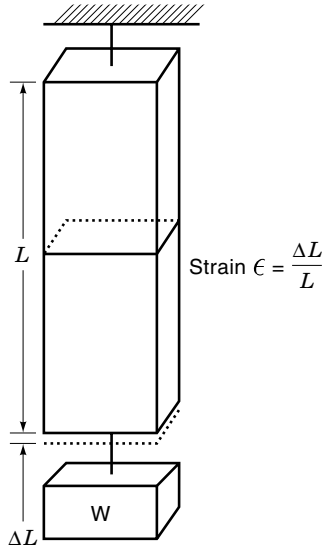


Figure 1. Schematic of the bar subjected to load.

load are very small. Hence it is convenient to measure strain in units of 10^{-6} that is, microstrain ($\mu\epsilon$).

BASIC OPERATING PRINCIPLE OF THE STRAIN GAUGE

The discovery of the basic operating principle (3) of the strain gauge dates back to 1856, when Lord Kelvin reported that certain metallic conductors subjected to mechanical strain exhibited a corresponding proportional change in electrical resistance. This property, namely change in resistance due to strain, is referred to as the piezoresistive effect. Generally the term *piezoresistive effect* is used in connection with semiconducting materials.

STRAIN SENSITIVITY/GAUGE FACTOR

In general, all electrically conducting materials possess strain sensitivity. The dimensionless number F is variously termed the electrical resistance–strain coefficient, the strain sensitivity factor, or the gauge factor and is expressed mathematically (3,4) as,

$$F = \frac{\Delta R/R}{\Delta L/L} \quad (3)$$

$$\text{since } \frac{\Delta L}{L} = \epsilon,$$

$$F = \frac{\Delta R/R}{\epsilon} \quad (4)$$

where R and L represent, respectively, the initial resistance and length, while ΔR and ΔL represent the small changes in resistance and length which occur as the gauge is strained. The gauge factor of a strain gauge is thus an index of the strain sensitivity of the gauge. The higher the gauge factor, the more sensitive the gauge and the greater the electrical output for indication or recording purposes.

The major milestones/events in the history of strain gauge development are indicated in Appendix 1.

FEATURES OF AN IDEAL STRAIN GAUGE

An ideal strain gauge should possess the following characteristics. It should:

- Have high strain sensitivity (gauge factor)
- Exhibit linear response to strain
- Have a very low temperature coefficient of resistance (TCR)
- Be insensitive to humidity and other ambient conditions likely to be encountered
- Have good temporal stability
- Be suitable for use as the sensor in other transducer systems where an unknown quantity is measured in terms of strain
- Have low hysteresis effect

Although in practice it is difficult to meet all the requirements of an ideal strain gauge, a great deal of effort has been expended in making a strain gauge having the characteristics close to that of an ideal one.

GENERAL CLASSES OF STRAIN GAUGES

Broadly the types of strain gauges developed over the years are

- Unbonded—wire gauges
- Bonded—wire gauges
- Bonded—foil gauges
- Semiconductor gauges
- Thin film strain gauges

The unbonded wire strain gauge consists of a strain-sensitive wire mounted on a mechanical frame whose parts can have slight movement with respect to each other. This relative movement causes a change in tension of the wire resulting in a change in electrical resistance. The electrical resistance change produced is a measure of the relative displacement or strain. This type of gauge can be made of entirely inorganic and high temperature materials, so that operation of such sensors is possible even in high dose radiation and high temperature environments.

The bonded wire/foil strain gauge also consists of a strain-sensitive wire or foil, but is entirely attached by an adhesive to the member (component) whose strain is to be measured. As the strain-sensitive wire or foil are basically electrically conducting, they have to be electrically isolated from the component (especially if the component is made of conducting material). Usually the required level of electrical insulation in these gauges is provided by the adhesive and/or insulating backing material. The commonly used adhesive and backing materials are of the phenolic or epoxy resin type, some of which may resist the use of the bonded gauges in radiation and very high temperature environments. Also, as the adhesive and backing material are in the force path, the accuracy of measurement is limited by the characteristics of these materials. Normally the force required to produce the displacement in bonded wire/foil type gauges is larger than that required in the case of an unbonded wire gauge because of the

additional stiffness of the member. The bonded wire/foil type gauge can operate in tension, compression, or bending mode. Bonded foil type strain gauge is a by-product of precisely photo-etched printed electronic circuits by photolithography process. Foil gauges have undergone intensive development and are available in many grid configurations (Fig. 2) and a variety of different foil alloys. Foil gauges are thinner than wire gauges. It is easier to make complicated grid designs with these gauges and they conform more easily to curved surface components.

Generally, wire or foil type gauges made of metal and metal alloys exhibit a gauge factor value typically about 2.0 to 5.0. They are basically low signal devices and require signal amplification. Although their signal output is small, their linearity and hysteresis characteristics are generally good.

Semiconductor gauge development resulted from the intensive research in solid-state physics relating to semiconductors. Semiconductor gauges offer higher sensitivity than wire and foil type gauges. By controlling the type and amount of dopant in the semiconductor material, the strain gauge properties can be controlled effectively making them suitable for specific applications. In the case of semiconductor gauges, gauge factors of about 10 to 120 are typical, either positive or negative. Although the gauge factor values are numerically large, they are not linear and greatly affected by temperature variations above 70°C. However, because of their high sensitivity they are attractive for detecting very small strains.

As indicated earlier, the resistance change due to strain in the case of semiconductors is generally referred to as the piezoresistive effect. On the other hand, piezoelectric type strain sensors convert mechanical strain into electrical output and are based on the piezoelectric effect observed in certain nonmetallic and insulating dielectric compounds. However, these types of strain sensors are not usually preferred for static strain measurements. In cases of strain monitoring over a prolonged period of time, better accuracy and stability can be achieved with strain gauges made of metallic alloy materials.

An important development in the field of strain gauge technology is the introduction of thin film strain gauges (5). These gauges can be made of any desirable resistor metal, metal alloy, semiconductor, or a combination of metal and dielectric (called cermet which are basically metal-dielectric composites). Thin film strain gauges are prepared mainly by vacuum deposition processes. These techniques provide the greatest flexibility to control the major strain gauge properties. It is possible to optimize specific properties by controlling the deposition conditions such as pressure, temperature, rate of deposition, and so on. Significant research has been reported in the literature regarding the evaluation of the strain gauge properties of various materials in thin film form. The extensive effort in this direction resulted in the development of special alloy/cermet films which exhibit the necessary stability, gauge factor, and resistance characteristics. Figure 3 shows the broad classification (6,7) of the strain gauges.

STRAIN SENSITIVITY DEPENDENCE

In order to have an idea of how the strain sensitivity of the material depends on other basic parameters (7), we can consider a conductor of uniform cross-sectional area A and length

L , made of a material with resistivity ρ . The resistance of such a conductor is given by

$$R = \frac{\rho L}{A} \quad (5)$$

Considering all the parameters in this equation as variables, if we differentiate and substitute in the equation for gauge factor, we obtain (after simplification),

$$F = \frac{\Delta R/R}{\Delta L/L} = [1 + 2\nu] + \left[\frac{\Delta \rho/\rho}{\Delta L/L} \right] \quad (6)$$

where ν is the Poisson's ratio of the material. In Eq. (6), the term $(1 + 2\nu)$ represents purely a geometrical effect of deformation. The term $(\Delta \rho/\rho)/(\Delta L/L)$ represents a physical effect, namely the change in specific resistance with elastic strain (which is related to number and mobility of free electrons). In fact, in metals the dimensional change (or the geometrical effect) is the major factor, whereas in semiconductors the resistivity change is predominant. Vacuum deposited thin film materials may have additional contributions to their resistivity terms, because their structure (especially grain size and separation of grains) can be significantly different from that of the common bulk material specimens.

It is important to bear in mind that a high gauge factor is not the only criterion for selecting a suitable material for fabrication of strain gauges. In addition, the material must also possess low TCR values and exhibit excellent thermal and temperature stability. The data on the strain sensitivity of commonly used strain gauge materials (metals and alloys) are provided in Table 1. Table 2 contains data of different classes of thin film materials relevant to strain gauges.

USE OF THE WHEATSTONE BRIDGE

The Wheatstone bridge is one of the most common configurations used with strain gauges. This is because the resistance change of strain gauge is very small, and precise instrumentation is required to measure it accurately. Figure 4 shows the Wheatstone bridge in its simplest form. The condition for bridge balance is,

$$\frac{R_1}{R_4} = \frac{R_2}{R_3} \quad (7)$$

Any change in resistance of the arms will unbalance the bridge and produce a voltage V_{out} across the output terminals. In strain gauge related instrumentation, usually each arm of the bridge is used as a transducer or strain gauge. The general output Eq. (4) used in these cases is the following.

$$V_{out} = \frac{F \epsilon N V_{in}}{4} \quad (8)$$

where

F = Gauge factor

V_{in} = Bridge input voltage

ϵ = Strain

N = Number of active arms of the bridge

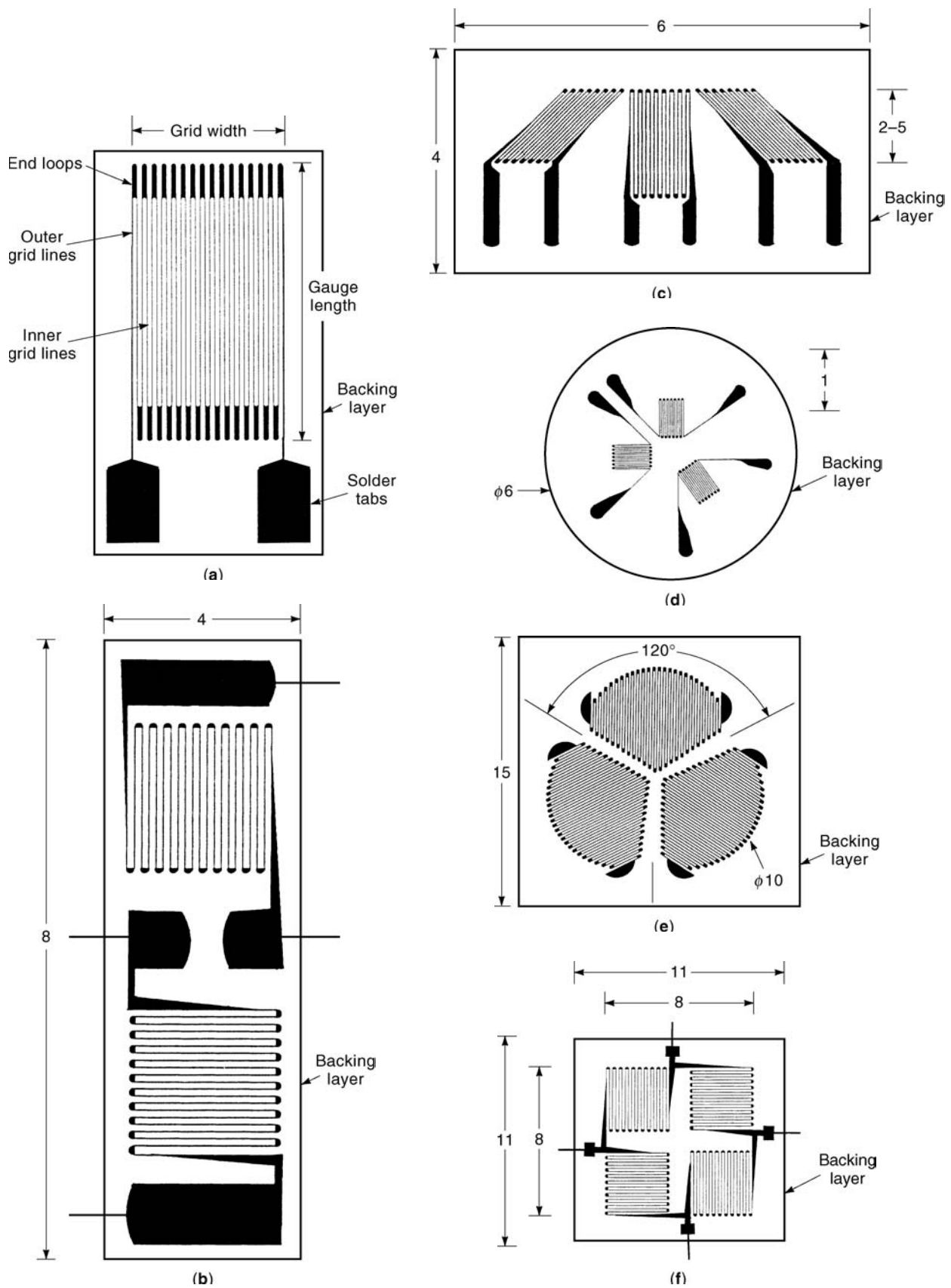


Figure 2. Foil type strain gauge configurations. (a) Single element. (b) Two element. (c), (d), (e) Three elements. (f) Four elements. Note: (1) All dimensions are in mm; (2) backing layer thickness 0.010 mm to 0.020 mm; (3) grid dimensions vary according to the type of gauge and application.

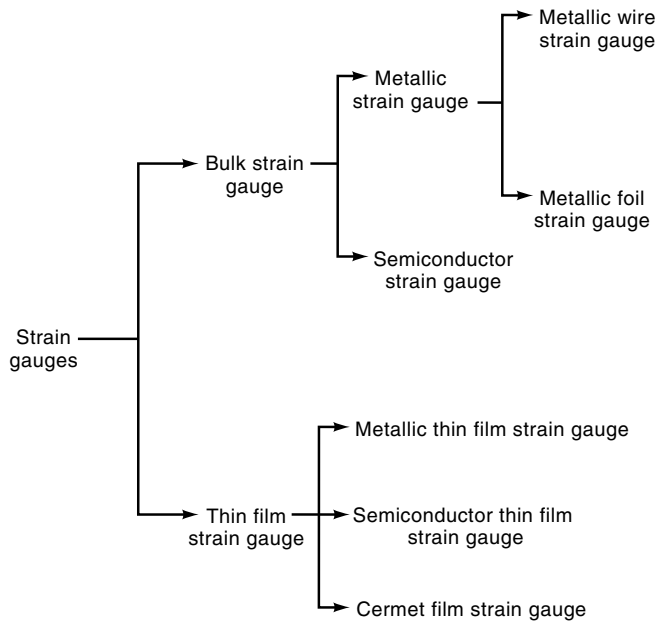


Figure 3. Broad classification of strain gauges.

The bridge output V_{out} obtained can be suitably amplified and processed. Details on bridge output voltage measurement, variations of wheatstone bridge configurations, bridge excitation, and associated aspects can be seen in Refs. 3, 4, 6, and 9–14.

Temperature Effects

Temperature is an important interfering input for strain gauges since resistance of strain gauges changes with both strain and temperature. Also, the material on to which the strain gauges are bonded/deposited will expand or contract with change in temperature. This causes an additional error resulting in apparent strain. Therefore, in order to carry out accurate strain measurements, temperature compensation must be employed. Several methods are available to compensate for temperature effects. One such method is to use a dummy gauge which is identical to the active gauge (in the Wheatstone bridge configuration) bonded/deposited on to a piece of similar

Table 2. Various Classes of Thin Film Materials and Their Characteristics Relevant to Strain Gauges (From Ref. 8)

Type and Class Film Material	Gauge Factor F	Temperature Coef cient of F (ppm/K)	Thermal Stability
Continuous, metals	2	400	Fair
Discontinuous, metals	Up to 100	~ -1000	Poor
Continuous, metal alloys	2	~ 100	Good
Discontinuous, metal alloys	Up to 100	~ 1000	Poor
Cermets	Up to 100	200–1000	Generally Good
Semiconductors	Up to 100	~ -1500	Good in some cases

Reproduced from Strain Gauge Technology, A. L. Window and G. S. Holister, (eds.), M/S Applied Science Publishers Inc. USA, 1982. Copyright © Kluwer Academic Publishers. Reproduced by permission.

material maintained at the same temperature. The dummy gauge and active gauge are placed in adjacent arms of the Wheatstone bridge, so that the resistance change due to the temperature and differential thermal expansion will have no influence on the bridge output voltage. Although in theory this is a simple and effective way of compensating, in practice because of inevitable differences from gauge to gauge and the fact that temperature of strain sensors is never precisely the same the inference is that it is possible to achieve superior performance with strain gauges having very low temperature co-efficient of resistance. It is important to note that, for the purpose of achieving very low temperature co-efficient of resistance for the strain gauges, thin film materials technology offers greater flexibility.

Another approach of temperature compensation involves the use of special gauges whose thermal properties are matched to the particular materials on which they are to be mounted—called self-temperature compensated (STC) gauges. STC gauges include those gauges made up of two parts; one with positive response to temperature and the other having negative response, and are so proportioned that the positive and negative responses essentially cancel each other over a given temperature range. STC gauges also include single element gauges made of various metallic or alloy materials which

Table 1. Strain Sensitivity of Various Materials (From Ref. 4)

Material	Trade Name	Typical Strain Sensitivity
Copper–nickel(55–45)	Constantan	+2.1
	Advance	
Nickel–chromium(80–20)	Nichrome V	+2.2
Nickel–chromium(75–20) plus iron & aluminium	Karma	+2.1
Iron–chromium–aluminium (70–20–10)	Armour D	+2.2
Nickel–chromium–iron –molybdenum (36–8–55.5–0.5)	Isoelastic	+3.5
Platinum–tungsten(92–8)	—	+4.0
Copper–nickel–manganese (84–4–12)	Manganin	+0.6
Nickel	—	–12.0
Iron	—	+4.0

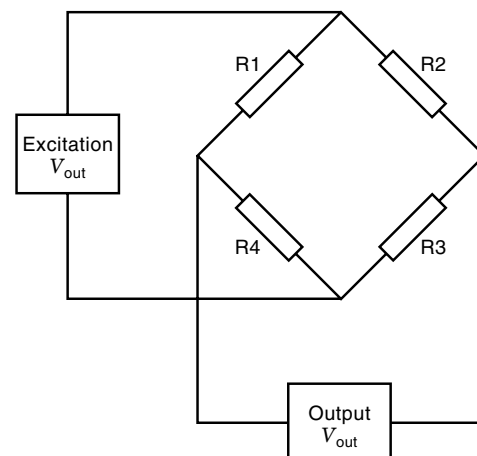


Figure 4. Wheatstone bridge configuration.

have been classified according to their temperature characteristics. Another variation of STC gauges is to produce a sensing element alloy which, when subjected to certain heat treatment and mechanical processing, will match the properties of the material on which the gauges are to be used.

For several applications the STC gauges are adequate and save the cost of an additional gauge and its associated installation and wiring. In situations when self compensation is not good enough, for example at higher temperatures, the bridge compensation with external compensation network approach can be employed. Detailed information on temperature compensation can be found in Refs. 3, 4, 6, 9, and 13–18.

Transverse Sensitivity

Ideally, a strain gauge should respond to strains of a specimen along a specific direction. But most strain gauges exhibit some degree of sensitivity to strains along directions other than the one to be measured. The transverse sensitivity of strain gauges refers to the behavior of gauges in responding to strains which are perpendicular to the primary sensing axis of the gauges. Normally, strain gauges have very low response to transverse strains. Therefore, the errors in strain measurement due to transverse sensitivity of strain gauges are generally quite small. However, if utmost accuracy in strain measurement is needed, then transverse sensitivity of the gauges must be taken into account. Also, the effects of transverse sensitivity should necessarily be considered in the experimental stress analysis of a biaxial stress field using strain gauges.

In fact, one of the important aspects of strain gauge technology is that in many applications both the magnitude and direction of the strain need to be measured. In such cases, the information on the directional sensitivity (both longitudinal and transverse sensitivity) of the gauges will be very helpful. More detailed aspects of transverse sensitivity (including the mathematical formula) and related information can be found in Refs. 3, 4, 19, and 20.

THIN FILM TECHNOLOGY FOR STRAIN GAUGES AND STRAIN GAUGE BASED SENSOR DEVELOPMENT

Although foil gauges are being used widely, in recent years thin film strain gauges and thin film strain gauge based transducers are gaining increasing popularity because of their several distinct advantages (1,21,22). Some of the important advantages (in comparison with the bonded foil/wire gauges) are, (1) elimination of the glue between the strain gauge and the straining member, (2) easier thermal compensation with minimal interference with the mechanical properties of the component material, (3) larger operating temperature range, (4) mass production with considerable cost reduction, and (5) complete freedom from external strain gauge suppliers. During the last decade, a number of companies have started adopting thin films technology for strain gauge transducers development. This clearly indicates that thin film techniques will play a leading role in strain gauge based device technology.

In view of this, a concise description of thin film deposition techniques is given next.

Thin films can be deposited by a variety of methods (23–27). The important techniques commonly employed (useful for strain sensors development) may be broadly classified into two

categories, namely, physical vapor deposition and chemical deposition methods.

PHYSICAL VAPOR DEPOSITION

The term physical vapor deposition denotes those vacuum deposition processes such as evaporation and sputtering where the coating material is passed in to vapor transport phase by physical mechanisms, that is, evaporation, sublimation, or ion bombardment.

Thermal Evaporation

In thermal evaporation, the material to be deposited (evaporant) and the substrates on which it is to be coated are placed in vacuum. The evaporant is loaded in a heating element. The required vaporization temperature is achieved by resistance heating of the filament or boat, which conducts heat to the evaporant. At that point, the material evaporates and coats everything in its vicinity. The subsequent condensation process, consisting of nucleation and film formation, is strongly dependent on the thermal energy, the rate of arrival and the incident angle of the vapor particles, as well as the physical, chemical, and thermal conditions of the receiving surface. The resistance heating approach is a very convenient method and is widely used.

E-beam Evaporation

Some materials cannot be used as evaporants as they have high melting points or because they will react with any material used to support them in the chamber, making the deposited coating impure. Many of these materials, however, can be evaporated from an electron beam gun. In E-beam evaporation (28), the material is heated by electron bombardment. By controlling the kinetic energy of the electron beam, the depth of the melt area can be controlled. As a result, the molten evaporant material does not come into contact and alloy with the crucible (supporting material). Scanning the E-beam over the surface of the melt prevents the nonuniform deposition that would otherwise occur due to the formation of cavity in a molten source. A schematic of the general vacuum evaporation system is shown in Fig. 5.

Flash Evaporation

If the evaporant is not a pure element, it will fractionate upon heating. This makes it difficult to control the stoichiometry of the deposited film. In such situations, flash evaporation, in which a fine-sized powder of the alloy is dropped at a controlled rate on to a hot 'boat', is employed. The temperature of the boat is held well above the minimum vaporization temperature of the individual elements of the alloy; thus the arriving powder grain instantly flashes off (totally vaporizing) without fractionation.

Sputtering

Besides thermal and E-beam evaporation, vapor species may also be created by knocking out the atoms or molecules from the surface of a solid material by bombarding it with energetic ions. The ejection process, known as sputtering, occurs as a result of momentum transfer between the impinging ions and the atoms

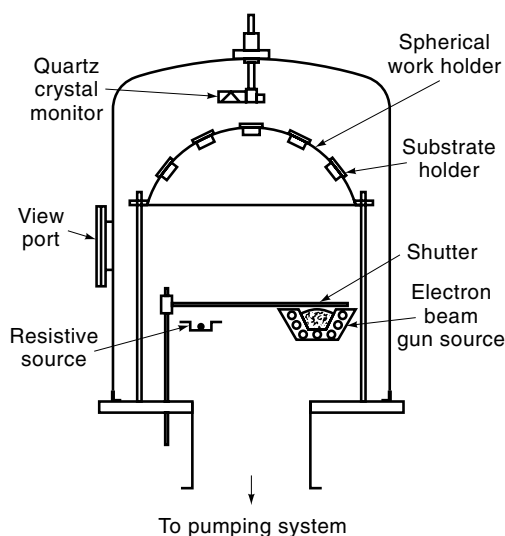


Figure 5. A schematic of the general vacuum evaporation system.

of the target being bombarded. The sputtered species can be condensed on a substrate to form a thin film. Many different materials (including alloys or compounds) can be sputtered.

Ions for sputtering may be produced by establishing a glow discharge between the target and the substrate holder. This is referred to as glow-discharge sputtering. However, in case of ion-beam sputtering, a separate ion source is utilized. Depending on the geometry of the target-substrate system and the mode of ion transport, a large number of sputtering variants have been developed (29,30). These are briefly discussed next.

Direct Current (dc) Diode Sputtering

In this arrangement, a plasma discharge is maintained between the anode (substrate) and the cathode (target). The chamber is evacuated to about 10^{-6} torr and then backfilled to the sputtering pressure with an inert gas, usually argon. A potential applied to the electrodes gives enough energy to the electrons to ionize the argon molecules, creating a plasma. The ions in the plasma near the cathode get accelerated across the potential drop of the cathode dark space, and hit the target with enough energy to eject target atoms. These target atoms spray in all directions and coat everything in their path, including the substrates. There are three factors that characterize dc sputtering with planar diode arrangement: (1) the cathode current densities and sputtering rate are low, (2) the working pressures are high, and (3) the substrates are in contact with the plasma. A schematic of the dc diode sputtering system is shown in Fig. 6.

If the partial pressure of contaminants is higher than about 10^{-5} torr, sputtering can still be induced but the deposited films will be less pure. Oxygen, nitrogen, water vapor, and other constituents of the background gas will be incorporated in the films as they are deposited. If the base pressure is low and argon is pure, the only gas present will be argon. Very little argon will be incorporated in the films, however, because argon, as an inert element, has a low sticking co-efficient. Any incorporated argon will not form compounds with the target atoms. It is also unlikely that it will alter the properties of the deposited films to any great extent.

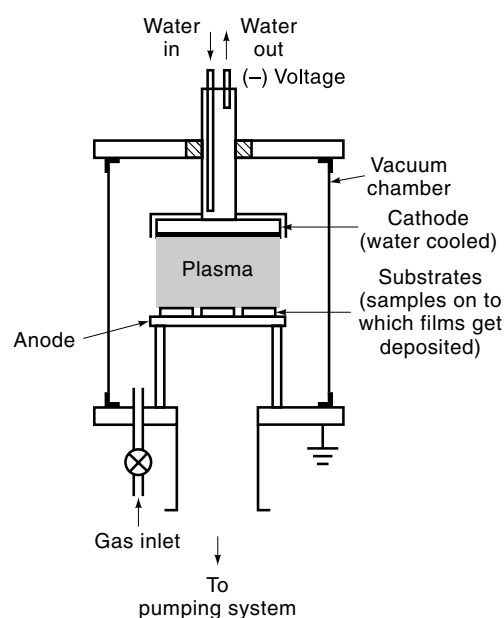


Figure 6. A schematic of the dc-diode sputtering system.

In a diode system, bombarding ion current density and bombarding ion energy cannot be varied independently because they both depend on the cathode potential. This inflexibility occasionally presents a problem. This problem can be overcome by using a triode system.

Triode Sputtering

In this configuration, sputtering rates are increased by supplying auxiliary electrons from a thermionically emitting filament. Both the total ionization and the ionization efficiencies are increased by accelerating the electrons by means of a third electrode and injecting them into the plasma. Consequently, triode sputtering system can be operated at low pressures. With this arrangement, the bombarding ion current density and the bombarding ion energy can be varied independently. The bombarding ion current density (the plasma density) can be adjusted by varying either the temperature of the thermionic cathode or the anode potential, whereas the bombarding ion energy can be adjusted by varying the negative potential on the target (Cathode).

Magnetron Sputtering

In magnetron sputtering system, the ionization efficiency of the electrons is increased by increasing their path length by applying a transverse magnetic field normal to the electric field. Magnetron sputtering makes it possible to utilize the cathode discharge power very efficiently to generate high current densities at relatively low voltages and lower operating pressure to yield deposition rates which are higher than those in the nonmagnetron sputtering systems.

BIAS SPUTTERING

The term *bias sputtering* is used to refer to the specific process of maintaining a negative bias on substrates during sputter deposition. In this case, the film is subjected to steady ion

bombardment throughout its growth, which effectively cleans the film of adsorbed gases otherwise trapped in it as impurities.

Radio Frequency Sputtering

Direct current methods cannot be used to sputter insulating targets due to the buildup of positively charged sputtering gas ions which repel the bombarding (sputtering) ions. This difficulty can be overcome by using radio frequency (RF) sputtering. In RF sputtering a high frequency alternating potential is used to neutralize surface charges periodically. RF sputtering apparatus can be used to deposit conducting, semiconducting, and insulating films. Therefore RF sputtering has found wide applications.

Ion Beam Sputtering/Deposition

This is a relatively newer technique. Ion beam sputtering permits independent control over the energy and current density of the bombarding ions. Ion beams are used for thin film deposition in a variety of configurations (31). Compared with other thin film deposition techniques, ion beams provide a controlled, collimated flux of energetic particles that may be directed at the substrate, a target material, or a growing film.

Ion Plating

Ion plating is the result of the combination of vacuum evaporation and sputtering. In this arrangement, the source of evaporation is placed in a vacuum chamber. Opposite to this source is placed a substrate holder. The high voltage applied to the substrate generates a discharge (plasma). When the evaporation source emits vapors, the vapor passes through a glow discharge on its way to the substrates. Ion plating technique combines certain advantages of both evaporation and sputtering.

CHEMICAL METHODS

Electrodeposition and chemical vapor deposition are the two important techniques that come under this category. These methods have a limited and specific usage. Chemical methods (32) require simple equipment, and thus may be more economical. These methods, however, are often complex and difficult to control. Also, some of these techniques demand that the supporting substrate withstand high temperature; others require that substrates be exposed to various solvents. A brief description of these methods is given next.

Electrodeposition is done in three ways, namely, electrolytic deposition, electroless deposition, and anodization. In the electrolytic deposition, two electrodes and a suitable electrolyte to pass a current are required. The deposition rate is dependent on the temperature of the electrolyte and the geometry of the cathode including other parameters.

In electroless deposition, the external potential source is replaced by a chemical reduction process. The deposition rate is highly affected by the temperature of the reaction which is rather difficult to control.

The production of a coating of metal oxide or metal hydroxide by the electrochemical oxidation of a metal anode in an electrolyte is called anodization. It is achieved by maintaining constant current or constant voltage. The sticking of oxide films on the parent metal depends on the nature of the metal. These metals are often referred to as "valve metals" because of rectify-

ing characteristics of their anodic oxides. The anodic films are invariably amorphous in nature, but crystalline structure may be obtained by suitably adjusting the conditions of anodization.

Chemical Vapor Deposition (CVD)

In this method, a volatile component of coating material is thermally decomposed, or it reacts with other vapors in the vapor phase or at the hot substrate surface so that the reaction product is deposited as a thin film.

Plasma Chemical Vapor Deposition

This method is also known as plasma assisted CVD (PACVD). In this technique for producing the glow discharge, radio frequency energy is used. Because the activation energy is provided by the plasma and not by heat, films can be produced at lower temperatures than with standard thermally activated atmospheric CVD (APCVD) and low pressure CVD (LPCVD).

Almost all the methods just outlined are useful for preparing the thin film strain gauges and strain gauge based transducers. In some cases of transducer development, more than one thin film deposition technique needs to be adopted.

METHODS TO DETERMINE THE GAUGE FACTOR OF STRAIN GAUGES

Normally the gauge factor of the commercially available strain gauges will be specified by the supplier along with the other parameters such as gauge resistance, TCR, STC number, grid dimensions, backing material, temperature range, fatigue life, maximum strain limit, and creep property. However, if the user desires to have an assessment about the suitability of the gauge for the specific practical applications and also to determine the gauge factor and resistance-strain characteristics of in-house developed gauges (especially true in the case of thin film strain gauges), it is necessary to measure the strain values by some other known standard method. A brief description of these methods is given here.

The methods to measure the gauge factor as well as the resistance-strain characteristics of the strain gauges/strain gauge materials can be broadly divided into (1) mechanical and (2) optical techniques. The mechanical methods are essentially based on the bending of beams in different configurations. They include:

- Four-point bending beam arrangement
- Beam supported at both ends
- Cantilever beam arrangement
- Beam bent over hard metal rods

In the four-point bending beam arrangement (33), the beam is held between four rolling pins, two at the top and two at the bottom (Fig. 7). This allows the application of the equal and opposite couples to both ends of a beam. As a result, the beam is subjected to pure end moments. The strain experienced by the strain gauge bonded to the surface of the beam at its center can be calculated, which involves measuring the maximum deflection at the center of the beam. This deflection measurement is possible by the use of a dial gauge or linearly variable differential transformer (LVDT).

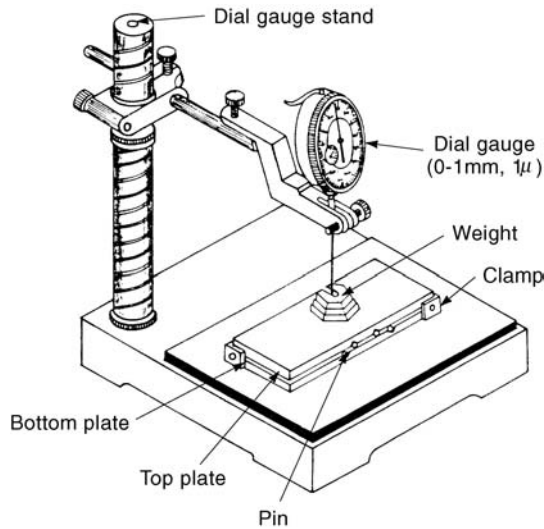


Figure 7. Schematic of the four-point bending set-up (from Ref. 33).

In the case of beam supported at both ends, it is deflected by applying a force at its center (34). The strain experienced by strain gauge that is bonded on a convex surface of the beam can be calculated by measuring the thickness, length, and deflection at the center of the beam. As already mentioned, in this case the maximum deflection of the beam at its center also can be measured using either dial gauge or LVDT.

In the cantilever technique (35), a bending moment is applied to the beam by fixing one end and loading the other end with the weights (Fig. 8). Due to loading, a strain gauge cemented to the beam at a typical distance from the fixed end experiences a strain which can be calculated by knowing the dimensions of the beam, the Young's modulus of the material of the beam (E) and the weight (W) applied at the free end of the beam. It is important to note that while measuring the length of the beam, it is the length from the center of the gauge to the point of application of the load (W) which has to be taken into account for calculating the strain (ϵ).

In an arrangement (Fig. 9) in which the beam is bent over a hard metal rod (36) the strain experienced by the strain gauge (cemented at the top surface of the beam) can be calculated by measuring the thickness (t), length (L), and deflection (d) of the beam. In this case, the maximum deflection is obviously equal to the diameter of the rod on which the beam is bent. In order

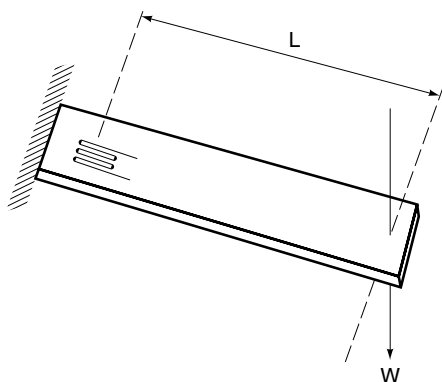


Figure 8. Cantilever set-up.

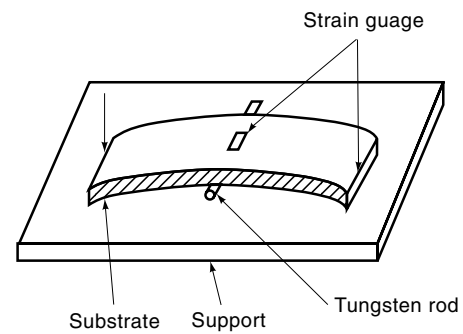


Figure 9. Schematic of the beam bent over a hard metal rod (from Ref. 36).

to subject the strain gauge to different strain values, rods of different diameters are needed.

The optical methods which are adopted to estimate the value of the strain are based on interference and diffraction phenomenon. One such method which can be employed is the diffraction method.

Diffraction Method

In this method (37), a slit with independent jaws is cemented to the test member (say a metal bar) such that its jaws are parallel as shown in Fig. 10. A laser beam is made incident on the slit. The slit diffracts the beam into various orders which are observed on the screen located at a suitable distance. Any change in the slit width due to loading will result in the corresponding change in the diffraction pattern. A tensile strain will contract the pattern, whereas a compressive strain will elongate it. Hence, the strain experienced by the strain gauge (bonded to the test member) due to loading can be calculated by making measurements on the change in the diffraction pattern produced.

APPLICATION OF STRAIN GAUGES IN OTHER MEASURING SYSTEMS

Strain gauges find application as sensors in other measuring systems such as pressure transducers, load cells, displacement measuring devices, accelerometers, position sensors, stress-strain analysis systems, and so on. A brief description of some of these measuring devices is given in what follows.

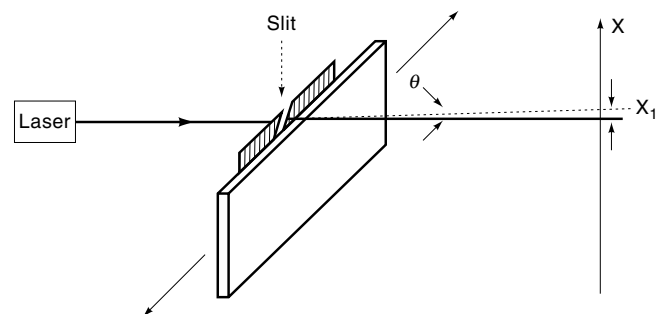


Figure 10. Schematic of the diffraction set-up to estimate the value of strain. From Sirohi R. S. and Radhakrishna H. C., *Mechanical Measurements*, 3/e, Copyright © M/S Wiley Eastern, Ltd. Reproduced by permission.

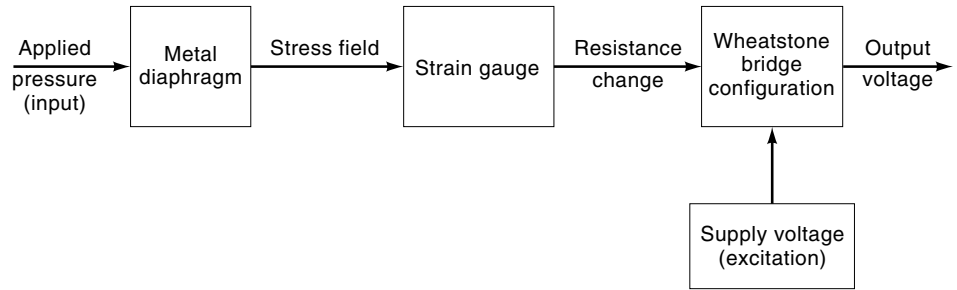


Figure 11. Block diagram of the principle of strain gauge pressure transducer.

PRESSURE TRANSDUCER

Pressure transducers are basically the electromechanical devices which are useful for a number of applications. Typical applications of pressure transducers include measurement of pressure in process industries, automobile engines, depth study in oceanography, wind tunnel experiments, gas turbine systems, hydraulic systems, nuclear propulsion systems, mining safety, and nuclear and aerospace applications. The primary function of the pressure transducer is to sense fluid pressure and provide an electrical output proportional to the input pressure. A pressure transducer essentially consists of an elastic element such as a metal diaphragm which undergoes deformation due to applied pressure. This mechanical deformation of the diaphragm is converted into an electrical response by a strain gauge bonded to it. Schematically, the functioning of the pressure transducer is shown in Fig. 11. There are three types of pressure transducers, namely, absolute, relative (gauge), and differential pressure transducers (Fig. 12).

Absolute Pressure Transducer. This measures pressure referenced to vacuum, hermetically sealed at about 10^{-5} m bar of Hg. When the pressure port is exposed to the atmosphere, the transducer will indicate atmospheric pressure; approximately 760 mm of Hg or 1 bar. This occurs because there is a vacuum on one side of the diaphragm and atmospheric pressure on the other.

Relative or Gauge Pressure Transducer. This measures pressure referenced to local atmospheric pressure and is vented to the atmosphere. When the pressure port is exposed to the atmosphere, the transducer will indicate 0 mm of Hg or 0 bar. This occurs because the pressure on both sides of the diaphragm is the same and there is no net output. Venting is accomplished by means of a small hole located near the transducer's electrical termination-connector/cable. The vent hole contains a porous, stainless steel disk designed to filter out harmful air-borne particles from entering the transducer in order to

safeguard the strain gauges from contamination, corrosion, and hence resistance/output variation.

Differential Pressure Transducer. This measures pressure differential between two pressure P_1 and P_2 as shown in Fig. 12. When both the pressure ports (P_1 and P_2) are exposed to the atmosphere, the transducer will indicate 0 mm of Hg or 0 bar. In other words, if the pressures P_1 and P_2 are the same, the net output is 0 bar. If they are not the same, then the net output will be a reading other than 0 bar.

Application of pressure results in deformation of the sensing element (diaphragm or other type of elastic sensing elements) on to which strain gauges are bonded and wired in the Wheatstone bridge configuration. The change in the output of the bridge is related to the magnitude of the pressure. Since the resistance change of the strain gauge is a function of surface strain, this strain is directly related to the applied pressure. Hence, strain gauges form an important component of the pressure transducers.

A cross-sectional view of the complete absolute type strain gauge pressure transducer assembly is shown in Fig. 13. Either

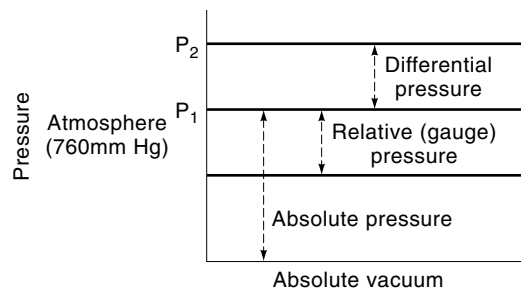
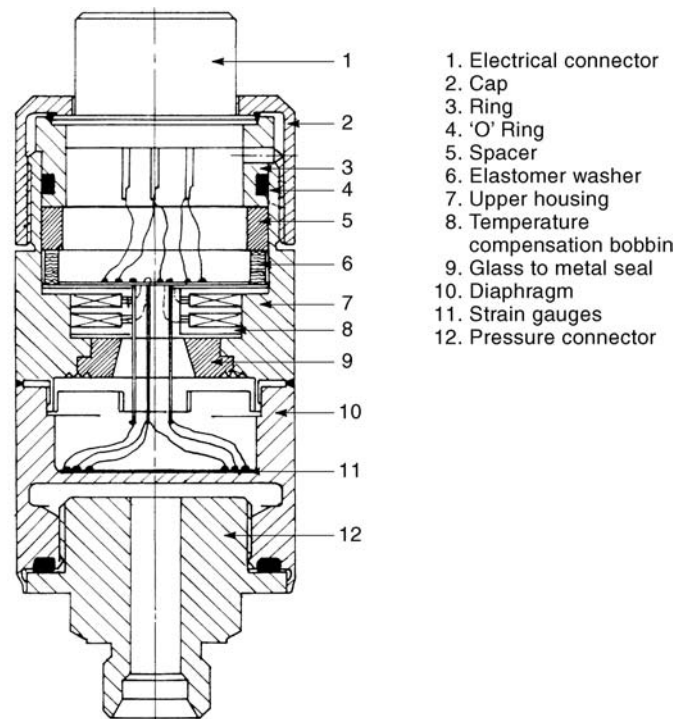


Figure 12. Types of pressure transducers.



1. Electrical connector
2. Cap
3. Ring
4. 'O' Ring
5. Spacer
6. Elastomer washer
7. Upper housing
8. Temperature compensation bobbin
9. Glass to metal seal
10. Diaphragm
11. Strain gauges
12. Pressure connector

Figure 13. Cross-sectional view of pressure transducer assembly (from Ref. 33).

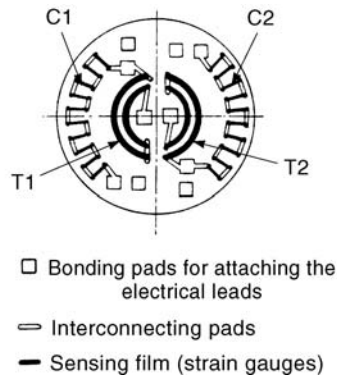


Figure 14. Schematic of the thin film strain gauge pattern deposited on the pressure transducer diaphragm. Reproduced from Performance study of pressure transducer with meandering—path thin film strain gauges, M. M. Nayak, K. Rajanna, and S. Mohan, *Thin Solid Films* 193/194 (1990), P. 1023–1029. Copyright © Elsevier Science. Reprinted with permission.

foil type strain gauges or thin film strain gauges can be utilized in these transducers. The use of thin film strain gauges for the measuring systems of this type have the additional advantage that the gauges can be directly deposited (with a dielectric film for insulation) on the diaphragm. This process enables elimination of the likely limitation of accuracy in the case of foil type gauges, because of the presence of adhesive and backing material. A schematic diagram of the thin film strain gauge pattern (38) deposited on the diaphragm is shown in Fig. 14. It

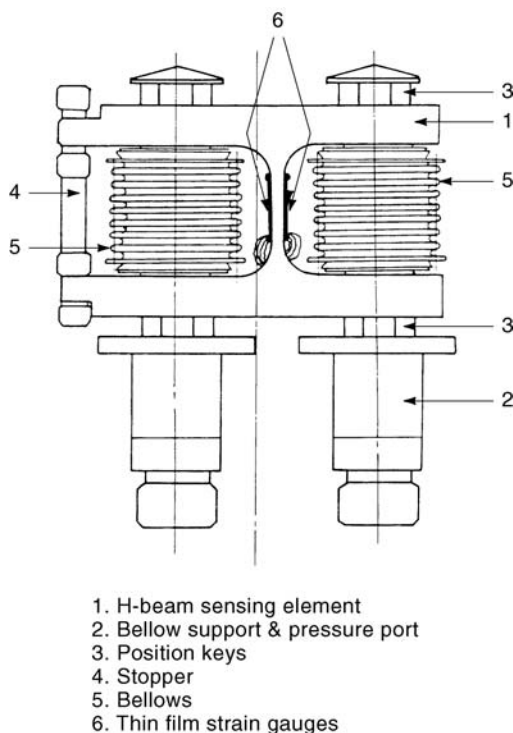


Figure 15. Strain gauge based differential pressure transducer assembly. Reproduced from Sputtered thin film strain gauges for differential pressure measurement, M. M. Nayak et al., *IEEE Trans. Instrum. and Meas.*, 45 (1) February 1996, P 335–339. Copyright © IEEE, Inc. Reprinted with permission.

is possible to obtain the required strain gauge pattern by using precision mechanical masks or photolithography technique (especially for very fine line patterns). Referring to Fig. 14, it is important to note that the location of the strain gauges is such that the gauges C_1 and C_2 at the diaphragm edge experience a compressive strain and those near the center (T_1 and T_2) undergo tensile strain. All the four gauges are made active by connecting them in Wheatstone bridge configuration. Gauges C_1 and C_2 experience compressive strain whose resistance decreases with pressure will form one opposite set of arms. The strain gauges whose resistance increases with pressure form the other set. Pressure transducers are normally calibrated using standard dead weight pressure calibration set-up.

Suitability of the pressure transducers for a specific application can be assessed from their general specifications as well as output performance characteristics such as variation of output with pressure at different excitation voltages, nonlinearity and hysteresis, stability, repeatability, temperature effects, and so on.

Similar to absolute and gauge pressure transducers, *differential pressure transducers* are made using strain gauges. A differential pressure transducer gives an output with increasing difference between two pressures, both of which may vary. Normally, the lower or less varying pressure is termed as the reference pressure or the line pressure and the other pressure is called the measured pressure. When the measured pressure is always higher than the reference pressure, the transducer has a unidirectional range. When the measured pressure is either lower or higher than the reference pressure, the transducer is said to have bidirectional range. However, in either case, the measurement of differential pressure is of great value. Figure 15 shows the schematic of the strain gauge based on

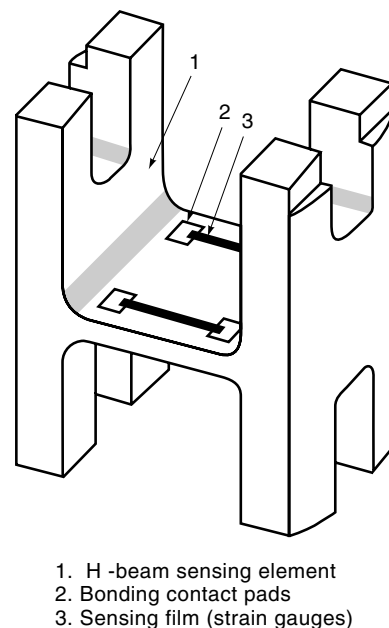


Figure 16. Schematic of the thin film strain gauges on the H beam. Reproduced from Sputtered thin film strain gauges for differential pressure measurement, M. M. Nayak et al., *IEEE Trans. Instrum. and Meas.*, 45 (1) February 1996, P 335–339. Copyright © IEEE, Inc. Reprinted with permission.

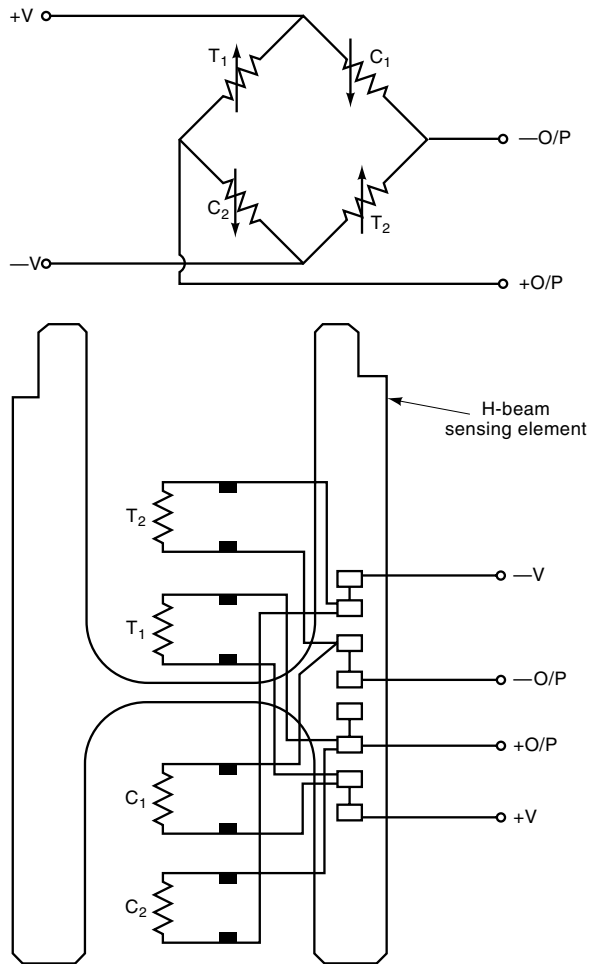


Figure 17. Four active gauges on the sensing element of the differential pressure transducer connected in the Wheatstone bridge configuration.

differential pressure transducer assembly (39). It essentially consists of an H beam configuration with a set of bellows as a sensing element. The two thin walled bellows used on either of the H beams convert the pressure difference into a linear displacement.

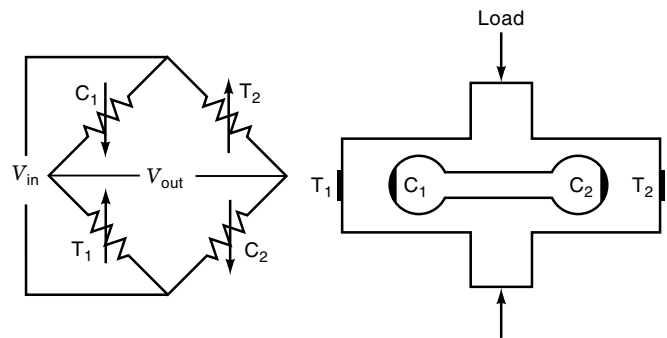


Figure 19. Binocular type sensing element configuration for load cell.

As in the case of an absolute or gauge pressure transducer, in this case foil type or thin film strain gauges (two gauges on either side of the beam) also can be adopted for converting the linear displacement into a proportional electrical output. Figure 16 shows the schematic of the thin film strain gauges deposited on the H-beam sensing element. The Wheatstone bridge configuration with all the four gauges active is shown in Fig. 17. Calibration of the device can be done using a standard differential pressure calibration system (Fig. 18).

LOAD CELLS

Basically the load cells are the force transducers which are used for force or thrust measurement and weighing purposes. Like pressure transducers, load cells can be made using strain gauges (3,40,41). Typical common areas of applications of strain gauge load cells include on-board weighing for trucks, postal and shipping scales, crane and laboratory weighing systems, agricultural applications, thrust measurement in static testing of rocket motors, high altitude testing systems, and others.

The construction of strain gauge load cells is based on three types of strain fields, namely bending, shearing, and direct stress. Accordingly, the different types of sensing elements are adopted in load cells. Some of the commonly used sensing element configurations are hollow cylinder, slotted-cylinder, binocular type, ring type, wheel-shaped configuration,

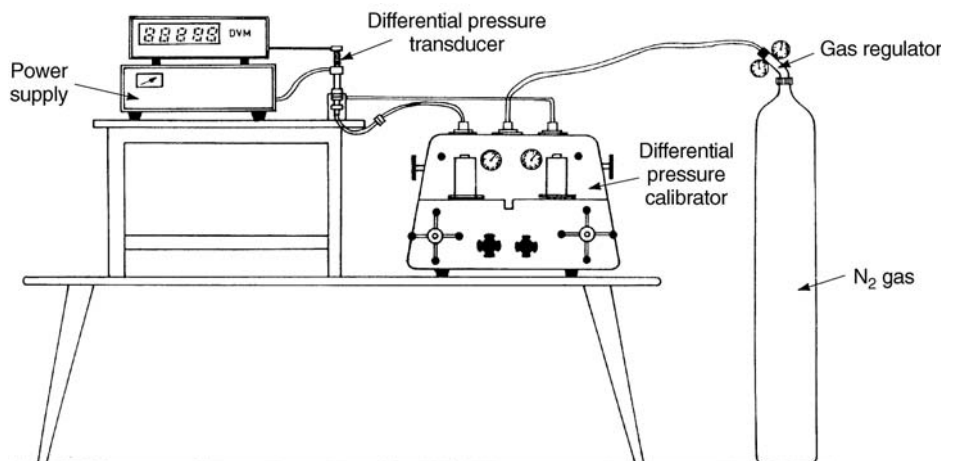


Figure 18. Schematic of the differential pressure calibration set-up. Reproduced from Sputtered thin film strain gauges for differential pressure measurement, M. M. Nayak et al., *IEEE Trans. Instrum. and Meas.*, **45** (1) February 1996, P 335–339. Copyright © IEEE, Inc. Reprinted with permission.

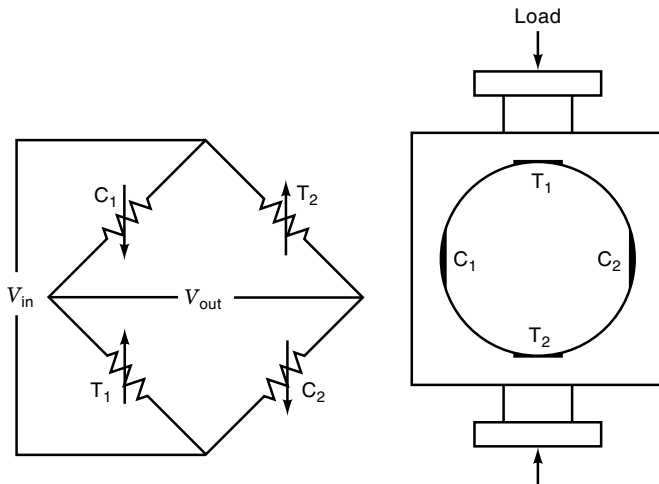


Figure 20. Ring type sensing element configuration for load cell.

coupled dual-beam, and cantilever beam type. A typical binocular type and ring type configuration are shown in Figs. 19 and 20 respectively. Depending on the range of load, appropriate materials and configurations are chosen for the sensing element. Also overload protection will be normally provided in load cells.

As pointed out earlier, strain gauges find application in several other measuring systems. Information on these as well as related aspects including analysis of strain gauge data can be found in Refs. 3, 6, 37, 42, and 43.

SUMMARY

Strain gauges and strain gauge based sensors/transducers find a wide variety of applications in many branches of science and engineering. In this article, most of the important aspects of strain sensors are presented. However, for some of the related aspects such as surface preparation of specimens, bonding of strain gauges, soldering and wiring, providing moisture and environmental protection, output standardization, shielding and grounding, curing and post-curing, and so on each manufacturer recommends its own standard procedures. This information is available as "Technical Notes" from the manufacturers. Some of the known strain gauge/strain gauge based transducers manufacturers are: (1) Micro Measurements, U.S.A., (2) Hottingen Baldwin Messtechnik (HBM), Germany, (3) Kulite, U.S.A., (4) Motorola, U.S.A., (5) BLH Company, U.S.A., (6) Rosemount, U.S.A., (7) Transamerica Instruments, U.S.A., (8) Kyowa Electronic Instruments, Japan, (9) Tokyo Sokki Lenkyujo Company, Japan, (10) Dynisco, U.S.A. and (11) Sensotec Inc., U.S.A.

In addition to the general applications indicated in this article, in recent years strain gauge transducers have become increasingly important in areas such as biomedical instrumentation, robotics, and space and defence applications. It is possible that the combination of advanced thin film materials technology, improved design approach for sensing elements of the transducer, and compatible signal conditioning electronics will result in the development of high performance strain gauge transducers for many more important applications.

APPENDIX 1. A CHRONOLOGY OF THE MAJOR MILESTONES IN THE HISTORY OF STRAIN GAUGE DEVELOPMENT

1856	Lord Kelvin, England
1908	Investigated the variation in electrical resistance of wires with increasing strain.
1923	S. Lindeck, Germany
1938–1945	Development of a method of measuring high pressure (800 PSI) using fine manganin wire wrapped around a copper tube.
1950	P. W. Bridgman, Massachusetts, U.S.A.
Free-rod approximation	Confirmed Lord Kelvin's results in a series of tests involving wires under hydrostatic pressure.
Thick-walled bulk-modulus type	A. C. Ruge, Massachusetts, U.S.A.
	E. E. Simmons, California, U.S.A.
	Considered as the co-inventors of the bonded wire strain gauge.
	First patent issued to E. E. Simmons on April 13, 1943.
	Subsequent patents issued to A. C. Ruge on strain gauges and transducers.
	Technograph LTD, England
	Foil strain gauge introduced using printed circuit technology.
Around 1958–1960	Introduction of semiconductor strain gauges.
Around 1970s	Introduction of thin film strain gauges.

BIBLIOGRAPHY

1. K. Rajanna, S. Mohan, and E. S. R. Gopal, Thin film strain gauges—an overview, *Indian J. Pure Appl. Phys.*, **27**: July–August 1989, 453–460.
2. E. J. Hearn, *Mechanics of Materials: An Introduction to the Mechanics of Elastic and Plastic Deformation of Solids and Structural Components*, Vol. 1, 2 Oxford: Pergamon, 1980.
3. C. C. Perry and H. R. Lissner, *The Strain Gauge Primer*, 2nd ed., New York: McGraw-Hill, 1963.
4. A. L. Window and G. S. Holister (eds.), *Strain Gauge Technology*, London: Applied Science Publishers, 1982.
5. P. R. Perino, Thin film strain-gauge transducers, *Instrum. Control Syst.*, **38**: December 1965, 119–121.
6. E. O. Doebelin, *Measurement Systems—Applications and Design*, 4th ed., New York: McGraw-Hill, 1990.
7. K. Rajanna and S. Mohan, Strain—Sensitive Property of Vacuum Evaporated Manganese Films, *Thin Solid Films*, **172**: 45–50, 1989.
8. K. L. Chopra and I. Kaur, *Thin Film Device Application*, New York: Plenum Press, 1983.

9. William M. Murray, Strain gauge types and basic circuits, *ISA J.*, February 1962, **9** (2): 47–51.
10. Jerome Catz, Basic strain gauge instrumentation, *ISA J.*, **9** (4): 50–55, April 1962.
11. Peter Mansfield, Electrical resistance strain gauge—theory and practice, *Transducer Technol.*, **8** (1): 17–19, January 1985.
12. Peter Mansfield, Electrical resistance strain gauge—theory and practice: 2, *Transducer Technol.*, **8** (2): 6–9, March 1985.
13. Peter Mansfield, Electrical resistance strain gauge—theory and practice, *Transducer Technol.*, **8** (3): 9–10, June 1985.
14. Peter Mansfield, Electrical resistance strain gauge: 4, *Transducer Technol.*, **9** (1): 7–11, January 1986.
15. Al Brendel, Temperature effects on transducers (part-1), *Meas. Control*, **13**: 54, February 1980.
16. Al Brendel, Temperature effects on transducers (part-2), *Meas. Control*, **13**: 90, March 1980.
17. Al Brendel, Temperature effects on transducers (part-3), *Meas. Control*, **13**: 138, April 1980.
18. Al Brendel, Temperature effects on transducers (part-4), *Meas. Control*, **13**: 172–173, May 1980.
19. Charles T. Wu, Transverse sensitivity of bonded strain gauges, *Experimental Mech.*, 338–344, November 1962.
20. J. C. Anderson, Thin film transducers and sensors, *J. Vac. Sci. Technol.*, **A4** (3): 610–616, May/June 1986.
21. K. Rajanna et al., Pressure transducer with Au-Ni thin film strain gauges, *IEEE Trans. Electron Devices*, **40** (3): 521–524, 1993.
22. T. Bravo, A. Tersalvi, and A. Tosi, Comparison of SiO_x and polyamide as a dielectric layer on stainless steel in thin film pressure sensor manufacture, *Sensors and Actuators A*, **32**: 611–615, 1992.
23. K. L. Chopra, *Thin Film Phenomena*, New York: Krieger Publishing Co., 1979.
24. L. Maissel and R. Glang (eds.), *Handbook of Thin Film Technology*, New York: McGraw-Hill, 1970.
25. L. Holland, *Vacuum Deposition of Thin Films*, London: Chapman & Hall, 1961.
26. D. R. Biswas, Review—deposition processes for films and coatings, *J. Mater. Sci.*, **21**: 2217–2223, 1986.
27. R. F. Bunshah (ed.), *Hand Book of Deposition Technologies for Films and Coatings—Science, Technology and Applications*, Park Ridge, NJ: Noyes publications, 1994.
28. H. K. Pulker, *Coatings on Glass*, New York: Elsevier Science, 1984.
29. B. Chapman, *Glow Discharge Processes: Sputtering and Plasma Etching*, New York: Wiley, 1980.
30. J. L. Vossen and W. Kern (ed.), *Thin Film Processes*, New York: Academic Press, 1978.
31. J. Harper, J. J. Cuomo, and H. R. Kaufman, Material processing with broad-beam ion sources, *Ann. Rev. Mater. Sci.*, **13**: 413–439, 1983.
32. R. W. Berry, P. M. Hall, and M. T. Harris, *Thin Film Technology*, New York: Van Nostrand, 1968.
33. M. M. Nayak, *Studies on sputtered thin film strain gauges and pressure transducers*, Ph.D. Thesis, Indian Institute of Science, Bangalore, India, 1994.
34. M. R. Neuman and W. G. Sutton, Structural dependence of strain gauge effect and surface resistivity of thin films, *J. Vac. Sci. Technol.*, **6**: 710–713, 1969.
35. K. Rajanna and S. Mohan, Studies on meandering path thin film gauges, *Sensors and Actuators* (Switzerland), **15** (3): 297–303, 1988.
36. R. S. Muller and J. Conragan, A Metal Insulator—Piezoelectric semiconductor electromechanical transducer, *IEEE Trans. Electron Devices*, **12**: 590, 1965.
37. R. S. Sirohi and H. C. Radhakrishna, *Mechanical Measurements*, 2nd Edition, New Delhi: Wiley Eastern Co., 1980.
38. M. M. Nayak, K. Rajanna, and S. Mohan, Performance study of pressure transducer with meandering—path thin film strain gauges, *Thin Solid Films*, **193/194**: 1023–1029, 1990.
39. M. M. Nayak et al., Sputtered thin film strain gauges for differential pressure measurement, *IEEE Trans. Instrum. Meas.*, **45** (1): 335–339, 1996.
40. Ural Erdem, Load cell technique for weighing accuracy, *Transducer Technol.*, **8** (1): 7, January 1985.
41. H. A. Nielsen, Jr., The ten dollar load cell, *Experimental Techniques*, 21–24, February 1988.
42. J. W. Dally and W. F. Riley, *Experimental Stress Analysis*, 2nd ed., New York: McGraw-Hill, 1978.
43. Thomas G. Beckwith, N. Lewis Buck, and Roy D. Maragoni, *Mechanical Measurements*, 3rd ed., New Delhi: Narosa Publishing House, 1982.

Reading List

- David W. A. Rees, The sensitivity of strain gauges when used in the plastic range, *Int. J. Plasticity*, **2** (3): 295, 1986.
- James Dorsey, Homegrown strain-gauge transducers, *Experimental Mech.*, **17** (7): 255, July 1977.
- R. Bertodo, Precious metal alloys for high temperature resistance strain gauges, *Brit. J. Appl. Phys. (J. Phys. D)*, SER. 2, **1**: 1743, 1968.
- W. H. Tuppeny, Jr. and A. S. Kobayashi (eds.), *Manual on experimental stress analysis*, 3rd ed., Society for Experimental Stress Analysis, 1978.
- R. J. Roark and W. C. Young, *Formulas for Stress and Strain*, 5th ed., New York: McGraw-Hill, 1975.
- J. C. Sanchez and W. V. Wright, Semiconductor strain gauges—what can they do?, *ISA Journal*, **9** (5): 38, May 1962.
- Paul Gay, Sputtered thin film method for high accuracy gauging, *Transducer Technol.*, **8** (1): 9, Jan. 1985.
- Ural Erdem, Load cell technique for weighing accuracy, *Transducer Technol.*, **8** (1): 7, Jan. 1985.
- B. S. S. Rao and M. Gopal Rao, A strain indicator for semiconductor strain gauges, *J. Phys. E: Sci. Instrum.*, **10**: 808, 1977.
- R. V. Milligan, The effect of high pressure on foil strain gauges, *Exp. Mech.*, **4** (2): 25, 1964.
- H. K. P. Neubert, *Strain Gauges: Kinds and Uses*, London: Macmillan, 1967.
- M. L. Meyer, A simple estimate of the effect of cross sensitivity on evaluated strain gauge measurements, *Exp. Mech.*, 476, Nov. 1967.
- C. S. Smith, Piezoresistive effect in germanium and silicon, *Phys. Rev.*, **94**: 42, 1954.
- G. R. Wilt, The electromechanical properties of thin films and the thin-film strain gauges, *Thin Solid Films*, **22**: 133, 1974.
- R. L. Parker and A. Krinsky, Electrical resistance-strain characteristics of thin evaporated metal films, *J. Appl. Phys.* **34**: 2700, 1963.
- Y. Onuma and K. Kamimura, Piezoresistive elements polycrystalline semiconductor thin films, *Sensors Actuators* (Switzerland), **13**: 71, 1988.
- W. Germer and W. Todt, Low-cost pressure/force transducer with silicon thin film strain gauges, *Sensors Actuators*, **4**: 83, 1983.
- K. Bethe and D. Schon, Thin film stain gauge transducer, *Philips Tech. Rev.*, **39** (314): 94, 1980.
- H. Yamadera and Y. Taga, Cr-O-X film as a strain gauge, *Thin Solid Films*, **206**: 107, 1991.
- D. Mariolli, P. Rolla, and A. Taroni, Strain gauge transducers: a evaluation of accuracy limits, *Measurement*, **10** (3): 98, Jul.–Sep. 1962.

- J. F. Lei, H. Okimura, and J. O. Brittain, Evaluation of some thin film transition metal compounds for high temperature resistance strain gauge application, *Mater. Sci. Eng.*, **A111**: 145, 1989.
- L. Clegg, Bonded foil strain gauge force transducers part 1, Materials and Design, *Sensors: J. Appl. Sensing Technol.*, **13** (9): 60, 1996.
- L. Clegg, Bonded foil strain gauge force transducer part 2, Performance, *Sensors: J. Appl. Sensing Technol.*, **13** (10): 68, Oct. 1, 1996.

K. RAJANNA
Indian Institute of Science
M. M. NAYAK
Indian Space Research Organisation
(ISRO)

TEMPERATURE SENSORS

TEMPERATURE AS A PHYSICAL QUANTITY

Every object or phenomenon existing in the real world may be described by a set of properties. Some of these properties are physical quantities, while others are descriptive ones. Physical quantities must be measurable. In order to make a property measurable, one has to establish a method with which to detect whether a state of a property is higher than another state, to detect if two states are the same, to propose a measure of the state, and finally to find a scale which transforms a given level of the property into an abstract symbol in the form of a number and a unit of measure. Considering temperature as a property of an object, all the problems mentioned above are rather complicated from both theoretical and practical points of view. The human sense of touch enables us to distinguish between higher and lower thermal levels, over a limited range of temperatures and with very limited repeatability, but nothing more, and there was a long way to go from our "feeling of heat" to the definition of temperature and temperature scales. The popular and often-quoted definition of temperature as an "intensity of heat" does not lead directly to solving the problem because of its lack of clarity.

Only the discovery of the fundamental laws of thermodynamics in the middle of nineteenth century allowed us to answer the question, What is temperature? The first law of thermodynamics says that thermal energy transfer is possible only from a system with higher temperature to a system with lower temperature. By observing the direction of thermal energy transfer, we are able both to tell which system is the one of a higher state of temperature and also to confirm the existence of equilibrium of temperature states when the heat transfer between two systems declines to zero. Furthermore, the works by Carnot, Lord Kelvin, and Clausius resulted in the formulation of the laws concerning the reversible thermodynamic cycle, called the Carnot cycle. The Carnot cycle consists of two isothermal heat conversion processes and two adiabatic heat transfer processes as illustrated in Fig. 1. By transferring a heat energy from the system with a higher state of temperature to a system with a lower temperature state, it is possible to transform a part of that energy (although a relatively small one) into mechanical energy. This constitutes a theoretical principle for all heat engines. The theoretical efficiency of a Carnot cycle

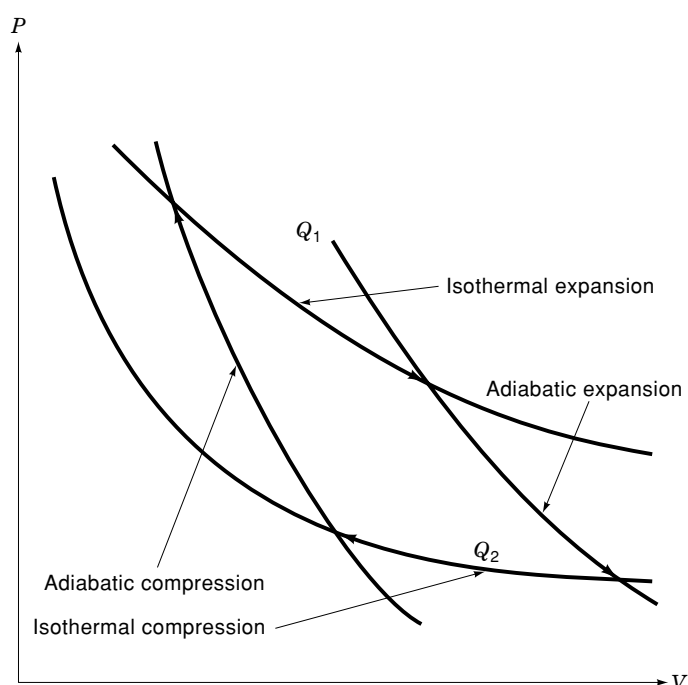


Figure 1. The Carnot cycle for an ideal heat engine. The arrows indicate the subsequent thermal processes.

engine is

$$\eta = \frac{Q_1 - Q_2}{Q_1} \quad (1)$$

where Q_1 is thermal energy transferred from the system with higher temperature in the isothermal expansion process, and Q_2 is thermal energy transferred to the system with lower temperature in the isothermal compression process. The theory of the Carnot cycle does not depend on the medium used, and Eq. (1) is a universal one. The engine efficiency depends only on the ratio Q_2/Q_1 and this ratio was proposed by Lord Kelvin as the basis of a new "absolute" thermodynamic measure of temperature in the form

$$\frac{T_2}{T_1} = \frac{Q_2}{Q_1} \quad (2)$$

Such a measure is independent of the thermal properties of any material and requires only one fixed temperature point to create the temperature scale. Equation (2) indicates very clearly that temperature has no physical zero point. In creating any temperature scale the zero point is to be assumed arbitrarily. That is, temperature, in the same manner as time, can be measured only by the interval scale but not by the metric scale. The point defined as 0 K is only a mathematical point on the scale but not a physical null temperature. Even in outer space we are not able to achieve physical null temperature, because the insertion of any material body changes the state of the temperature previously existing.

In Lord Kelvin's lifetime the Celsius scale of temperature had been in use for 100 years, and the temperature differences (e.g., between the ice point and the boiling point of water) had already been expressed in Celsius degrees. Gay-Lussac's gas law had also been known for more than 100 years

and expressed as

$$V = V_0[1 + \alpha(t - t_0)]_{p=\text{const}} \quad (3)$$

where V and V_0 are the volumes of an ideal gas at temperatures t and t_0 , respectively. Equation (3) proved to give results fitting quite well to the experimental results, especially for rarefied gases at low pressure, and the numerical value of the coefficient α was known with a good accuracy. Taking a Celsius degree as a unit of temperature difference, assuming the ice point of water as a point of $t_0 = 0$, and accepting particularly for these conditions an experimentally obtained value of the coefficient $\alpha = 0.003661 \text{ } 1/^{\circ}\text{C}$, we are able to create a new temperature scale, with the zero value at the point when an ideal gas volume decreases to zero (Fig. 2). The scale is now known as the absolute or Kelvin scale and is shifted by a value of $1/\alpha = 273.15$ with respect to the Celsius scale. All currently used temperature scales are in linear relations with each other:

$$\begin{aligned} T[\text{K}] &= t[^{\circ}\text{C}] + 273.15 = (t[^{\circ}\text{F}] + 459.67) \frac{5}{9} \\ t[^{\circ}\text{C}] &= T[\text{K}] - 273.15 = (t[^{\circ}\text{F}] - 32) \frac{5}{9} \\ t[^{\circ}\text{F}] &= 1.8(t[^{\circ}\text{C}] + 17.778) = 1.8(T[\text{K}] - 255.37) \end{aligned}$$

It is evident that Gay-Lussac's law is an excellent basis for developing not only the scale but also a thermometer: a gas thermometer. There is only one problem, albeit a significant one: An ideal gas does not exist. This inconvenience may be overcome either by the use of gases with properties close to the ideal gas (^4He , H_2) or by the use of rarefied gases at low pressures and by applying a very special measurement procedure. It has to be pointed out that in Eq. (3) there is an implied condition that $p = \text{const}$. In practically built gas thermometers, however, it is easier to fulfill the requirement of

$v = \text{const}$ and to observe the changes in p , instead of v . Therefore the majority of gas thermometers work on the constant volume principle.

INTERNATIONAL TEMPERATURE SCALE OF 90

In spite of the great progress in measurement techniques achieved since the days of Lord Kelvin, the use of gas thermometers is not a way to establish a contemporary temperature scale because of the great difficulties regarding their performance. Absolute Kelvin temperature according to Eq. (2) remains the theoretical definition of temperature, but the temperature scale is reproduced with the highest accuracy by means of the International Temperature Scale (ITS) established first in 1927 and recently modified in 1990. Most high accuracy industrial temperature measurement requires the production of a reproducible temperature state rather than its value in terms of its absolute temperature value. ITS-90 allows this by setting up a scale that is highly reproducible but only approximates to an absolute Kelvin scale to the limits of technology available in the late 1980s. Temperatures corresponding to that scale are sometimes marked using the subscript 90 in order to distinguish the differences with respect to the former scales (1927, 1948, 1968) and to the theoretical values of the thermodynamic scale (T [K] and t [$^{\circ}\text{C}$] with no subscripts).

The ITS-90 describes a set of fixed temperature points (17 points) and the methods of measuring the temperature between these points. Fixed points are freezing or triple points of selected materials (except Gallium's melting point). The points and the methods have been chosen to ensure, according to the actual knowledge and technology, the best conformance to the absolute temperature scale. An additional document, "Supplementary Information for the ITS-90," gives a very in-depth and exhaustive description of the instruments and the procedures which ensure the highest accuracy and traceability of

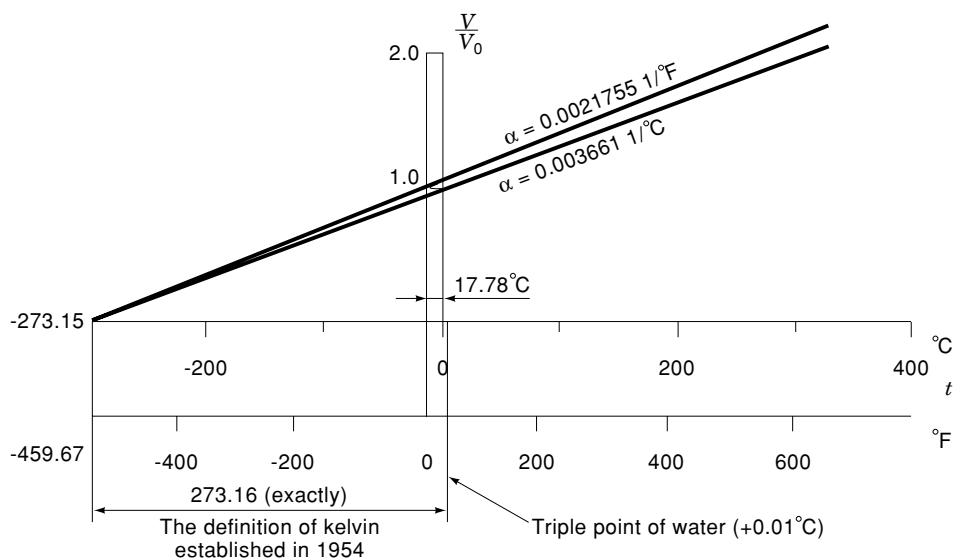


Figure 2. The meaning of absolute temperature scale and its relation to the ideal gas law. The difference between ice point and triple point of water on the temperature scale is excessively enlarged in order to enhance clarity.

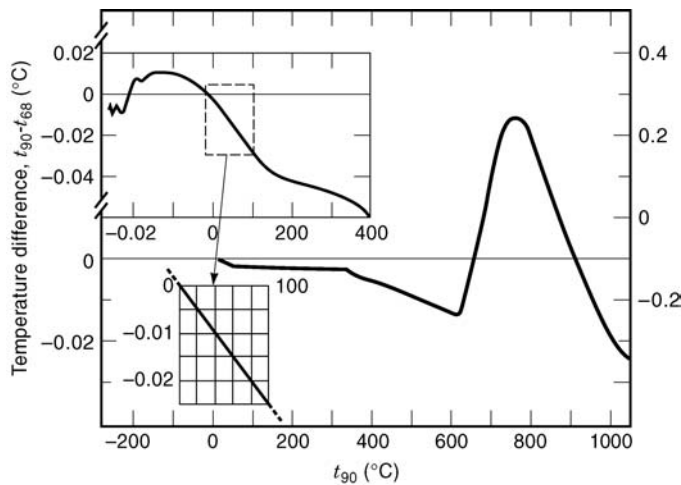


Figure 3. The differences in temperatures expressed by IPTS-68 and ITS-90. (After Ref. 1.)

temperature standard measurements. Figure 3 presents the difference in temperature values expressed by the previous IPTS-68 scale and by the present ITS-90. In some ranges, especially above 630°C, the differences are really great and are mainly caused by the incorrect reference function accepted by the IPTS-68 for the type S standard thermocouple. The correction of that reference function allows us to reduce the differences. Boiling points are rejected by the ITS-90 because of their poor stability and great sensitivity to pressure. The boiling point of water is no longer a fixed point. Some differences in standard instruments and methods have been introduced, too. The most important is that the standard platinum resistance thermometer now covers a much wider range of temperatures than previously, extending from about -260°C up to the freezing point of silver $+962^{\circ}\text{C}$. Above that temperature the Planck radiation law is used as a principle for standard measurements. Thus, the PtRh–Pt thermocouple is no longer a standard thermometer (1).

Extremely complex reference functions have been defined in order to express sufficiently precisely the ratio of the resistance at a given temperature T_{90} to the resistance at the triple point of water ($T_{90} = 273.16\text{ K}$) and vice versa, to express the temperature as a function of that ratio. The convenience of computerized calculations justify the complex forms of these functions.

They are not used, however, for numerical corrections of temperature values in the industrial, microprocessor-based instruments (e.g., for linearity correction). For that purpose much simpler, yet not so accurate, equations have been developed. Equation (4) is an adequate example.

ITS-90 serves as the best approximation of a realized temperature scale to the absolute thermodynamic scale and determines a highest level of the temperature standards. In every country, several levels of temperature standards are available, which are used for comparisons and to calibrate the thermometers in practical use. The uncertainty is higher as far as the calibration of the technical thermometers is concerned, and it ranges from a few millikelvin or less at the highest accuracy level and in the medium temperature range to the tenths of Kelvin for the case of industrial thermometers and thermometers used in everyday life. The uncertainty of both standard and practical thermometers in the higher temperature range (above 650°C) is always greater. According to the general idea of the ITS-90, it is evident that some modifications are inevitable in the future.

THE GENERAL PRINCIPLES OF TEMPERATURE MEASUREMENTS

The measurement of temperature differs from the measurement of other fundamental quantities such as mass, force, longitude, or voltage not only because of the lack of physical zero point of temperature, but primarily because of the inconvenience in direct comparison of the thermal state of the system of unknown temperature with the thermal state of the standard. The temperature is an intrinsic property of a material and hence does not permit scaling in the way of an extrinsic property such as length or mass. To measure temperature it is necessary to find an extrinsic property that varies in a predictable way with temperature and use this to construct a thermometer. That is why the practical measurements of temperature are always performed indirectly. The temperature sensor interfaces with the system whose temperature is to be measured. The interface may be realized by insertion, by point contact, or by visual contact with the system (Fig. 4). The sensor converts the thermal state of a system to a determined state of another quantity, which is defined as an output signal from the sensor. The output signal is then processed in the transducer T and finally presented in numerical form as a result of the temperature measurement. However, it is not the only function that contemporary transducers perform. They are more and more

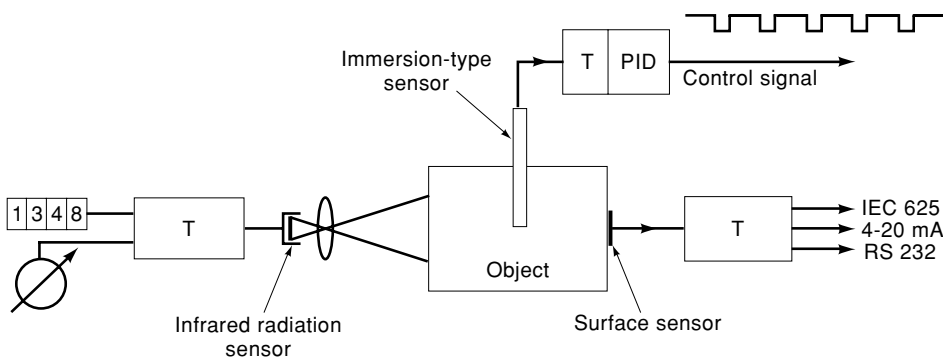


Figure 4. Temperature sensors and transducers. The outputs are arbitrarily assigned to the transducers.

frequently equipped with microprocessors and constitute a system which controls the measurement process, controls range changes, performs the numerical result correction, presents the results in appropriate units, and controls the standard interfaces such as RS 232, IEC 625, or others. Many control systems work according to the two-wire 4–20 mA standard. Therefore it happens very often that a temperature-measuring transducer provides the 4–20 mA output too, or even has an incorporated PID or on/off controller. Sometimes sensors are integrated with transducers either mechanically only or both mechanically and electrically on the same chip. At a large enough scale of integration, the term *integrated sensor* (IS) is justified. (The reading of the acronym IS as “intelligent sensor” is a commercial and marketing abuse.)

It is not the aim of this article to describe transducers but sensors. Therefore only input stages of the transducers—those stages which are directly interfacing with the sensors—will be presented. They are usually analog parts of measuring circuits, because all temperature sensors are analog devices. The principle of sensor operation depends on the physical phenomena used for conversion of temperature to the output signal. There are a lot of physical phenomena where temperature plays a significant role, but only a few of them are used in temperature sensors because they have to meet some additional requirements. These are as follows:

1. The monotonic calibration curve—that is, the relation between the temperature and an output signal over a sufficiently wide temperature range.
2. Sensitivity to temperature that is much higher than the sensitivity to all other influencing variables.
3. The output signal easily measurable with sufficiently low uncertainty.
4. Interchangeability of the sensors at least within the same type or the same manufacturing technology.
5. Repeatability of the calibration curve over the whole range of operating conditions.

Repeatability is mainly disturbed by hysteresis, relaxation, and aging. Hysteresis is observed when the calibration curve taken for increasing temperatures differs from that taken for decreasing temperatures. Relaxation means a delay in approaching the stable value of measured temperature after a rapid change in the sensor temperature. Aging is a long time process which causes the shift of the calibration curve due to many, sometimes unknown, reasons. Recalibration reduces, at least for a limited period of time, the errors caused by aging.

The linearity of the calibration curve and a large output signal are no longer essential requirements because of the progress in signal conditioning technology.

In order to produce a good sensor, the above requirements have to be fulfilled, but they are not sufficient to ensure a proper temperature measurement. The quality of temperature measurement depends to a great degree on the design of the sensor adequate to the conditions where the temperature has to be measured and on a proper measurement procedure. These two aspects are general ones, and are valid for all measurement techniques, but for the temperature measurements their importance is particularly great. It is due to the fact that every temperature sensor measures its own temperature and, more precisely, the temperature of its own sensitive part (the

thermometric body). The designer and the user of a sensor have to ensure the sameness of that temperature with the temperature which is defined as a measurand (that one which has to be measured in the given particular circumstances). For that purpose the sensor should be brought into as close thermal equilibrium with the measurand as possible without disturbing the measurand's thermal state. It requires a good thermal connection to the thermometric body and a poor thermal connection to the environment. The difference in thermal conductivity between thermal isolating and conducting materials is not very high. This involves some difficulties in design of a good thermometer and leads to measurement errors. Errors and uncertainties in temperature measurements will be discussed in more detail in the last section of this article. The problem of the thermal burdening by a sensor does not exist in radiation thermometry, but other sources of errors occur in this case.

RESISTANCE SENSORS

In all resistance sensors the change of their resistance follows the temperature changes, but the way it happens is different in metal sensors and in semiconductor sensors. It is therefore reasonable to separate the considerations about those two groups of resistance sensors. Furthermore, within the group of metal resistance sensors there is a great difference in the design of precise and industrial thermometers. This difference justifies further subdivision of the discussion.

Precise Resistance Sensors

The resistivity of almost all metals depends on temperature, but only a few of them are used in resistance thermometers: those which meet the requirements listed in the previous section. Pure platinum is considered the best material for temperature sensors. It has a relatively high resistivity (15 times greater than copper); thus wires needed to form resistors do not need to be particularly thin. Platinum can be obtained in a pure form with few impurities ensuring repeatability and interchangeability of the sensors. However, the most important reason why platinum is so widely used for temperature sensors is its ability to withstand even severe environmental conditions at high temperatures. For this reason, only pure platinum is used in the standard temperature sensors.

The progress in the technology and in the design of platinum temperature sensors achieved in the past decades made it possible to eliminate the PtRh–Pt thermocouple from the list of standard thermometers, which define temperatures according to ITS-90. Now, the temperatures T_{90} in the range between 13 K and 1233 K (960°C: silver point) are defined by the standard platinum resistance thermometer (SPRT). SPRTs are situated at the top of the hierarchical system of propagation of standards and are used as a first tie, which links ITS with all other temperature standards. They are used only occasionally for measurement of unknown temperature but more frequently for calibration purposes only. It is evident that to satisfy such high demands the quality of SPRTs must be the highest one. The purity of platinum used is secured by meeting two requirements: $R_{\text{Hg}}/R_{\text{TP}} \leq 0.844235$ and $R_{\text{Ga}}/R_{\text{TP}} \geq 1.11807$, where R_{Hg} , R_{Ga} , and R_{TP} are resistances at Hg point, Ga point, and triple point of water, respectively. These requirements are much greater than those required for industrial thermometers. In order to achieve such high values, the purity of platinum

has to be greater than 99.999%. The influence of impurities is much stronger at lower temperatures, limiting the temperature range of SPRTs. The next limitation is a very low resistance at 13 K with respect to the resistance at the triple point of water, R_{TP} (approximately one-thousandth), which makes the calibration process more complicated and results in decreasing sensitivity.

Standard resistors are always wire-wound, but the cores differ according to the temperature range. For the lowest temperatures, resistors are encapsulated in a hermetically sealed platinum sheath filled with helium under the pressure of 30 kPa. Such a design makes the sensor short and vacuum-protected as required for calibration in cryostats used for realization of ITS-90 fixed points. For higher temperatures, SPRT sensors are fixed at the end of long Inconel or quartz glass tubes, because of the necessity of providing a deeper penetration in the calibration device. For temperatures above 650°C, some special materials such as silica, alumina, or sapphire must be used. The wire diameter of high-temperature SPRTs is greater, exceeding 0.5 mm, resulting in a lower resistance of 2.5 Ω or even 0.25 Ω as compared with 25 Ω for SPRTs used at lower temperatures and wound from 0.05 mm wire.

Industrial Metal Resistance Thermometers

The industrial resistance thermometers used for measurement and control purposes in manufacturing plants, in the automotive industry, in housekeeping equipment for environmental measurements, and for many other everyday purposes are much less accurate than standard resistive thermometers. Industrial thermometers differ from SPRTs not only by design, technology, and material used, but also by the idea of its implementation. For the purposes of the ITS and of the calibration performed with SPRTs, the ratio of two resistances at two different temperatures is taken as a measure of the temperature. In “normal,” not standard, temperature sensors the value of the resistance of the thermometer becomes a measure of temperature. In other words, the output of a standard thermometer is the resistance ratio, and the output of an industrial thermometer is its resistance. The abbreviation for industrial thermometers is PRT (without “S”) or more frequently RTD (resistance temperature detector), indicating that sensors in use are not only platinum. Most of the RTDs all over the world are adjusted to the nominal value equal to 100.00 Ω at 0°C,

and hence termed as Pt-100 or Ni-100. The relationship between resistance and temperature for platinum RTDs is much simpler than for SPRTs and may be expressed in the form of the following equation:

$$R(t) = R_0[1 + At + Bt^2 + Ct^3(t - 100)] \quad (4)$$

where $A = 3.90802 \times 10^{-3} \text{ } 1/^{\circ}\text{C}$, $B = -5.802 \times 10^{-7} \text{ } 1/^{\circ}\text{C}^2$, $C = -4.27 \times 10^{-12} \text{ } 1/^{\circ}\text{C}^3$ for $t < 0$, and $C = 0$ for $t > 0$. However, it is not the equation but the values of resistances corresponding to appropriate temperatures that are the subject to national and international (IEC) standards, in the form of reference tables. Sometimes a distinction is introduced between “European” sensors with $R_{100}/R_0 = 1.385$ and “American” sensors with $R_{100}/R_0 = 1.392$. Furthermore, the uncertainties allowable for those sensors are also set in national standards, which are normally very close to the international IEC standards (Fig. 5). Standardization secures the reproducibility and hence interchangeability of RTDs, which is one of the most significant advantages of these sensors over all other temperature sensors.

It is a common practice that the repeatability of each individual sensor—especially over a limited temperature span—is in fact much better than the standard uncertainty limits. Therefore the individual recalibration of RTD sensors is recommended, because it allows further improvement of the accuracy. Due to the cost and available technical equipment, such calibration is commonly performed at one or two fixed points only. One-point calibration enables us to take into account an additive component of the difference between the nominal and the actual value of the resistor (i.e., additive error). Two-point calibration enables us to account for the sensitivity difference (i.e., the multiplicative error too). Nonlinearity error remains unknown. If the range of measured temperatures is limited (as is usually the case in practical situations), the nonlinearity of the RTDs has little influence on the total uncertainty and may be neglected. For wider temperature ranges the reference tables or appropriate $R(t)$ equations like Eq. (4) are useful for identifying the nonlinearity component and for applying it for the correction of the measurement result, together with the correction of the additive and multiplicative errors determined during the two-point calibration.

Besides platinum, some other metals and alloys are also used for temperature sensors. Nickel and copper are utilized

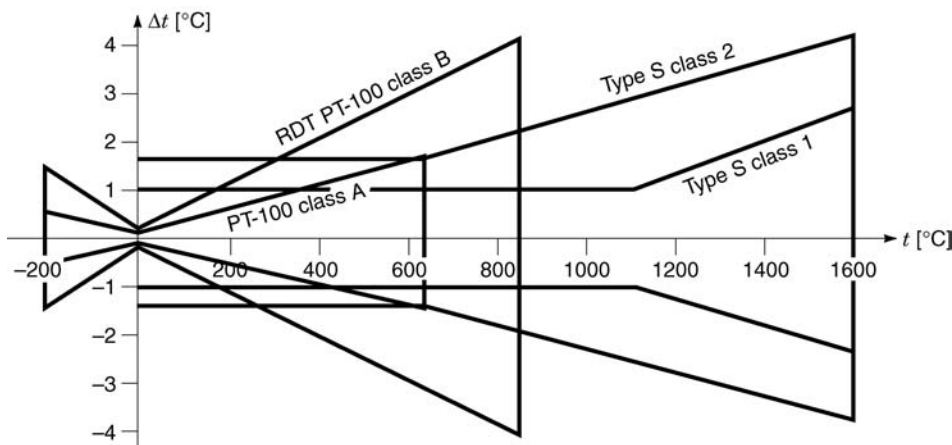


Figure 5. Comparison of allowable uncertainties of RTD Pt-100 after IEC Publication 751 and thermocouple type S after IEC Publication 584. Higher accuracy of resistance thermometers at lower temperatures is evident.

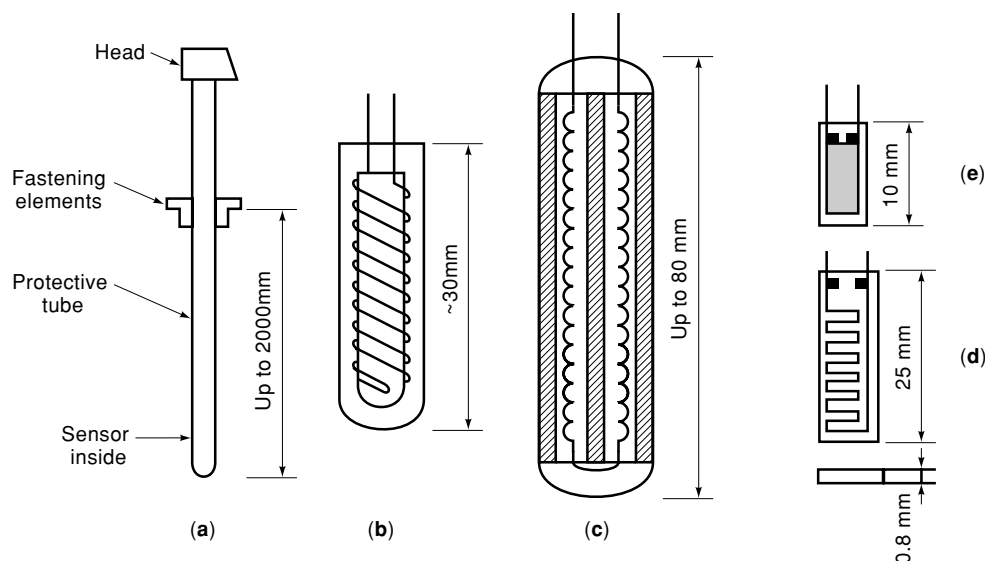


Figure 6. Immersion-type thermometer (a) and four types of sensors: ceramic sealed type (b), free wire type (c), thick film (d), and thin film (e). Types b and c are commonly used in industrial thermometers.

for temperature measurements over a narrow range. The sensitivity of nickel sensors is higher than that of the platinum sensors, but their nonlinearity is greater. Copper sensors are known to be extremely linear, but due to their low resistivity, it takes very long and thin wires to produce a 100 Ω resistor. Therefore, lower nominal values of copper sensors are also allowed by standards.

With the wire wound sensors two designs are usually used. In the first one the wire is bifilarly wound on a glass, quartz, or ceramic rod or pipe and coated with fired glass layer [Fig. 6(b)]. Glass other than quartz glasses is unsuitable for sensors above about 250°C as electrical insulation properties begin to deteriorate rapidly. For sensors working at temperatures above 600°C, glass is not a proper material and is replaced by alumina (Al_2O_3). The difference in the thermal expansion coefficient of core and platinum causes stress, which influences the long-term stability of the sensors. The second design is stress-free because the helical winding is placed in two holes drilled along the ceramic core and only sealed at the both ends of the core [Fig. 6(c)]. Often, two independent sensors are placed in four separate holes in the body. One of them may be replaced by the second in case of damage, or, more frequently, one serves for measurement and recording purposes while the second serves as a control. It is also very important to ensure the high shunting resistance from the internal mounting structures of the RTDs.

Sensors are protected from mechanical and chemical influences by metal tubes of different length (up to 2 m) made of stainless steel, nickel alloys, and sintered alumina and equipped with fixing elements for fastening them to the objects where the temperature should be measured [Fig. 6(a)]. The choice depends on the kind of object and the range of measured temperatures, pressures, and other environmental conditions. Because the sensor is placed at the bottom of the tube, it is convenient to lead out all the wires from one end of the sensor. In fact, this is the reason why the bifilar winding is used. It

protects against the induced noise voltage too. The best thermal contact between the sensor and the protecting tube is highly recommended.

Besides wire-wound RTDs, a group of sensors exists where a metallic layer is deposited on a flat or cylindrical core. The core material most frequently used is alumina, and a metal layer is deposited either as a thick film (5 μm to 10 μm) in a screen printing process [Fig. 6(d)] or as a thin film (1 μm) by sputtering [Fig. 6(e)]. The laser cutting provides the adjustment to the required resistance value. Sensors are coated with a thin protective layer of overglaze. Short response times of such sensors result from the small dimensions and small mass of the sensors. Long-term stability is a bit worse, and the temperature range is restricted to 500°C, but it probably will change with the advances in technology. Deposited RTDs are used in laboratory and service hand-held thermometers and in instruments requiring relatively accurate but small sensors for thermal control or correction. Psychrometric humidity sensing instruments also use deposited sensors. The electric signal which is obtained from these sensors is smaller than the one from traditional sensors because of the lower magnitudes of applied supply current.

SEMICONDUCTOR SENSORS

Semiconductor material may be used in resistance thermometers such as negative temperature coefficient thermistors, linear KTY sensors, and germanium resistance sensors used in cryogenic temperatures, but also in semiconductor devices such as diodes, transistors, and integrated circuits, the operation of which is related to the properties of p - n junction being the essential part of each semiconductor device. Classification and terminology are not established, but for the purposes of this article the distinction will be made between semiconductor resistance sensors and semiconductor active sensors.

Thermistors

Negative temperature coefficient (NTC) thermistors are prepared from a mixture of powdered metal oxides such as MgTiO_3 , MgO , CuO , NiO , and CoO , along with others sintered at the temperature of 1300°C . During that process, some p - and n -type semiconductor centers are created, thus enabling resistance–temperature relations to be described as semiconductorlike. In semiconductors, both electrons and holes are responsible for the conductivity.

$$s = \frac{1}{\rho} = \frac{ne^2\tau_e}{m_e} + \frac{pe^2\tau_p}{m_p} \quad (5)$$

where: s is the conductivity, reciprocal to resistivity ρ ; n and p are the numbers of electrons and holes in the valence band, respectively; τ_p and τ_e are their relaxation times; and m_e and m_p are their effective masses. In the semiconductors, τ_p and τ_e remain constant but n and p values change with temperature according to the relationship

$$n = p = 2 \left(\frac{kT}{2\pi h} \right)^{3/2} (m_e m_p)^{3/4} e^{-E_g/2kT} \quad (6)$$

where E_g is the energy of the band gap, and k and h are Boltzmann's and Planck's constants, respectively. From Eqs. (5) and (6) we obtain

$$\rho = CT^{-3/2} e^{E_g/2kT} \quad (7)$$

In the range of temperature in question (250 K to 400 K), the last term of Eq. (7) dominates. This fact leads to the well-known relationship between temperature and resistance of NTC thermistors in the form of

$$R = R_\infty e^{B/T} \quad (8)$$

R_∞ has a very small value and no physical interpretation. More practical therefore is the equation

$$R = R_{25} e^{B(T-25)/298} \quad (9)$$

where R_{25} is the thermistor resistance value at 25°C (500 Ω to 20 k Ω are typical values), and B is a material constant (Fig. 7). The value of B does not correspond strictly to $E_g/2k$ because many other factors influence the resistivity of the semiconductor, and additionally other mechanisms of conduction exist in the thermistor structure. Therefore the value of that constant depends on the material and manufacturing technology of the thermistor, and normally we obtain $B \approx 3000$ K to 5000 K. By describing the relative sensitivity of the thermistor in the same way as for the metal sensors, one obtains

$$S = \frac{1}{R} \frac{dR}{dT} = -\frac{B}{T^2} \quad (10)$$

from which the value of $S \approx -0.03 \text{ K}^{-1}$ at 25°C is approximately 10 times greater than that for metal sensors.

There are two principal types of thermistors commercially available. The first is the bead type, where the sintered material is formed into a bead of 0.1 mm to 1 mm diameter and sealed

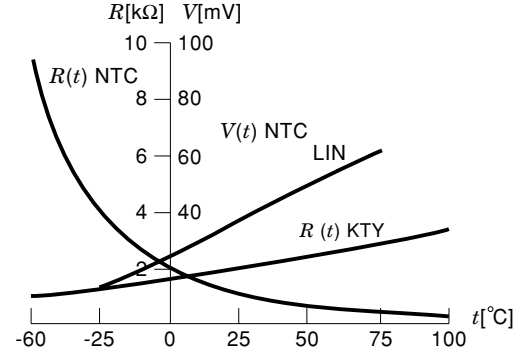


Figure 7. Calibration curves of an NTC thermistor, a KTY temperature sensor, and the output voltage of linearized circuit according to Eq. (15) for temperature span -25° to $+75^\circ\text{C}$. The sensitivity of linearized NTC sensor is twice the sensitivity of KTY which does not need any linearization procedure.

in glass stick or epoxy sheath together with two platinum connecting leads. The second and cheaper type is a disk thermistor, where the metal oxides are pressed at 1000°C into the forms of disks, tablets, and bars. Disk-type thermistors are less stable and are used for temperature compensation in electronic circuits. Only bead-type thermistors may be used as temperature sensors because their stability is much better, and after annealing at 60°C repeatability level of ± 10 mK may be achieved. Unfortunately, the interchangeability of thermistors is rather poor. Both parameters R_{25} and B differ for individual thermistors, even taken from the same batch (3). This is of particular importance because of the nonlinearity of the thermistors calibration curves. The methods of matching the $R(t)$ characteristics are much more complicated for nonlinear characteristics than for the linear ones. The International Standardization Organization (ISO) has attempted to unify thermal calibration curves by introducing the so-called ISO curve thermistors. The standardization concerned the R_{25} value (0.5, 1, 2, 4, 15, 100 k Ω), the shape of the calibration curve, and the admissible limits of interchangeability (from $\pm 0.25\%$ to $\pm 5\%$ of resistance span). Such thermistors are much more expensive than the ordinary ones.

Positive temperature coefficient (PTC) thermistors are used as switching elements rather than as temperature measuring sensors, because of their bistable calibration curve.

Bulk Silicon Sensors

Extremely low doped bulk silicon material shows a different mechanism of conductivity. At temperatures above 100 K, all free electrons become ionized and the temperature influences only the relaxation times, which decrease with the increase of temperature. As a consequence, the resistivity of doped silicon increases, creating a positive slope of the calibration curve of a respective sensor. At higher temperatures, however, the process of thermally excited electrons dislocating from the valence band to the conductivity band becomes more evident and stops the increase in resistivity. The mechanism described above may be practically used only when no p - n junction is created in the bulk material. This is why a special technique of resistance measurement has to be used. The technique is based on a great difference in size between the electrodes used for the measurement.

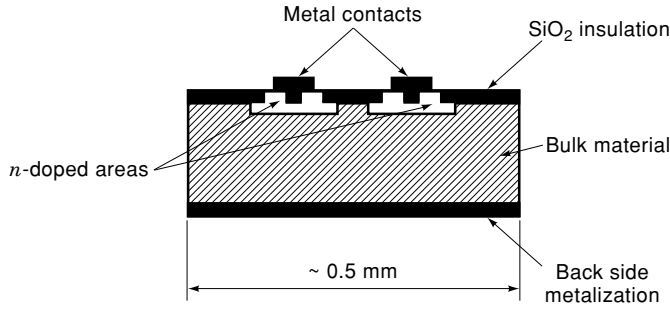


Figure 8. The design of a KTY sensor.

One electrode is only micrometers in diameter, while the other covers the whole counter surface of the semiconductor piece (Fig. 8). This creates a great difference in the electric field density at both electrodes, and therefore only the part of the material with high field density is responsible for the measured resistance R . The relationship between the resistance and the semiconductor resistivity is given in the form

$$\rho = \frac{R}{\pi d} \quad (11)$$

where d is the fine electrode diameter.

The commercially available temperature sensors, which work according to the described principle, are known as KTY linear sensors. In fact there are two small electrodes of $22 \mu\text{m}$ diameter, and the “back side” of the bulk silicon material is coated by a conductive layer. Such a design makes the sensor insensitive to the current direction. The very precise doping control of the material is realized by means of neutron implantation in which the silicon atoms are replaced by phosphorus atoms with excellent homogeneity over the whole bulk material. The resistance of KTY sensors at ambient temperatures is about $1 \text{ k}\Omega$ to $2 \text{ k}\Omega$, their sensitivity is about $1\%/K$, and the operation temperature ranges from -50°C to $+125^\circ\text{C}$ (Fig. 7).

Active Semiconductor Sensors

Active semiconductor sensors are those whose operating principle relies on the potential barrier between the conducting layer and the valence layer in a semiconductor, as in diodes and transistors. The simplest semiconductor device is a diode. According to the Shockley theory, the relationship between current I and

voltage V in the forward polarized diode may be expressed as

$$V = V_B + \frac{kT}{nq} \ln \frac{I + I_S(T)}{I_S(T)} \quad (12)$$

where V_B is a barrier voltage, n is a coefficient (not considered in the simplest form of the theory), and $I_S(T)$ is reverse saturation current, many times smaller than forward current I , but very strongly dependent on the temperature. Every temperature increase by 10 K results in its value doubling, and this behavior determines the temperature properties of the diode. In the temperature range of 150 K to 400 K the thermometric calibration curve of the diode is nearly linear, with the negative slope of approximately 2 mV/K to 2.3 mV/K . Unfortunately the I_S value depends not only on the temperature but also on many factors hardly controllable in the manufacturing process. Therefore, the diode sensor's interchangeability is poor. A single diode is the cheapest temperature sensor, but each one has to be individually calibrated at one or better at two points. Even in that case the uncertainty, including nonlinearity and hysteresis, is at the level of $\pm 2 \text{ K}$. In order to improve the properties of diode temperature sensors, two integrated diodes fed from two different current sources I_1 and I_2 should be used (Fig. 9). The difference in voltage drop over these diodes is

$$\Delta V = V_1 - V_2 = \frac{kT}{nq} \ln \frac{I_1}{I_2} \quad (13)$$

and does not depend on the reverse saturation currents because of their similarity due to the integration. Furthermore, the output voltage bias is significantly reduced, which leads to simpler measuring circuits.

Transistors are often used instead of diodes, and the base-emitter voltage difference is the output signal of the sensor. Integrated circuit (IC) technology allows us not only to produce temperature-sensitive pairs of transistors but also to include amplifiers and signal conditioning circuits on the same chip. In this way, integrated sensors with precisely trimmed output signal can be produced. The most popular are IC temperature sensors with $1 \mu\text{A/K}$ output (e.g., Analog Devices AD592), but sensors with voltage output of 1.5 mV/K or even 10 mV/K are manufactured too. The LM75 temperature IC sensor, produced by National Semiconductor, has a silicon band gap sensing element and is equipped with a sigma-delta A/D converter, programmable alarms, and a two-wire I^2C interface. The operating temperature span of diode- or transistor-based IC sensors ranges from -55°C to $+125^\circ\text{C}$.

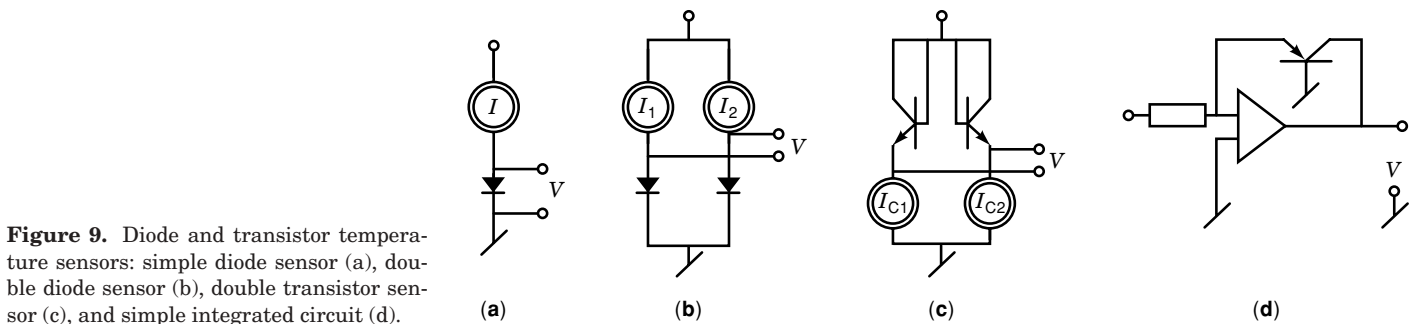


Figure 9. Diode and transistor temperature sensors: simple diode sensor (a), double diode sensor (b), double transistor sensor (c), and simple integrated circuit (d).

THE MEASUREMENT OF SENSOR RESISTANCE

Common Problems

The first stage of every transducer consists of the measuring circuit, directly connected to a temperature sensor. A few different measuring circuits are used to transform the resistance changes $R_x(t)$ to the output signal V . $R(t)$ forms that part of the sensor resistance which changes its value according to the measured temperature: $R_x(t) = R(t) - R_0$, where R_0 is a constant value of the sensor resistance corresponding to the reference temperature. Three problems seem to be common for all those circuits. These are (1) sensor self-heating, (2) lead resistance, and (3) linearity.

Self-heating of resistance sensors is unavoidable because the flow of the current creating the output signal causes automatic heat dissipation in the sensor, subsequent increase of its temperature, and consequent measurement error Δt :

$$\Delta t = P k_w = I^2 R(t) k_w \quad (14)$$

where k_w is a dissipation factor. The dissipation factor depends on the design of the sensor and on its materials, dimensions, and shape, but it depends primarily on the environment of the sensor. Its magnitude changes dramatically with the kind of medium surrounding the sensor and with the velocity of that medium, as presented in Table 1. Theoretically, the error due to self-heating can be removed from the measurement result by means of a correction procedure, but our knowledge about the value k_w is insufficient to calculate the correction value because of the unstability of the environmental conditions. It is sometimes possible to correct for self-heating effects by measurement at two currents and extrapolating to zero current. The best way, however, is to limit the error due to self-heating by keeping the current at the allowable level, but it results in lowering of the output signal.

The second problem is the change of lead resistances with temperature. The problem becomes serious when the distance between the sensor and the transducer reaches up to hundreds of meters and the long leads are exposed to great temperature differences (e.g., outdoor temperatures in summer and winter). For a 10 Ω copper lead the temperature change of 60 K (from -30°C to $+30^\circ\text{C}$) causes a 2.4 Ω resistance change which is observed as a 6 K temperature error if a Pt 100 sensor is used. The best way to avoid this kind of error is to feed the sensor from a current source by one pair of leads and to sense the voltage from the sensor by another pair of leads. This solution is called a four-wire line and is commonly used in transducers with standard analog 4–20 mA output and in all high

accuracy transducers. A three-wire line instead of a four-wire line is also used, especially in bridgelike measuring circuits. Three-wire installation cancels the additive errors caused by the thermal change of lead resistance, but the multiplicative part of the error remains. The higher the sensor resistance, the lower the influence of the lead resistance. There is no need to use four- or three-wire lines for thermistors or KTY sensors.

The third problem, linearity, is common for all transducers working with more or less nonlinear sensors. While most transducers are equipped with microprocessor-controlled systems, the linearity corrections are commonly performed numerically. The look-up table method is preferred. In that method, appropriate corrected values or the values of corrections which have to be added to the directly measured uncorrected results are written in memory. The linearization algorithm consists of a simple readout from the memory. At 0.1% resolution the method requires only 1 kB of memory. Some other methods of numerical linearity correction, utilizing the reduced tables containing only node point correction values, are also used. The correction data for all the results falling in between the node points are calculated by linear interpolation.

When the numerical linearization is performed together with other numerical corrections realized by the microprocessor system, the procedure becomes complicated and needs multidimensional tables. In that case the possibility of analog linearization ought to be considered. Two extremely simple methods may be used for that purpose. The methods are different in their principles, but both lead to the same results because of the same mathematical model.

The first method is based on the use of voltage controlled current sources in the structure presented in Table 2, part E. The supply currents depend linearly on the output voltage $I = I_0(1 + \alpha V)$. Linearization is performed by the proper selection of the coefficient α according to the sensor nonlinearity. The relation between V and R_x is then nonlinear and enables a successful linearization of NTC thermistors and nickel RTD sensors while choosing at positive α coefficient and of platinum RTD sensors while choosing it negative. The magnitude of α coefficient depends on the kind of the sensor and on the temperature range.

The second method is even simpler, because it does not require the voltage controlled sources. It depends on the proper choice of the Thevenin equivalent resistance R_{Th} seen from the sensor terminals, that is from the virtual terminals of the effective part of the sensor $R_x(t)$. According to such a definition of R_{Th} , the R_0 component of the sensor belongs to R_{Th} . The method is based on the theorem that any voltage between two points of each linear electrical circuit containing one variable parameter $R_x(t)$ may be presented in a form

$$U = \frac{U_0 R_{Th} + U_\infty R_x(t)}{R_{Th} + R_x(t)} \quad (15)$$

where U_0 is the voltage at $R_x = 0$, U_∞ is the voltage at the open circuit $R_x = \infty$. Some examples of linearizing circuits are presented in Table 2, A to D.

The above described methods do not lead to the canceling of the nonlinearity but only to its reduction (11). In practice more complicated circuits are used in integrated temperature measuring transducers like AD1B41 or B-B IXR 100, but the idea of linearization procedure depends on the above given principles.

Table 1. Dissipation Factors of Resistance Sensors Without Protective Sheath

Sensor	Environment	k_w (mW/K)
Wire-wound RTD ^a	Still air	3–5
	Air 1 m/s	10–20
Thin-film RTD	Still air	2
	Still water	75
NTC bead-type thermistor	Still air	1
	Stirred oil	8

^a RTD, resistance temperature detector.

^b NTC, negative temperature coefficient.

The Most Popular Measuring Circuits for Resistance Temperature Sensors

Generally the resistance sensors are manufactured with high accuracy. Transducers have to be matched to the sensors in order not to increase the total uncertainty. It is evident that for SPRTs the most accurate methods and instruments should be used for measuring the resistance and more exactly the resistance ratio. Costs and compactness are less important. The uncertainty of modern precise resistance ratio measurements is as low as a few parts per million, but only with very special apparatus used in advanced well-equipped laboratories. Such measurement circuits will be not presented here. However, with conventional temperature measuring transducers the accuracy of resistance measurements has to be high too. Let us note that according to IEC 751 standard, the Pt-100 uncertainty at 0°C is only ± 0.1 K, which means ± 0.04 Ω . In order to protect the sensor accuracy, the uncertainty of the transducer ought to be less than, say, ± 0.01 Ω , which gives 0.01% with respect to a 100 Ω sensor. For resistance-measuring instruments in common use, it is a rather high requirement and a bridge circuit is therefore the one which has to be primarily considered as the input stage of the transducer.

Balanced bridges are contemporarily used almost only in self-balancing chart recorders or x - y - t recorders. The complicated mechanical design of such instruments together with the need for the precise potentiometer, makes these instruments rather expensive. Such instruments have been formerly widely used not only in industrial applications, but also in laboratories and research. High cost and the absence of an electrical output signal (which may eventually be obtained from an additional potentiometer) make those instruments not very suitable for modern instrumentation systems.

To the contrary, unbalanced bridges are very often used as the first stages of contemporary transducers working with resistance temperature sensors [Fig. 10(a)]. The differential structure of any unbalanced bridge circuit enables easy adjustment to the desired temperature range. The output voltage is not strictly proportional to the resistance, because the unbalanced bridge belongs to the class of the circuits described by Eq. (15) and presented in Table 2 as cases B and C. Therefore, an unbalanced bridge may also be used as a linearizing circuit for some types of sensors. To do that, an appropriate matching of the bridge branches have to be performed. Unbalanced bridges are supplied either from a voltage source or from current sources. The constant current supply is preferred especially for low-resistance sensors, as Pt-100 or Ni-100, where the three-wire connection between the sensor and the transducer is needed in order to reduce the line temperature error. The reduction is twice better using a current source than using a voltage source.

The output voltage from a bridge is fed to a direct-current (dc) differential amplifier. The signal is usually high enough for a conventional low-noise operational amplifier with a proper compensation of bias currents. In some extremely precise instruments the switched-capacitor-based instrumentation amplifiers are used (i.e., Linear Technology LTC 1043). The aim of the amplifier is not only to increase the signal but also to allow the transition from differential to a single-ended signal. It is a general requirement, especially in the industrial measuring systems that the sensor must be grounded. When the three-wire configuration is used, two of them are close to ground

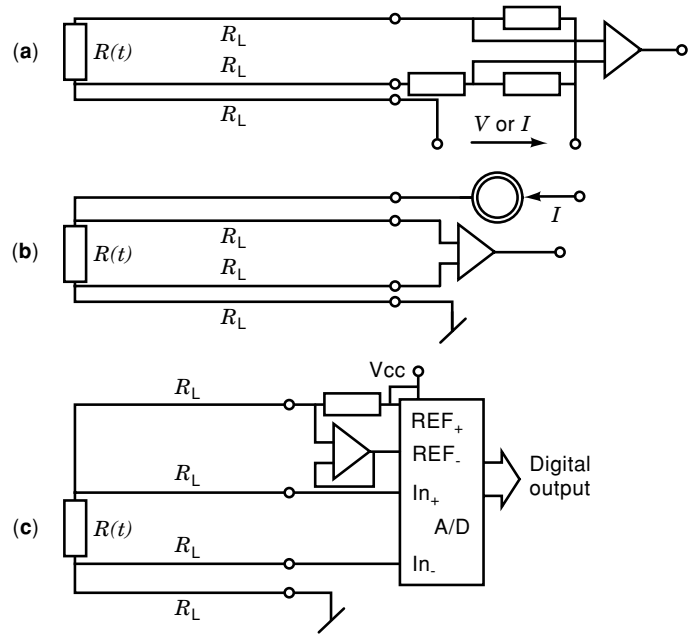


Figure 10. The measuring circuits which reduce the influence of the leads resistances R_L . (a) Three-wire bridge circuit makes it possible to connect one lead to the sensor and the second to the resistor in opposite bridge arm and then to reduce their influence. (b) Four-wire circuit with the current source enables canceling of the lead resistance influence. (c) The implementation of an A/D converter in the four-wire circuit provides direct conversion of the analog signal to the digital one.

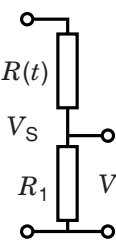
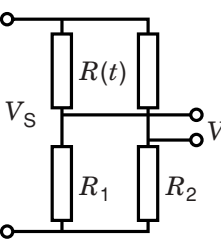
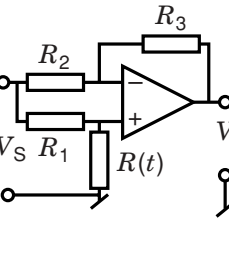
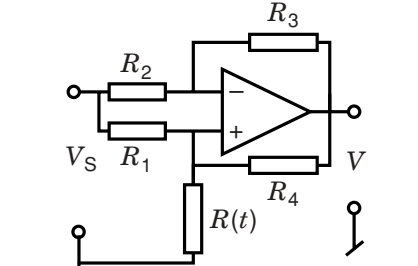
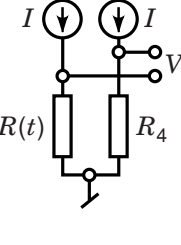
potential, but according to the principles of noise protection they must not to be grounded at two points (at sensor's side and at transducer's side) in order to avoid the ground loop which introduces additional unknown voltages. The circuit with floating voltage supply and grounded one-amplifier input is less convenient because of the limitations in scaling the circuit parameters. The greatest comfort in circuit parameters scaling is provided by a four-wire installation because it consists of two almost separated circuits [Fig. 10(b)]. The only problem to solve is the subtraction of that part of voltage which corresponds to the low limit of the measured temperature. It may be done either by a bias voltage or by another differential structure containing a sufficiently stable voltage source. Integrated circuits, which incorporate a controlled gain amplifier, a linearization circuit, and isolated output (i.e., Analog Devices 1B41), facilitate the design of the measuring system. A/D converters with reference input may be used for direct four-wire connection to the sensor supplied from the voltage source instead of a more complicated current source [Fig. 10(c)].

Some completely different temperature measuring circuits—that is, circuits with frequency output, where the sensor resistance influences either the oscillator frequency or the duty cycle of square-wave output voltage—are also known. The practical implementation of such circuits are limited mostly to those in a form of integrated circuits—for example, the SMT 160-30 produced by Smartec.

RESISTIVE SENSORS FOR LOW TEMPERATURE RANGES

The range of temperatures below 20 K becomes more and more interesting not only for the researchers but also for

Table 2. Some simplified examples of linearizing circuits for resistance sensors. Circuit A (potentiometric), B (bridge circuit) and C (active bridge circuit) are for NTC thermistors or RTD nickel sensors. Circuit D (positive feedback circuit) is for RTD platinum sensor. Circuit E (with controlled current) can be used for all sensors regarding the proper sign of α coefficient. The balance state of bridge circuits is assumed at $R_x(t) = 0$, ($R_1 = R_2 = R_3 = R_0$).

 (a)	 (b)	 (c)	 (d)	 (e)
$R(t) = R_0 + R_x(t) \quad R_{Th} = R_0 + R_1$			$I = I_0 (1 + \alpha V)$	
$V = \frac{V_S R_1}{R_1 + R_0 + R_x(t)}$	$V = \frac{V_S}{2} \frac{R_x(t)}{2R_0 + R_x(t)}$	$V = V_S \frac{R_x(t)}{2R_0 + R_x(t)}$	$V = V_S \frac{R_1 R_3 R_4 + (R_1 R_3 - R_2 R_4) R(t)}{R_1 R_2 R_4 + (R_1 R_3 - R_2 R_4) R(t)}$	
			$V = \frac{I_0 R_x}{1 + \alpha I_0 R_x}$	

the technologists. The practical use of the superconductivity requires the precise temperature measurements and control of temperatures as low as 4 K. In some cryogenic technologies the high magnetic fields and nuclear radiation are simultaneously present. Temperature sensors destined for low-temperature applications have to be resistant to those environmental conditions too. It is reasonable to distinguish a special group of sensors working at low temperatures in spite of their different principles of operation and design.

As stated before, a platinum resistance thermometer does not work properly at temperatures below 10 K. For that range a different alloy has been developed, namely rhodium with 0.5% iron (2,3). The technology of preparing thin, 0.05-mm-diameter rhodium-iron wires is complicated. It includes chemical iron deposition on powdered rhodium and then a series of metallurgical processes. The helically wound sensor is hermetically encapsulated similarly to SPRT sensors. The most useful operating range is 0.3 K to 30 K; but due to its relatively low slope of resistance versus temperature, it may be used up to the normal ambient temperatures too. The stability of an Rh-Fe sensor is relatively good, much better than that of low-temperature semiconductor sensors. Semiconductor sensors, however, are much simpler and smaller, and for that reason they are used too.

Some specially prepared and composed thermistors, usually made from iron oxide, are able to measure temperatures as low as 5 K. According to Eq. (8), thermistor sensitivity and nonlinearity increases dramatically at lower temperatures, creating problems with covering a wider range of temperatures. This is a common problem of all low-temperature sensors related to the "wrong" representation of temperatures on the linear scale, as discussed in the first section of this article. The greatest advantage of thermistors is their low sensitivity to the magnetic fields. Germanium was previously used in

electronic semiconductor devices to the same degree as silicon. The physical principles of its conductivity remain the same as those described for silicon. Germanium temperature sensors have been used for measurement purposes for a much longer time than bulk silicon sensors. Their operating area, however, is shifted toward the very low, cryogenic temperature range. The bulk germanium with a very small amount of added impurities forms a low-temperature sensor which may be used down to 1.6 K; but due to the very strong dependence of its properties on the amount of the impurities introduced, the individual calibration of each sensor is necessary. The calibration process at extremely low temperatures is always a complicated and expensive one. Other problems related to the use of the germanium sensor are (1) rapid and unexpected changes in resistance of the magnitude corresponding to a few millikelvins, (2) high sensitivity to the mechanical stress, and (3) a strong Peltier effect causing an additional heat distribution along the sensor. The instability exhibited by many low temperature Ge resistance thermometers is thought to be due to the difficulty of attaching leads in a way that defines the resistor without subjecting it to strain effects. A long-lasting experience with this kind of sensor (since 1960) has resulted in gathering a considerable amount of information, which enables us to reduce the effects of all inconveniences listed above. Specially doped germanium resistors are insensitive to magnetic fields (3).

Individually calibrated diode sensors may also be used in the very low temperature region, down to 10 K. Sensitivity is not the same as at medium temperatures, and it increases rapidly below a certain temperature (approximately 25 K for silicon diodes), but the sensor calibration curve remains repeatable with the uncertainty of ± 10 mK. Commercially available diode sensors are produced with a wider uncertainty span, exceeding ± 0.25 K but with quite good reproducibility of ± 50 mK (4).

THERMOCOUPLE SENSORS

Physical Principles

A temperature difference between two points of a conducting wire forces free electron diffusion from the point of higher temperature to the point of lower temperature. Such dislocation of electrons produces a voltage difference, which forces the electron flow in the opposite direction. In the state of dynamic equilibrium, both processes are in balance. A voltage difference caused by the temperature difference is known as thermal electromotive force (emf), and it provides a measure of temperature difference between any two points of the wire. Thermal conductivity of the metal wire causes temperature distribution along the wire, and hence the thermal emf may be considered as continuously distributed along the wire too. The problem of how to measure the thermal emf arises because each electrical contact of the connecting leads with the heated wire is also a thermal contact and generates subsequent thermal emf corresponding to the temperature difference at the ends of the connecting leads. If the materials of the heated wire and connecting leads are the same, two emfs appear in the loop with the opposite signs and are reduced to zero. However, the magnitude of the thermal emf depends on the material; and it is possible to find a pair of materials (A and B) with essentially different emfs, thereby generating relatively high voltages in the loop (Fig. 11).

By denoting the thermal emf sensitivities to the temperature of both wires as $S_A(t)$ and $S_B(t)$, one obtains

$$E_{AB} = \int_{t_1}^{t_2} [S_A(t) - S_B(t)] dt = \int_{t_1}^{t_2} S_{AB}(t) dt \quad (16)$$

where $S_{AB}(t)$ is the sensitivity of a thermocouple known as a Seebeck coefficient or “thermoelectric power.” In other words, a temperature difference produces thermal emf, and the inhomogeneity of the materials in the loop allows us to detect it. The greatest inhomogeneity appears at the contact points of two different materials. These points constitute “hot junction” and “cold junction,” or more properly, a measuring junction and reference junction. Any other material (C in Fig. 11) included in the thermocouple loop but not exposed to the temperature difference does not produce any additional emf in the loop. Otherwise, each unexpected inhomogeneity, caused not only by inclusion of any additional material D but also by chemical reactions or annealing processes in the material under temperature difference, is a source of additional thermal emf. Therefore the whole of the thermocouple loop ought to be considered as the sensor not just the tip and this makes the consideration of sources of uncertainty much different to most other temperature sensors.

The thermal emf effect discovered by Seebeck in 1821 is superposed by two other effects related to the current flow in the

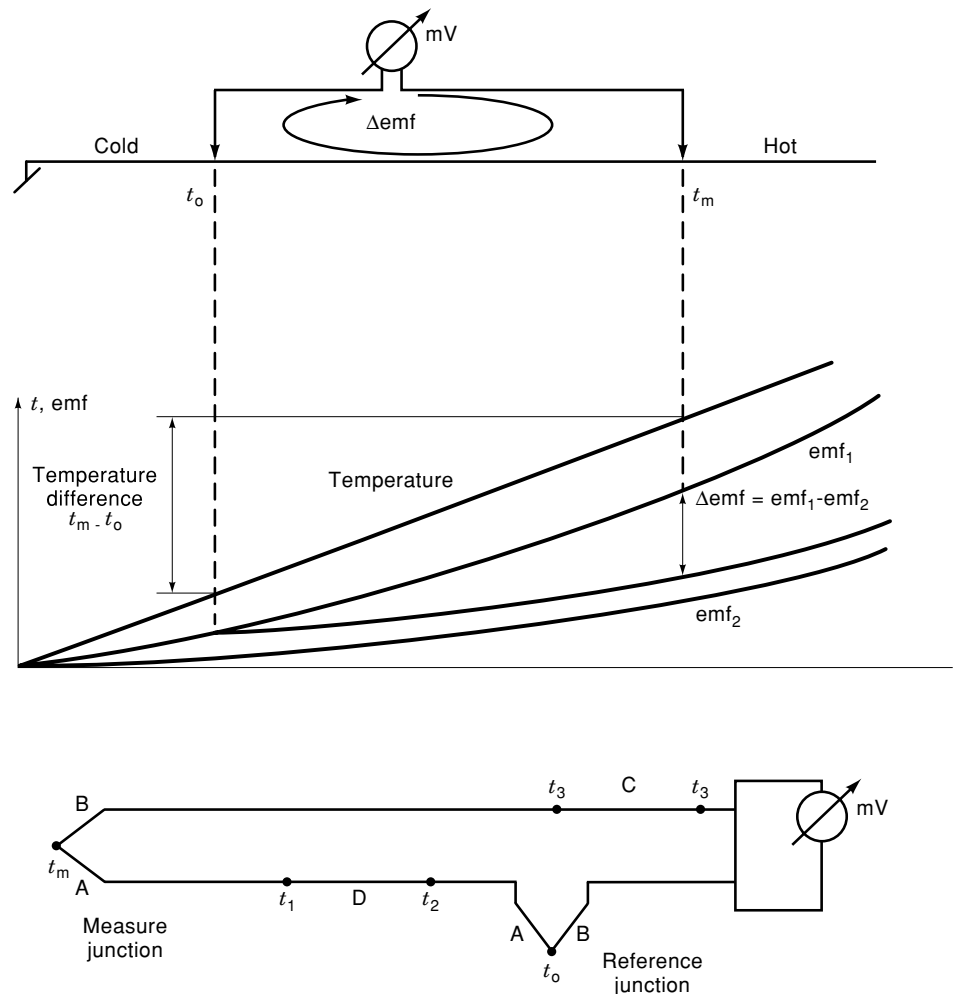


Figure 11. Temperature difference measurement by a thermocouple circuit. Temperature difference $t_m - t_0$ corresponds to the emf difference Δemf .

Table 3. Thermocouple Data

Type	+Wire	−Wire	Long-Time Temperature Span (°C)	Short Time Allowable Temperature (°C)	Thermal emf Span at 0°C Reference Junction (mV)	Allowable Uncertainty (IEC)				Properties
						Class 1 at 0°C (°C)	Class 1 at Upper Limit (°C)	Class 2 at 0°C (°C)	Class 2 at Upper Limit (°C)	
S	PtRh10	Pt	−50 to +1600	1760	−0.23 to +18.6	1	2.5	1.5	4	Stable. For oxidizing and inert atmosphere. Sensitive to contamination.
R	PtRh13	Pt	−50 to +1600	1760	−0.23 to +21	1	2.5	1.5	4	
B	PtRh30	PtRh6	+100 to +1600	1800	− to 13.8	—	—	—	4.27	Negligible thermal emf up to 60°C.
K	NiCr Chromel	NiAl Alumel	−270 to +1000	1250	−6.46 to +50.6	1.5	4	2.5	9	Most frequently used. Linear for $t > 0^\circ\text{C}$. For clean oxidizing and inert atmosphere.
N	NiCrSi	NiSi	−270 to +1000	1300	−4.3 to +47.5	1.5	4	2.5	9	More stable than type K at high temperatures.
T	Cu	CuNi Constantan	−270 to +350	400	−6.26 to +20.9	0.5	1.4	1	2.6	For oxidizing and reducing atmosphere at low temperatures. Moisture-resistant.
J	Fe	CuNi	−210 to +750	1200	−8.1 to +69.5	1.5	3	2.5	5.6	For reducing and inert atmospheres.
E	NiCr	CuNi	−270 to +700	1000	−9.8 to +76.4	1.5	3.2	2.5	6.75	High thermal emf, also at low temperatures.
—	WRe5	WRe26	0 to +2400	2700	0 to +40.7	No standards				For highest temperatures. For neutral and reducing atmospheres only.
—	NiCr	Au	−273 to 0	—	−5.3 to 0					Linear for low temperatures, also below 4.2 K.

loop: (1) Thomson effect and (2) Peltier effect. In the Thomson effect an additional emf is induced by current flow in the presence of a temperature difference in the conductor, involving heat liberation or absorption by the conductor at the rate

$$Q = \int_{t_1}^{t_2} S_{TA} I dt \quad (17)$$

where S_{TA} is a Thomson coefficient of the particular material A. For the same reason as with Seebeck effect, the Thomson emf in the whole loop is different from zero only in a case when two different materials make two branches of the loop. The Peltier effect reveals in the additional heat absorption or heat generation forced by the current flowing through a junction of different materials $Q_P = V_{PAB} I$, where V_P is called the Peltier coefficient (Peltier emf).

Three effects described above are interrelated according to the equation

$$E_{AB} = V_{PAB}(t_1) - V_{PAB}(t_2) + \int_{t_1}^{t_2} S_A(t) dt + \int_{t_2}^{t_1} S_B(t) dt \quad (18)$$

Because of the low currents flowing in the temperature-measuring thermocouple loops, the effects of additional heat emitted in the Thomson and Peltier effects may be normally neglected.

Thermocouples

It is evident that a thermocouple has to be composed of thermowires with very different thermal emfs. In measurement practice, however, a lot of additional requirements are of a great importance. Most of them stem from the fact that thermocouples are usually used at high temperatures. Therefore the wires themselves have to be resistant to high temperatures. Their

melting points must be at least by 150 K higher than the upper temperature of the measuring range. At high temperatures the atmosphere is usually corrosive and hence much more aggressive than at low temperatures. The degeneration of the lattice structure and the chemical changes observed at the surface of the wires cause the inhomogeneity of the wire materials and lead to the successively arising measurement errors. This is a reason why thermocouple wires are rather thick, and the upper temperature limit for continuous work case depends on their diameters. The cost of wires, especially those made of noble metals, is important too. Very special alloys have been developed in order to meet the above-mentioned requirements. The work in this area is still going on, and the result has been continuous improvements and modifications of the thermocouple wires. Table 3 presents the most popular thermocouples and their essential properties.

The calibration curves of some thermocouples are subject to standardization in the form of reference tables, similar to the corresponding tables for resistance sensors. Worse stability of thermocouples results in their much greater uncertainties as compared to resistance sensors (see Fig. 5). When a lower uncertainty is required, the individual calibration of thermocouples is not recommended because the validity of the results is rather short-lived.

Thermocouple Sensors

A great variety of thermocouple sensors with different sizes and designs are destined for a wide range of applications, such as power plants, nuclear plants, metallurgy, chemical reactors, and the glass industry, as well as laboratories, research works, and communal applications (5).

Industrial-immersion-type thermocouple sensors look like thick resistance sensors because they are manufactured in a similar form, with a long protective pipe and a head on one

end. The sensor element (wires with the junction) are mounted inside and may be replaced if necessary. However, the materials used for shields differ considerably. For high-temperature sensors a conventional carbon steel protective pipe is insufficient. Either stainless steel (18% Cr, 8% Ni), inconel (NiCr, 15% Fe), hastelloy, or bronze has to be used depending on the environmental conditions. Sometimes there is a need for molybdenum or tungsten sheath (for highest temperatures). Noble metal thermocouples have to be additionally protected by means of an internal ceramic (alumina Al_2O_3) coating, against the contamination of the thermocouple wires by the particles of pipe material that occurs at high temperatures. Some outer porous ceramic protection tubes are used with the sensors for open fire furnaces.

A special type of sensor is produced in the form of a metal shielded double thermowire cable with MgO or Al_2O_3 insulation. These are called shielded thermocouples or mineral insulated metal sheathed (MIMS) sensors. The same type of insulation is used in resistance heaters. The thermocouple junction is formed by connecting the both wires. The external diameter of the MIMS may be as low as 0.25 mm (more commonly 0.6 mm to 3 mm), and the bonding radius allowed is normally twice that of the diameter. This constitutes a great advantage of the sensor, being an ability to penetrate hardly accessible spots. This kind of sensor is now obtainable in lengths of up to a tenth of a meter, with the sensing junction, as well as the plug on the opposite end, formed and sealed by the manufacturer. Former MIMS were produced in a form simply cut from one piece of cable, but the hygroscopic properties of the insulation material made it very hard for the user to provide the proper hermetic sealing.

The next group of thermocouple sensors are those designated for the measurements of moving surface temperatures and designed as free thermowires or thermostrips, either suspended on elastic arms or shaped into a form of an elastic arch. The measuring junction is situated in the middle of the free part of the thermostrips and should be pressed to the surface during measurement. The smoothness of the thermojunction allows the measurement of the moving or rotating elements without heat generating by friction. The elasticity of the sensor ensures a good thermal contact with the surfaces of different shapes (e.g., with rollers of different diameters).

In metallurgy, two kinds of immersion-type sensors are commonly used for measurement of the molten metals temperature. Both kinds work under the transient state conditions. The construction must be strong enough to pierce the layer of the blast furnace slag. In the first design, two sharp-cut thick bars from thermocouple materials are placed near each other at the end of a long handle. The stick is immersed in the molten metal, thereby creating a junction. In the second design the

exchangeable cap with very thin thermocouple wires (50 μm to 100 μm) and a protective cover is placed at the end of the handle. The cover is damaged when immersed in the molten metal, and after each measurement the cap is replaced by the new one.

Many other thermocouple sensors are present on the market or are custom-designed for particular purposes. The use of a thermocouple sensor instead of another sensor type is recommended in a case when most important are (1) high temperature range, (2) small dimensions of the sensor, and (3) relatively low cost.

Thermocouple Measuring Circuits

At a first glance, the task seems to be easy: Create a reference junction and measure a dc voltage. However, some problems arise particularly in the industrial environment and at high temperatures. In the large area plants the distance between the sensors and the transducers is long, and at high temperatures the sensor head temperature is usually unstable to such a degree that it is impossible to treat it as a reference junction. Therefore, the thermocouple wires have to be brought to the place where the temperature is constant, say in the transducer (6,7). For evident reasons it is not a good solution, especially if noble metal wires are used. In such a case, extension wires are used as a connection between the sensor head and the reference junction. These are special compensation leads having the same thermal emf as the thermocouple wires, but with a much lower temperature range, namely that expected to occur at the sensor head (Fig. 12). Compensation leads have to be matched to the thermocouple; and in order to avoid misconnections, the colors of their insulation are subject to standardization. Compensating wires are much cheaper than thermocouple wires. A special noble metal thermocouple has been developed (Type B, Table 3), which does not require any compensation leads because its thermal emf at temperatures up to 50°C is practically equal to zero and with temperatures up to 120°C it is very low. For that thermocouple, neither a reference junction nor the compensating leads are needed, assuming that the ambient temperature of the transducer and the sensor head temperature do not exceed 50°C and 120°C, respectively.

For all other thermocouples, however, the reference junction is necessary. In laboratory practice ice-water baths, and in industrial measurements, thermostats may be used. Both are impractical. Instead of stabilizing the temperature of a reference junction, it is more convenient to measure it and to introduce a compensating voltage into an emf measurement loop. Such a method is now used in almost all instruments and transducers. The most common compensating circuit is shown in Fig. 12. At nominal temperature of the reference junction (say 25°C),

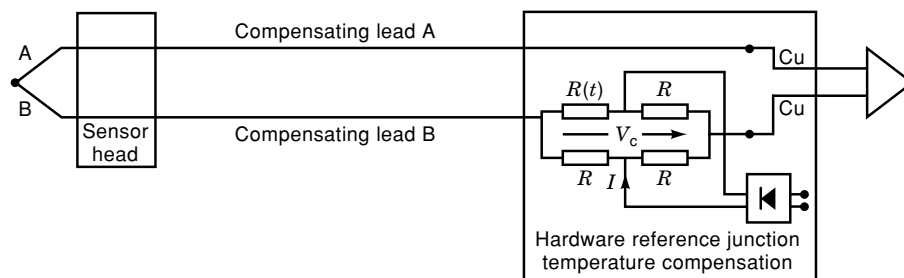


Figure 12. Basic thermocouple circuit with compensating leads which eliminate the influence of the head temperature variations and with reference junction temperature compensation circuit. Resistances R and supply current I are matched according to the type of the thermocouple.

$R(t) = R$ and $V_C = 0$. As the reference temperature increases, $R(t)$ increases accordingly, producing a compensating voltage V_C , equal to the change of the thermal emf corresponding to the difference between the actual temperature and 25°C. The supply current I is matched according to the sensitivity of particular thermocouple. There exist also a great number of integrated circuits for compensation of the reference junction temperature where a diode sensor is used instead of the temperature-sensitive resistor (i.e., Linear Technology 1025). The amplification of a thermocouple signal, together with the reference junction compensation and with some additional functions, is performed by integrated circuits such as Analog Devices AD594, or Linear Technology LTK001.

A method of reference junction voltage correction instead of compensation is also used. It is based on the measurement of the reference junction temperature (for example, by means of a semiconductor sensor), followed by a numerical calculation of the correction value. There is also a common need for the numerical correction of a result in all instruments working with thermocouples, because of the nonlinearity of these sensors. The correction is usually performed by the look-up table method described before.

Another problem is caused by the noise superposing on a relatively weak dc signal transmitted over long compensating leads. In order to avoid the electromagnetically induced voltages, the wires in the compensating leads should be twisted. The protection against the common mode noise is provided by shielding the wires and connecting the shield to a guard terminal of the instrument or transducer. In this way the current flowing through the stray capacitance between the leads and the supply power lines or induced by any source in the grounding loop is shunted and does not affect the measured voltage. The noise voltage may also be suppressed by filtering of the output signal.

A very successful method eliminating all problems due to long compensation leads is to place the whole transducer in the thermometer's head. The current developments in electronic components technology enables design of the compact and temperature-resistant transducers comprising all compensating and linearizing elements and delivering the standard 4–20 mA output signal. Many companies offer such a solution now, and this design seems to be very promising for all immersion-type thermocouple thermometers.

QUARTZ TEMPERATURE SENSORS

The piezoelectric properties of quartz crystal (SiO_2) are applied in the design of extremely precise, stable, and relatively cheap oscillators. The applications of those oscillators are very widespread, from counters and frequency meters to precise timers in clocks and computers. The most important requirement for all those purposes is temperature stability. It is achieved by appropriate cut of the oscillator plate from the quartz crystal. For temperature-invariant oscillators the so-called AT cut is used with the cutting plane inclined to a z axis (optical axis) of the crystal at $+35^\circ$. Any other cut results in a smaller or greater dependence of the oscillator frequency on the temperature. This very property is used in quartz temperature sensors. A plate obtained by a Y cut with the inclination of -4° called HT cut has the highest value of temperature coefficient. The relation between the frequency and temperature may be

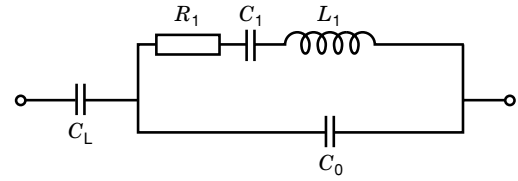


Figure 13. An equivalent circuit of a piezoelectric resonator. C_0 is a geometrical capacity between two electrodes and L_1 , C_1 , R_1 are quartz plate parameters which depend on its mechanical properties. C_L is a load capacity for tuning of the resonance frequency.

written as

$$f(t) = f_0(1 + 90 \times 10^{-6} \Delta t + 60 \times 10^{-9} \Delta t^2 + 30 \times 10^{-12} \Delta t^3) \quad (19)$$

where f_0 is frequency at temperature $t = t_0$ and $\Delta t = t - t_0$. The third and the fourth terms in parentheses introduce nonlinearity, which in this particular case is a great disadvantage because a conventional frequency meter cannot be used as a temperature indicating instrument. Therefore, the LC cut, instead of the HT cut, is used for quartz temperature sensors. The LC cut of a quartz crystal with the cut plane inclined at $+11^\circ 10'$ to the x axis and at $+9^\circ 24'$ to the z axis forms an oscillator with frequency linearly depending on the temperature but with a lower sensitivity (35×10^{-6} 1/K instead of about 100×10^{-6} 1/K with HT cut).

A quartz plate with two electrodes forms a resonator which may be presented in the simplest form as an equivalent electrical circuit, as shown in Fig 13. In the circuit, C_0 is a geometrical capacity between two electrodes and L_1 , C_1 , and R_1 are quartz plate parameters, which depend on its mechanical properties. Two resonance frequencies exist for this circuit: Serial resonance frequency $f_S = 1/2\pi\sqrt{L_1 C_1}$ and parallel resonance frequency $f_G = 1/2\pi\sqrt{L_1 C_E}$, where $C_E = C_0 C_1 / C_0 + C_1$. Both frequencies have close values, because the capacities C_1 and C_E are of the same order. Using of an additional load capacity C_L , the resonance frequency of the plate may be tuned in a limited range between f_S and f_G . Two opposite surfaces of the resonator are coated with thin gold layers. The resonator is mounted in a hermetical case protecting it from atmospheric air moisture. Spring contact elements ensure a low level of mechanical damping. The oscillator frequency f_0 depends on the dimensions of the resonator. It is very convenient to adjust that frequency to such a value that the relationship between the frequency and the temperature may be obtained by simply shifting a decimal point on the frequency meter. Hence if the sensitivity coefficient α is equal to 35×10^{-6} 1/K, the condition is fulfilled for $f_0 \approx 28.6$ MHz, because in that case $f_0 \alpha = 1000$ Hz/K. The tuning feature of the oscillator allows us to meet the above requirement in spite of some differences in individual plates parameters (3). The connection between the sensor and the oscillator must be performed by high-frequency coaxial cable. By the use of frequency meters with high resolution, high resolution of temperature measurements may be achieved too. Much better solution, however, is application of a differential structure of the measuring circuit (Fig. 14) where a mixer forms a low-frequency signal Δf , which corresponds to the difference between the measured temperature t and a reference temperature t_0 : $\Delta f = f - f_0 = f_0 \alpha (t - t_0)$. In such a state the resolution of the frequency meter may be much more

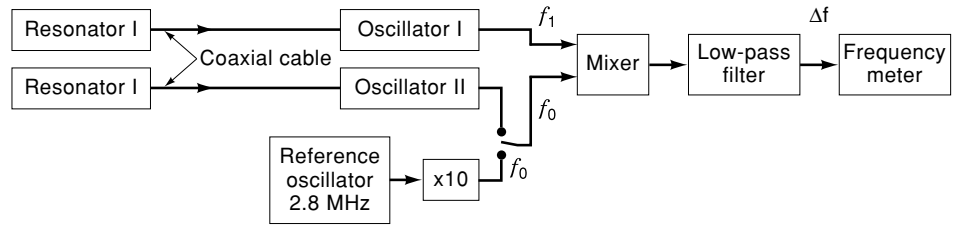


Figure 14. The differential structure of a quartz thermometer.

effectively used. Taking as an example the temperature range from 0°C to 200°C, the value of $f_0\alpha$ equal to 1000 Hz/K, and the 61/2 digit resolution of the frequency meter (which is a common practice), a 0.1 mK temperature resolution is achieved. This extremely high resolution is a reason why quartz thermometers are commonly equipped with two sensors allowing the measurement of temperature difference. In many practical cases, it is not the absolute value of temperature but the difference of temperatures that has to be known with a great accuracy. A double quartz thermometer is an excellent instrument for this purpose. Please note that the meaning of the term “absolute value” is used here differently than “absolute temperature scale.” The uncertainty of the quartz thermometers depends primarily on aging and relaxation. Single-point recalibration from time to time and avoidance rapid temperature shocks are therefore highly recommended. With these conditions met, the uncertainty of ± 50 mK may be sustained for a long time.

RADIATION THERMOMETRY

The principle of radiation thermometry is the measurement of the thermal energy emitted by radiation from the object under inspection. It is not the whole energy that is measured but only its very little part, corresponding to the radiation focused on the radiation-sensitive element placed in the thermometer. The essential difference between the radiation thermometry and all other methods of temperature measurements is the lack of the thermal equilibrium between the object and the sensor. With radiation thermometry the only way of thermal energy transfer from the object to a sensor would be electromagnetic wave propagation in the range from ultra-violet radiation (0.2 μm wavelength) through visible spectrum (0.38 μm to 0.78 μm) up to far-infrared radiation (50 μm). Thermal energy transfer does not lead to equalizing of the temperatures but only excites the sensor. The excitation level, and consequently the output signal of the sensor, depends on the portion of the received energy. Proper design, together with the proper use of a radiation thermometer, ensures a strictly defined dependence of that portion of energy and the temperature of the object. All other object properties and radiation parameters such as emissivity, dimensions, the distance to the sensor, atmosphere composition and temperature, and radiation of other objects including the thermometer body and many others, either have to be kept at the same level as during the thermometer calibration or must have no influence on the sensor. These requirements seem to be more serious than in conventional thermometers since they are related to the environment in which the measurement is performed, rather than to the sensor and the instrument themselves.

In order to answer the question of how to meet the above requirements, some essential properties of thermal radiation have to be considered. The spectral density of heat energy emitted by a black body is given by Planck's law

$$M(\lambda) = C_1 \lambda^{-5} \frac{1}{e^{C_2/\lambda T} - 1} \quad (20)$$

or in a simplified (but sufficient for our discussion) form of Wien's law

$$M(\lambda) = C_1 \lambda^{-5} e^{-C_2/\lambda T} \quad (21)$$

In both equations, $C_1 = 37.4 \text{ mW} \cdot \mu\text{m}^4/\text{cm}^2$ and $C_2 = 14,388 \text{ K} \cdot \mu\text{m}$. The lower the temperature of the radiation source, the longer the wavelength of the emitted radiation (Fig. 15). The product of the temperature and the wavelength corresponding to the maximum of spectral density remains constant, according to Wien's law of displacement: $T\lambda_{\text{max}} = 2899 \text{ K} \cdot \mu\text{m}$. The measurement of low temperatures requires the use of the sensors and of the methods which detect the radiation in far-infrared spectrum.

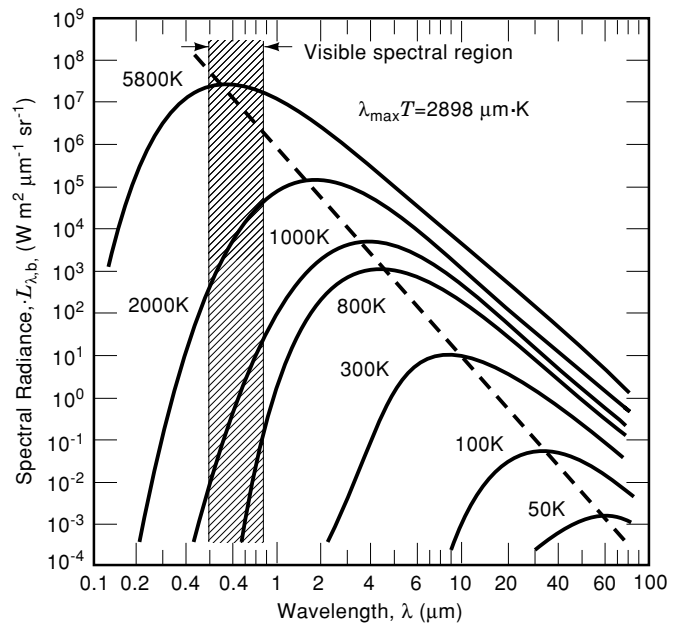


Figure 15. Spectral heat radiation density. Dotted line indicates the Wien's law. After Ref. 8. The radiance energy declines dramatically with the temperature. It is the main reason why low temperature radiation thermometers were developed many years later than high temperatures ones.

The total thermal energy emitted by a surface of a black body with the area F is an integral of the Wien's equation:

$$E_{BB} = F \int_0^\infty M(\lambda) d\lambda \quad (22)$$

The black body is defined as an object which does not reflect thermal radiation and therefore, according to Kirchhoff's law of radiation (absorptive power = emissive power), emits the whole radiant energy relative to its temperature. The emissivity factor ϵ of a black body is equal to 1. To simulate a black body (e.g., for the calibration of radiation thermometer), a closed cavity with a relatively small aperture may be used. The inner surface of the cavity has to be specially shaped. In the radiation thermometry practice, only some objects (such as a hearth of a furnace) may be treated as black cavities. All other objects, and especially all objects in the air, have the emissivity ϵ smaller than 1, and their radiation density has to be multiplied by ϵ . The magnitude of emissivity depends on the material, surface finish (polishing, oxidization, roughening, deadening), its temperature, incident and viewed angles of heat flux direction, and polarization of radiation. Furthermore, the emissivity depends on the wavelength too. Therefore, the whole heat energy emitted by a uniform surface F observed by the radiation thermometer may be expressed as

$$E = F \int_0^\infty \epsilon(\lambda) M(\lambda) d\lambda \quad (23)$$

where $\epsilon(\lambda)$ is usually known with a very poor accuracy.

Next we take into consideration the properties of the atmosphere that the radiation is passing through. Application of radiation thermometers with high-temperature objects is always disturbed by the presence of smoke and dust particles absorbing the radiation. The blow of purging air is used to clear the optical path between the object and the thermometer and to protect the optical parts from contamination with dust. Nevertheless, the outer surface of the instrument optics has to be cleaned from time to time.

In measurements performed in open air the concentration of absorbing gases is considerably lower, but the distance between the object and the instrument is much greater so the absorption plays a significant role too. The contents of CO, CO₂, and water vapor in the air are most significant. The spectral distribution of the absorption caused by these gases is not uniform and shows extremely great variations. Only two bands of thermal radiation wavelength may be indicated as almost free from absorption. These are 3.5 μm to 4.5 μm (near atmospheric window) and 8 μm to 13 μm (far atmospheric window). Hot gases disturb the measurement process by their own radiation too. According to Kirchhoff's law, the spectral distribution of emissivity is the same as the distribution of absorption, and it declines in the spectral ranges of atmospheric window. For these two reasons, it is highly recommended that radiation thermometers operate within one of those windows.

Considering the spectral disturbances, two additional aspects must not be overlooked: (1) the spectral transmittance of the materials used for optical parts of the thermometers (lenses, windows, and filters) and (2) the spectral sensitivity of radiation detectors. Glass optic elements may be used in the range of visible and near-infrared radiation up to 2 μm

wavelength that corresponds to measured temperatures over above 500°C. Quartz lenses enable slight widening of that range (up to 3.5 μm). For these thermometers, however, which cover a much lower temperature range, very special materials have to be used. These are ZnSe (0.6 μm to 16 μm), GaAs (1.2 μm to 12 μm), or CdTe (1.7 μm to 25 μm) and some other materials, such as Chalcogenid, KRS-5, or specially prepared ZnS. All of them (except ZnSe) are not transparent to visible light, and therefore the optical path for visual radiation must be simultaneously used to aim at the target.

Infrared Radiation Detectors

Two groups of infrared (IR) detectors are presently used in radiation thermometry. These are thermal detectors with low but spectrum-independent sensitivity and semiconductor photon detectors (IR diodes and IR photovoltaic sensors), much more sensitive but working in the limited spectral zones. Thermopiles, bolometers, and pyroelectric detectors are the thermal detectors.

Thermopiles consists of a large number (of up to 66) of thermocouples with hot junctions concentrated on a small surface and exposed to the thermal radiation flux, along with reference junctions kept at a temperature close to the ambient temperature (Fig. 16). The thermocouple materials are Bi-Sb or Ag-poly Si, with a hot junction deposited on a very thin (0.7 μm to 1 μm) silicon membrane isolated with an SiO₂ layer and with a reference junction deposited on bulk silicon material, which forms a frame around the membrane. A low thermal conductivity of the thin membrane secures proper thermal insulation between hot and reference junctions. The detector is fabricated by Micro Systems Technology (MST) and may be integrated with signal conditioning elements or even A/D converters on one chip with dimensions not exceeding a few millimeters. These thermocouples possess a very high thermoelectric power of approximately 100 $\mu\text{V/K}$, comparing with a few $\mu\text{V/K}$ for conventional metal thermocouples. This ensures high sensitivity of the detector expressed in volts per watt of thermal energy:

$$S_D = \frac{V_{\text{out}}}{E} \quad (24)$$

Thermal noise is the factor limiting the possibilities of thermal energy measurement. The Johnson noise equivalent voltage is given by

$$V_N = 2\sqrt{kTR\Delta f} \quad (25)$$

where k is Boltzmann's constant, R is the detector resistance, T is the detector temperature, and Δf is the bandwidth of the associated amplifier, determined either by the chopper frequency or by the detector speed. Substituting Eq. (24) in Eq. (25), one achieves the noise equivalent power (NEP) in the form

$$\text{NEP} = E_{N\text{min}} = \frac{2\sqrt{kTR\Delta f}}{S_D} \quad (26)$$

The reciprocal of NEP is called detectivity $D(\lambda)$. Hence the sensitivity S_D is proportional to the square root of detector area A , and the frequency band Δf is determined by the amplifier rather than by the detector itself. The properties of the detector are better described by specific spectral detectivity $D(\lambda)$,

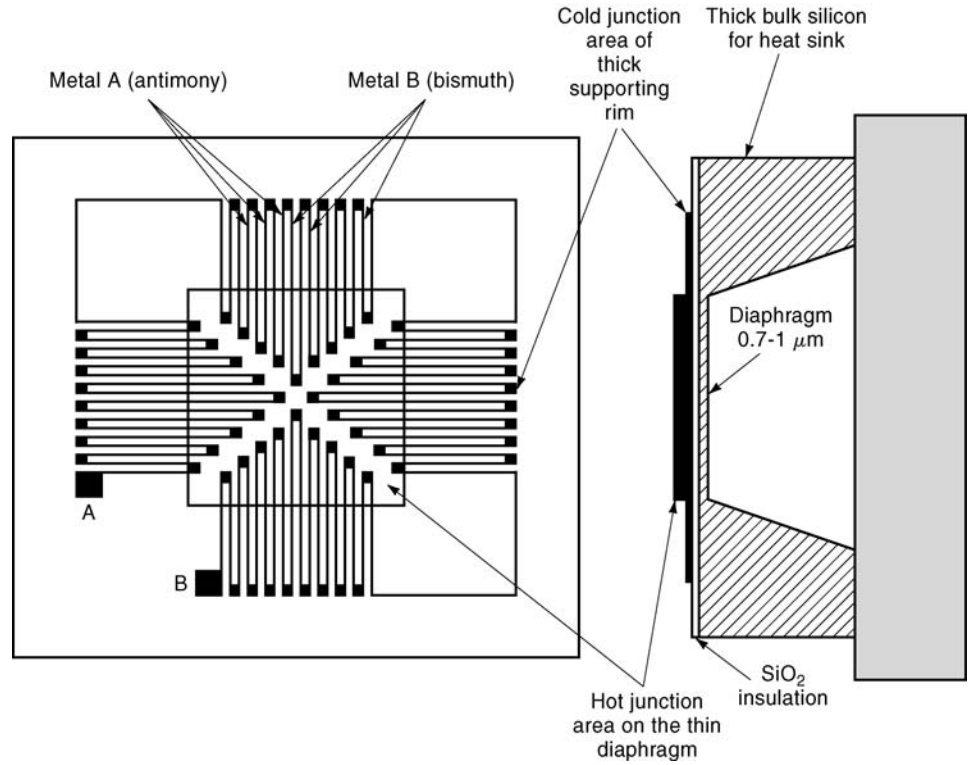


Figure 16. Thermopile manufactured in Micro Machining Technology.

defined as

$$D^*(\lambda) = D(\lambda)\sqrt{A\Delta f} \quad (27)$$

Thermopile specific spectral detectivity is usually about $10^8 \text{ cm}\cdot\text{Hz}^{1/2} \text{ W}^{-1}$ (Fig. 17).

Bolometers with thermistors used as temperature-sensing elements have the same detectivity but are less frequently used. Pyroelectric detectors where the electrical charge evoked on the piezoelectric plate follow the temperature changes are a few times more sensitive, but they do not work at steady-state conditions and the thermal flux must be mechanically modulated by means of a rotating disk with slots or holes. Mechanical chopping enables the use of alternating-current (ac) amplifiers; this is more convenient than using direct-current (dc) ones and additionally cuts off a large portion of low-frequency noise (red noise), thus increasing the specific spectral detectivity. For this reason, the chopping technique is in common use now.

Photon detectors are much more sensitive than thermal detectors and are mostly used when the limited range of wavelength is preferred. They are manufactured as semiconductor photodiodes or photovoltaic elements. The most sensitive Si detectors [$D(\lambda) = 10^{13} \text{ cm}\cdot\text{Hz}^{1/2} \text{ W}^{-1}$] are suitable for visual spectra only; PbS, InAs, and InSb detectors [$D(\lambda) = 10^{11} \text{ cm}\cdot\text{Hz}^{1/2} \text{ W}^{-1}$] cover the range up to $3 \mu\text{m}$ to $5 \mu\text{m}$; and only an Hg–Cd–Te detector may be used up to the $20 \mu\text{m}$ range, achieving its maximum detectivity [$D(\lambda) = 5 \times 10^{10} \text{ cm}\cdot\text{Hz}^{1/2} \text{ W}^{-1}$] in the far-infrared radiation region. Cooling of detectors in order to improve their detectivity by noise reduction results sometimes in shifting of the spectral characteristics toward longer wavelengths. This effect is observed in particular for Ge detectors. Therefore, in comparing the spectral detectivities of various

detectors, precise specification of their operating temperatures must be taken into account (8). Detector cooling is frequently performed by single-stage or multistage Peltier coolers, which are compact and easy to use. The only problem to solve in their

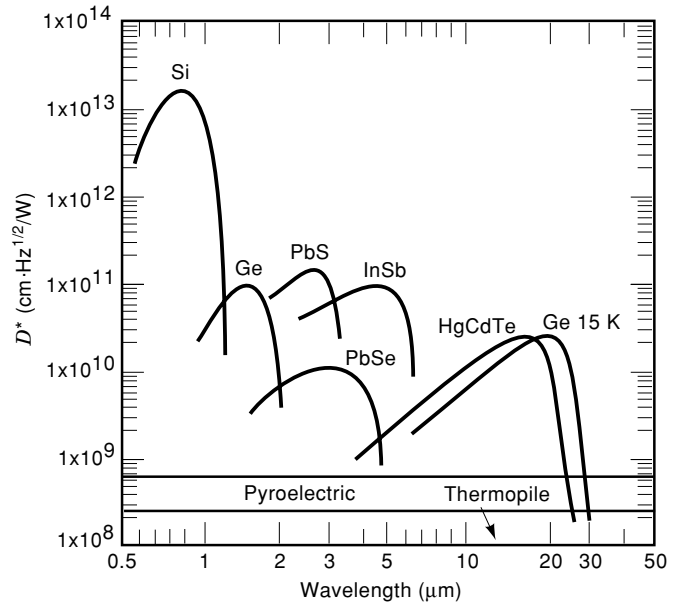


Figure 17. Specific spectral detectivity of thermal and photon infrared detectors. The design of low temperature radiation thermometers is more complicated not only because of the lower radiance energy but by the lower sensitivity of the photon detectors working in the far IR region.

design is providing quick dissipation of the heat to enable the continuous cooling of the detector. Stirling coolers which require pumps are seldom used. It is quite possible, however, that in the near future, micropumps for that purpose will be produced by Micro Machining Technology.

Wide-Band Radiation Thermometers

Conceptually the simplest method of radiation thermometry is to detect the thermal energy emitted by an object toward a sensor regardless of its spectral distribution. This energy depends on the temperature according to the Stefan–Boltzmann law:

$$E = \epsilon_0 \sigma T^4 - \epsilon_t \sigma T_t^4 \quad (28)$$

where σ is the Stefan–Boltzmann constant, T and T_t are the object and thermometer temperatures, respectively, and ϵ_0 and ϵ_t are the emissivities of the object and the thermometer, respectively. The exponent 4 in Eq. (28) makes the calibration curve nonlinear and increases the uncertainty at the lower range of measured temperatures where the sensitivity is also low. This form of relationship, however, decreases the influence of the object emissivity ϵ_0 . The thermometer calibrated with the use of a black body ($\epsilon_0 = 1$) always indicates lower temperature T_{in} than is actually existent at the observed surface with the emittance ϵ_0 . By the comparison of equal states of the detector output signals during calibration and measurement, and assuming that all other factors are the same, one obtains

$$T = \frac{1}{4\sqrt{\epsilon_0}} T_{in} \quad (29)$$

[The second term in Eq. (28) has been neglected, which is allowed in case of higher object temperatures.] For example, taking $\epsilon_0 = 0.5$, the real temperature is not twice the indicated temperature but only 19% higher, as expressed in kelvins. Nevertheless, such a great difference is not negligible and has to be corrected. The simplest and the most common way is to adjust the emissivity value in the instrument which calculates and introduces the appropriate result correction. It would be a good method if the values of ϵ_0 were known with sufficient accuracy, but usually this does not take place. Furthermore, ϵ_0 is used here in the meaning of average emissivity over the whole spectral range. It has been shown before that for many reasons the heat flux incident at the sensor is spectrally disturbed, and hence the averaging should be weighted with regard to all these disturbances. Calculations become complicated to such a degree that their practical usefulness becomes questionable. The only way is then to calibrate the thermometer at the same environmental conditions as its operational conditions. The calibration is more effective when influence variables are stable. Higher stability may be achieved by narrowing the wavelength band. For this reason the class of wide-band thermometers includes not only total radiation thermometers but also thermometers with intentionally limited spectral bands. Total radiation thermometers are commonly equipped with thermal detectors and are used for measurements inside closed cavities as tanks or ovens where the emissivity is equal to 1 and the distance to the target is relatively low and stable. For applications in open air the thermometers with narrowed band and adjustable emissivity are used. They work in the first atmospheric window (for higher temperatures) or in the second atmospheric window

(for lower temperatures). For lower temperature ranges the compensation of ambient temperature [T_t in Eq. (28)] is commonly performed. Some of the devices are referred to as “near touch” radiation thermometers, in which an infrared detector is placed immediately over the surface, and the mirror optic is used to focus the radiant energy at the detector.

The majority of wide-band radiation thermometers are produced as hand-held or tripod-based instruments, commonly equipped with optical or laser aiming facility. Because of the progress in noiseless amplification of weak electric signals, the lowest temperature range of the discussed instruments has been pushed down to the level of -50°C or even -100°C . It is no longer true that those radiation thermometers are destined for high temperatures only, as it was in the past. It still remains true, however, that accurate measurements may be obtained only in the case of sufficiently stable conditions. The responsibility for ensuring repeatable conditions is with the user. The accuracy of the measurement depends rather on his skills and experience than on the quality of an instrument.

Monochromatic Radiation Thermometers

The name for the device derives from the times when radiation thermometers were used in the visual band only and indicates that a very narrow spectral band $\Delta\lambda$ of emitted heat flux is used for measurement purposes. In optical pyrometers the narrow band was being filtered using colored windows. Now the infrared interference filters are used for that purpose. Interference filters consist of a number of thin transparent layers deposited on a substrate. The optical properties and the thicknesses of the layers are specially matched to transmit through only a desired wavelength band $\Delta\lambda$ with the middle wavelength equal to λ_1 , as well as to reflect all higher and lower band frequencies. The filter remains cool because the undesired heat energy is reflected and not absorbed. The filtered bands are matched first of all with respect to the temperature range of the thermometer, but also with respect to the atmospheric windows. The part of energy emitted in wavelength band $\Delta\lambda$ may be described as

$$E_{\lambda_1} = \epsilon_{\lambda_1} C_1 \lambda_1^{-5} e^{-C_2/\lambda_1 T} \Delta\lambda \quad (30)$$

where ϵ_{λ_1} is the emissivity at wavelength λ_1 . The volume of energy incident at the detector is much lower than that in wide-band thermometers, but narrow-band photon detectors with much higher detectivity may be used in this case. Assuming the same calibration procedure as described before for wide-band thermometers, one obtains the relationship between the real temperature T and indicated temperature T_{in} in the form of

$$\frac{1}{T} = \frac{1}{T_{in}} + \frac{\lambda_1}{C_2} \ln \epsilon_{\lambda_1} \quad (31)$$

and for the small differences between T and T_{in} in the form of

$$T \approx T_{in} \left(1 - \frac{T_{in} \lambda_1}{C_2} \ln \epsilon_{\lambda_1} \right) \quad (32)$$

The differences between real and indicated temperatures are lower than in wide-band thermometers [Eq. (29)]. However, it is more important that the value of ϵ_{λ} be better defined and more stable than an average emissivity ϵ_0 which has been used

for calculations in the case of wide-band thermometers. Furthermore, all disturbances such as atmospheric absorption, hot gas radiation, contamination of optical parts by dust particles, and variations in detector properties have no influence on the output signal so far, because they do not concern directly the filtered band $\Delta\lambda$. The maintenance of the monochromatic thermometers is therefore simpler and their stability is much better, but the problem of emissivity correction remains.

Radiation Ratio Thermometers

The idea of radiation ratio thermometers is based on simultaneous measurement of radiation in two narrow wavelength bands followed by calculating the temperature from the ratio of the results. Both bands have to be chosen on the “increasing” part of the Wien’s spectral energy distribution curve (see Fig. 15), and they differ for each temperature range. Denoting the energy ratio at wavelength λ_2 and λ_1 by R , one obtains

$$T = \frac{\lambda_2 - \lambda_1}{\lambda_2 \lambda_1} C_2 \left[\ln R \left(\frac{\lambda_2}{\lambda_1} \right)^5 \right]^{-1} \quad (33)$$

for all gray bodies independent from their emissivity. For nongray bodies, like all metals, in which the emissivity is a function of wavelength, a very small correction is needed:

$$T \approx T_{\text{in}} \left(1 - T_{\text{in}} \frac{\lambda_2 \lambda_1}{C_2(\lambda_2 - \lambda_1)} \ln \frac{\epsilon_1}{\epsilon_2} \right) \quad (34)$$

where ϵ_1 and ϵ_2 are emissivities for wavelength λ_1 and λ_2 , respectively. The difference between ϵ_1 and ϵ_2 is very low, even in metals, which causes, according to Eq. (34), a very little difference between T and T_{in} . The ratio thermometers may be considered emissivity-independent instruments. The design of radiation ratio thermometers is much more complicated than other radiation thermometers, and special requirements for low chrominance optic elements have to be fulfilled. Therefore this kind of instruments is rather expensive.

In summary, it is worthwhile to point out that in the last decade an intensive progress in radiation thermometry has been achieved. The basic ideas remain unchanged, an optical pyrometer with disappearing filament remains a prototype of narrow band thermometers, and a two-color pyrometer is an ancestor of the ratio thermometers, but the progress in detector technology, the introduction of interference filters, the possibilities of low noise signal amplification, and numerical result corrections make the use of these instruments simpler and allow moving the operational temperature range toward lower temperatures. The use of fiber optics to transmit the thermal energy from hardly accessible places or through the areas with nuclear radiation or strong electromagnetic fields creates the next step of development which cannot be underestimated. Up to now the optical fibers work at short wavelengths only (up to $2.5 \mu\text{m}$), but with low distances the temperatures as low as 100°C may be measured; further progress in that field is expected.

ERRORS AND UNCERTAINTIES

First of all we have to distinguish between errors and uncertainties. According to its definition, an error is a difference

between a result of the measurement and a measurand true value. Errors always have certain values and a sign, but they remain unknown. In order to determine an error, we can try to calculate it or to measure it. Errors are measured in the calibration process. A determined error added to the result with the opposite sign as a correction makes the result free from that particular error. Many errors exist which are very hard to determine, even if their sources are theoretically known. In radiation thermometry we know exactly how the emissivity influences the result, but we do not know the particular value of the target emissivity. We are only able to estimate the range of results which is affected by more or less unknown sources of errors. It is a range of uncertainty, and the most probable value of the result is commonly (but not always) placed in the middle of this range. Therefore, uncertainty is denoted as $\pm\Delta t$, where Δt is half of the estimated range.

All predictable errors in sensors or transducers are normally accounted by the producer by means of compensations and corrections. The accuracy data which are found in transducer specifications or in standards are uncertainty limits. ISO Guide For Expression of Uncertainties (10) distinguishes two kinds of uncertainties: type A and type B. Type A uncertainties are easy to detect and estimate by repetition of measurements performed in the same conditions and then by calculating the standard deviation of the results (classical random uncertainties). Type A uncertainties dominate in high accuracy measurements. Random effects observed in industrial measurements are caused by measurand instability and environmental variations rather than by measuring instruments themselves. Type B uncertainties are those which remain constant by repetition of measurements. They may be caused by residuals remaining after nonlinearity compensation, hysteresis effects, or they may be caused by the influence of variables such as pressure, nuclear radiation, electromagnetic fields, humidity, dust, lead temperature variations, ambient temperature variations, velocity of the medium under measurement, aging, and many others. Type B uncertainties dominate in medium accuracy thermometers. Estimation of type B uncertainties needs some investigations performed either by the manufacturer of the thermometer or by the user, and some amount of experience too. It is a more complicated task than a type A estimation but the uncertainty data given in thermometer specification may be adequately used here. The calibration of a particular sensor or transducer transforms some uncertainties into errors, allowing us to account for them in a form of correction. All the rest remains as a part of a type B uncertainty. Laboratory calibration of temperature sensors is expensive, but calibration performed at the site of thermometer installation is more expensive and sometimes even impossible. On the other hand, it is impossible to restore all measurement conditions during laboratory calibration, and thus the improvement of accuracy by the recalibration is always limited.

Two remarks dealing with specification sheets ought to be pointed out here. In thermometry the uncertainty data referred to as “% of reading” or “% of full scale” (%FS) have no sense because temperature scales have no physical zero, and $\pm 2\%$ of 273 K gives $\pm 6 \text{ K}$ but $\pm 2\%$ of 0°C gives $\pm 0^\circ\text{C}$. All uncertainties have to be referred to in temperature units.

The next remark is concerned with resolution. Resolution is not a measure of accuracy and normally has nothing to do with uncertainty. It is only the ability to indicate small differences in temperature. High resolution may sometimes be very useful,

but we ought to have in our minds that by the use of numerical result display the resolution is commonly 10 times better than the uncertainty involved.

In the sensor-transducer pair the uncertainty of sensor usually dominates. Excluding standards from our considerations, it may be stated that of all temperature sensors, metal resistance thermometers are the ones with the lowest uncertainty. In the temperature range of -50°C to $+150^{\circ}\text{C}$, quartz thermometers are much more accurate and stable than all semiconductor sensors (the fact is reflected in their price). Traditional NTC thermistors seem to be less attractive in comparison with modern linear semiconductor sensors and more and more cheap integrated sensors which have the advantage of standardized calibrated output signal. However, the uncertainty of all semiconductor sensors is not better than $\pm 0.5^{\circ}\text{C}$ to 1°C . Thermocouples are the less accurate sensors. Their uncertainty is rapidly increasing, with the measured temperature reaching up to ± 10 K at temperatures above 1000°C . This is a much higher value than for radiation thermometers where the uncertainty is rather optimistically referred to as ± 1 K to 2 K. The advantage of thermocouple sensors is their lower cost, but the costs of transducers are comparable now. Some radiation sensors are equipped with such a signal conditioning system that their output signal fits to conventional thermocouple transducers. The thermocouples, however, are the unique sensors which have to be immersed in the high-temperature medium.

The discussion presented above deals with the instrument uncertainty, but more serious sources of uncertainty are related to the methods of temperature measurement. Generally, the problem is in the difference between the temperature of a sensor and the temperature which ought to be measured. This problem will be considered separately for steady-state conditions and for dynamic conditions.

Steady-State Errors of Temperature Measurement

In order to provide the same temperature of an object and of a sensor, the conditions for easy heat transfer from the object to the sensor and simultaneously for the difficult heat transfer from the sensor to the ambient environment should be created. For immersion-type sensors the length of the immersed part should be as large as possible because the heat runs away from the sensor through the protective tube, and a higher temperature gradient along the shield facilitates the transfer. Thermal insulation of the tank or pipe line where the temperature is measured lowers the gradient, and therefore the error too. The problem is much more serious in the case of gas media because of their low heat transmission coefficient and the possibility of heat escape to the pipe walls by radiation. The problem weight increases with the measured temperature, not only because of the higher temperature gradient but also due to the need for increasing shield thickness and for better thermal insulation of the sensor by protective ceramics.

The measurement of surface temperature is the best case for obtaining a false result. The sources of errors are the same, but the available remedies are very limited. The typical situation is schematically presented in Fig. 18. The sensor ought to have the best thermal contact with the surface but simultaneously should be insulated from the environment. The thermal insulation provides protection from the temperature decrease at the point of measurement, but perhaps it may cause a local increase

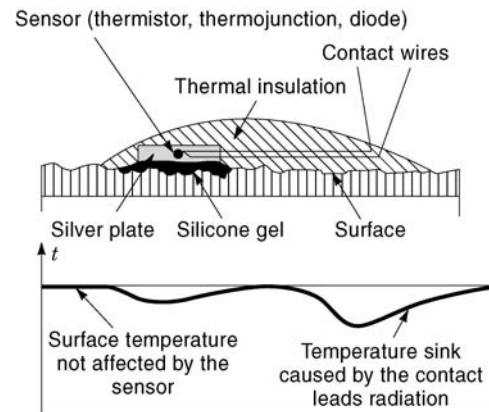


Figure 18. The measurement of surface temperature. The heat is accumulated in the silver plate with high thermal conductivity in order to equalize the surface and sensor temperatures. The air gap between the surface and the plate ought to be filled with silicon gel. The connection leads act as radiators and therefore run along the surface to a distance far enough to avoid the surface temperature disturbance near the sensor.

of the temperature especially when the surface was intensively cooled. Touching of the surface by the sensor involves thermal effects which are different in each particular case and practically unpredictable. From this point of view, the advantage of radiation thermometry for surface temperature measurements is evident.

Dynamic Errors

Dynamic errors are caused by the thermal inertia of the sensors and become important while transient temperatures have to be measured. The dynamic properties of a sensor are described by the response time of the output signal after a rapid change of sensor temperature, or by the time constant. The response time is usually defined as the time elapsing between 10% and 90% of the output signal change, but other definitions are also used. The response time depends on the sensor properties such as its material, shape, and dimensions, but depends first of all on the environment surrounding the sensor, characterized by the heat transmission coefficient. That coefficient is a few times greater in liquids than in gases, and it increases with the velocity of the medium. With any comparison of the sensor dynamic properties, exactly the same conditions have to be secured.

Assuming the linearity of the thermometer the idea of Laplace transformation and transfer function may be used for describing the thermometer dynamic properties. The simplest model of the dynamic behavior of a sensor may be presented in a form of the first-order transfer function:

$$K(s) = \frac{T_T(s)}{T_O(s)} = \frac{1}{1 + s\tau_1} \quad (35)$$

where $T_T(s)$ is the Laplace transformation of thermometer temperature, $T_O(s)$ is the Laplace transformation of the object temperature, s is the Laplace operator and τ_1 is a time constant. The model is valid for nonembedded sensors only, which are rather seldom used. Two other models are also used. The second-order

model

$$K_{II}(s) = \frac{1}{(1 + s\tau_1)(1 + s\tau_2)} \quad (36)$$

accounts for the delay of the output signal and better describes the thermometer with the shielded sensors. A model with the differential action

$$K_{III}(s) = \frac{1 + s\tau_D}{(1 + s\tau_1)(1 + s\tau_2)} \quad (37)$$

may in turn be used for surface temperature thermometers.

All those models may be treated as rough approximations because in fact the dynamic properties of temperature sensors are nonlinear. The experiment has shown that even for a nonembedded thermocouple sensor the time constant at 400°C has been three times larger than at that 1200°C (5). Also the warming curves and quenching curves always differ from one another.

For dynamically nonlinear sensors, it is not so easy to transform the results from one operating condition to other. One of the methods for estimation of dynamic behaviors of a sensor under its operating conditions uses an intentionally generated self-heating impulse in order to record the thermometer answer and then to calculate its dynamic parameters.

The dynamic behavior of the radiation thermometers depends on the dynamic properties of the infrared sensors which are rather slow, with the response time varying from 0.2 s to 2 s, and which are dynamically nonlinear too.

BIBLIOGRAPHY

1. H. Preston-Thomas, The international temperature scale of 1990 (ITS-90), *Metrologia*, **27**: 3–10, 1990.
2. T. J. Quinn, *Temperature*, 2nd ed., New York: Academic Press, 1990. Deep study of ITS and high accuracy temperature measurements.
3. W. Göpel, J. Hesse, and J. N. Zemel (eds.), *Sensors, A Comprehensive Survey*, Vol. 4, *Thermal Sensors*, T. Ricolfi, J. Scholz (eds.), Weinheim: VCH, 1990.
4. *The Temperature Handbook*, Stanford, CA: Omega Engineering, 1992. A comprehensive review of the market of thermometers and thermal sensors.
5. L. Von Körtrelyessy, *Thermoelement Praxis*, 2nd ed., Essen: Vulkan Verlag, 1987. In German. Comprehensive work with a lot of practical information.
6. J. F. Schooley, *Thermometry*, New York: CRC Press, 1986.
7. L. Michalski, K. Eckersdorf, and J. McGhee, *Temperature Measurement*, New York: Wiley, 1989.
8. D. P. De Witt and G. D. Nutter (eds.), *Theory and Practice of Radiation Thermometry*, New York: Wiley, 1988.
9. *The Infrared Temperature Handbook*, Stanford, CA: Omega Engineering, 1994.
10. *Guide to the Expression of Uncertainty in Measurement*, ISO/IEC/OIML/BIMP, Printed in Switzerland 1993.
11. J. Zakrzewski, Temperature Sensors, in J. G. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 21, New York: Wiley, 1997, p. 605.

JAN ZAKRZEWSKI
Silesian Technology, University
of Poland

THERMISTORS

In the conventional Ohm's law equation

$$v = Ri \quad (1)$$

it is well known that the resistance value R corresponds to a specific temperature. Every two-terminal resistor invariably possesses a resistance–temperature (R – T) characteristic. Three typical shapes of such a characteristic are shown in Fig. 1 over a temperature range 0° to 100°C (273.15 to 373.15 K). In curve (a), the resistance is almost invariant with temperature, this property being exhibited by materials like manganin or constantan. Curve (b) corresponds to a material having a positive temperature coefficient of resistance. Examples of such materials are metals like platinum, copper, nickel, and others. A characteristic as in curve (c) is indicative of a large negative temperature coefficient of resistance and is encountered with resistors made out of semiconducting materials such as Mn_2O_3 , Co_2O_3 , and NiO (1–2). Thermistors belong to this class of resistors and derive their name from the phrase “thermally sensitive resistors.” Their properties and applications are discussed in greater detail in the following sections.

TEMPERATURE-DEPENDENCE OF RESISTIVITY: METALS AND SEMICONDUCTORS (3)

The value of the resistance R referred to in Eq. (1) depends on the physical dimensions of the resistor and the resistivity ρ of the material used. The change in resistance with temperature is due mainly to the change in ρ or its reciprocal, the conductivity σ . The nature of the change is different for metals and semiconductors. A given solid is classified as a conductor if, in its atomic model, the energy gap between the valence and the conduction bands is absent, with the two bands overlapping. For semiconductors, however, there exists an energy gap ranging from 0.1 eV to 3 eV between the two bands. The value of σ depends on the number of charge carriers n_c available in

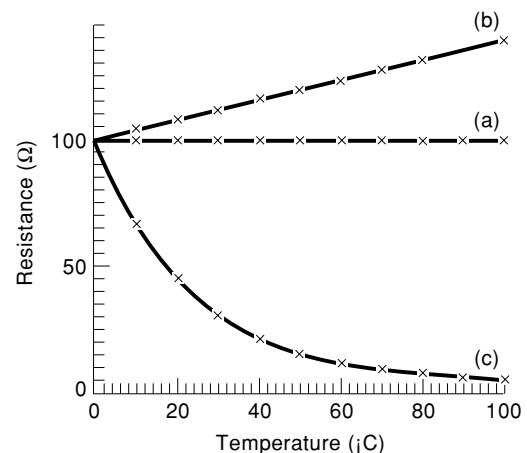


Figure 1. Typical shapes of R – T characteristics of resistors: (a) with negligible temperature coefficient of resistance, (b) with positive temperature coefficient of resistance, (c) with negative temperature coefficient of resistance.

the conduction band and the velocity with which these carriers move under the application of an electric field. The latter is directly related to the mobility μ of the charge carriers. In the case of metals, n_c does not vary appreciably with temperature. The contribution to the change in resistivity comes from a change in μ . As temperature is increased, the enhanced thermal agitation of the atoms in the crystal lattice decreases the mean free path of a charge carrier between two successive collisions. This causes a decrease in μ and a consequent increase in the resistivity ρ . For semiconductors, on the other hand, an increase in temperature causes a large number of charge carriers to move into the conduction band. The resultant increase in n_c more than offsets the effect of the decrease in μ . Semiconductors thus exhibit a negative temperature coefficient of resistance, whose magnitude is several orders higher than that observed in the case of metals. The temperature coefficient of resistance of a semiconductor would lie between -1 and -5% per K, compared with a value of around $+0.4\%$ per K for copper and platinum.

CONSTRUCTION OF THERMISTORS

It is only through special processing that germanium or silicon can be had in pure form. Such intrinsic semiconductors, no doubt, possess a large temperature coefficient of resistance. Their conductivity at ordinary temperatures is, however, too low for practical resistors to be made out of them. Commercial thermistors are therefore basically compound semiconductors, which are made of oxides of cobalt, copper, manganese, nickel, tin, titanium, and others. While the R - T variation of a compound semiconductor is similar to that of intrinsic germanium or silicon, the increase in number of charge carriers is traceable in this case to a different phenomenon involving oxygen atoms (1).

Thermistors are manufactured in different shapes such as disk, rod, bead, and washer (1,4). A few sample shapes are illustrated in the sketches of Fig. 2. The various stages involved in the manufacture of thermistors are:

- Mixing of the various ingredients (metallic oxides) and grinding of the mixture to obtain a fine homogeneous powder.
- Binding of the powder by using a suitable organic chemical and shaping to the required form.
- Sintering at a controlled temperature. (Sintering is a process in which a powder mix of various ingredients is heated to a temperature below the melting point of the major constituent to achieve inter-particle bonding).
- Fixing of connecting leads and encapsulating in glass, epoxy, or a ceramic sheath.

Table 1. Typical Parameters for a Few Commercial Thermistors

Type of Thermistor	R_0	A	B
Siemens K 19	10 k Ω	9.754 m Ω	3440 K
Omega 44005	3 k Ω	5.586 m Ω	3934 K
YSI part No. 44033	2.25 k Ω	3.817 m Ω	3958 K

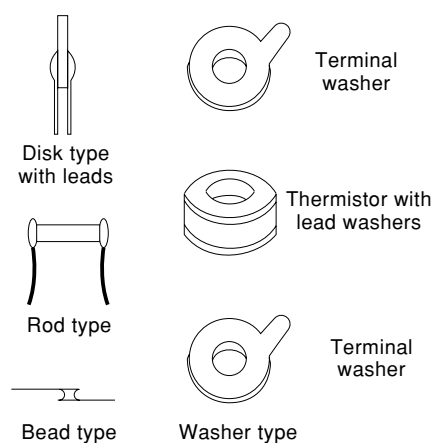


Figure 2. Various types of thermistors.

The nominal resistance of the thermistor, its temperature sensitivity and other relevant properties depend on the proportions of the constituents, the size, and the sintering temperature.

A QUANTITATIVE RESISTANCE-TEMPERATURE RELATIONSHIP WITH SIMPLIFIED THEORY (1,5)

The charge carriers contributing to current in a semiconductor include both electrons and holes. The conductivity due to each type of charge carrier is proportional to the product of its concentration (number of charge-carriers per unit volume) and mobility. The charge-carrier concentration at any temperature is governed by the Fermi-Dirac distribution theory. Application of this theory to electrons would yield the following expression, known as the Boltzmann equation, for the electron-concentration n_e :

$$n_e = N e^{-E_g/2kT} \quad (2)$$

where N is the concentration in valence band, E_g is the energy gap between the conduction and valence band, k is the Boltzmann constant and T is the absolute temperature. A similar expression is valid for holes too. As mentioned in the previous section, the variation of μ with T can be ignored in semiconductors. The conductivity is thus directly proportional to the charge concentration and hence bears an exponential relation to temperature similar to Eq. (2). With such a relation inserted into the expression for the resistance R of the device, we would get

$$R = \frac{\ell}{a\sigma} = \frac{\ell}{a\sigma_\infty} \cdot e^{E_g/2kT} \quad (3)$$

In Eq. (3), ℓ and a are respectively the usual length and area of cross-section and σ_∞ is the conductivity at a very high temperature (ideally infinity), when all the valence electrons will have moved into the conduction band. Now, Eq. (3) can be rewritten in the form

$$R = R_\infty e^{B/T} \quad (4)$$

where R_∞ and B are constants associated with the thermistor made out of the semiconductor. R_∞ will be in ohms and

$B = (E_g/2k)$ has dimensions of temperature. With E_g around 0.6 eV and $k = 8.625 \times 10^{-5}$ eV/K, we would get a B of about 3500 K for a typical thermistor. R_∞ , which is theoretically the resistance at infinite temperature, can take values ranging from fractions of an ohm to several ohms, depending on the material and size of the thermistor. Equation (4) is normally written in the form

$$R = Ae^{B/T} \quad (5)$$

with A being equal to R_∞ .

In Eq. (5), let us consider two temperatures T_0 and T . If the thermistor resistances at these two temperatures are R_0 and R , we would get

$$\frac{R}{R_0} = \frac{Ae^{B/T}}{Ae^{B/T_0}} = e^{(B/T - B/T_0)}$$

This leads to a commonly used R - T relationship, namely,

$$R = R_0 e^{B(1/T - 1/T_0)} \quad (6)$$

R_0 here is usually the thermistor resistance at a room temperature of 298 K (25°C). R_0 and B for a few thermistors are given in Table 1 (6,7).

APPLICATIONS OF THERMISTORS

General

The features of a thermistor that favor its use as a temperature sensing element are:

- high sensitivity
- availability in small sizes (beads as small as 0.7 mm in diameter), which facilitates measurement at a point and with fast response.
- wide range of resistance values
- possibility of covering a large temperature range from 120 K to 470 K and higher.

Because of these advantages, thermistors are extensively used for a variety of applications involving nonelectrical and electrical variables. The most common applications are in the measurement and control of temperature in process industries. Thermal conductivity, fluid flow, and gas composition are some of the other nonelectrical quantities for the measurement of which thermistors can be employed (1,6,8,9). One of the earliest known uses of the thermistor has been in the measurement of power in RF (radio frequency) and microwave circuits and devices (10). A problem frequently faced in the design of high-performance electronic circuits is the sensitivity of their response to ambient temperature variations. The thermistor with its high negative temperature coefficient comes in handy to minimize or nullify the above-mentioned undesirable effect. Millivoltmeters, biasing circuits of bipolar transistors, and log-antilog amplifiers (11) are typical examples where this property is used. Thermistors also serve as vital components in feedback systems used for the automatic control of amplifier gain and stabilization of output amplitude in electronic circuits.

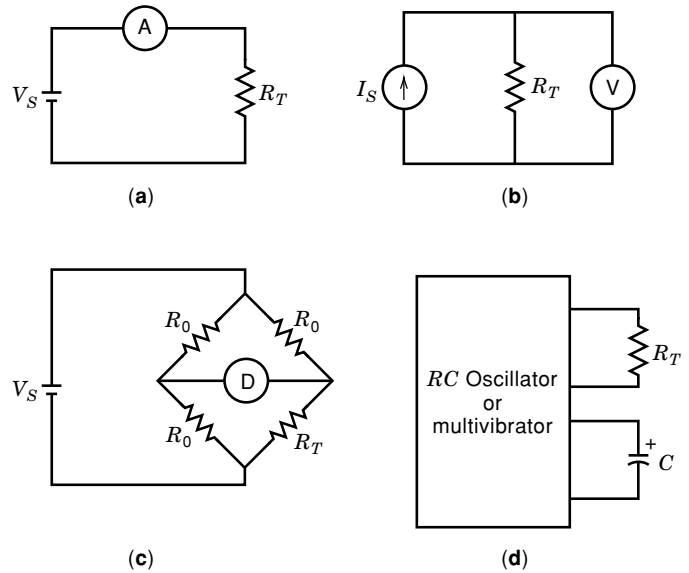


Figure 3. Thermistor converts temperature into an electrical output.

Methods of Deriving a Temperature-Dependent Output (5,12)

A thermistor is essentially a passive transducer and hence requires energization by an external power source for deriving a temperature-dependent electrical output. The various techniques employed for this purpose are illustrated in simple forms in Fig. 3(a) to (d). In Fig. 3(a), the thermistor is excited by a constant voltage V_S and the resulting current indicated by the ammeter is a function of temperature. In Fig. 3(b), a constant current I_S is passed through the thermistor across which a temperature-dependent voltage is developed. The thermistor is used as one of the arms of a Wheatstone bridge in the circuit of Fig. 3(c). The bridge is initially in a balanced condition at a temperature, say, T_0 . A change in temperature from this value causes an unbalanced voltage or current that can be detected by D and related to the change in temperature.

It is also possible to configure the thermistor in a circuit so that a temperature-dependent frequency or time-interval is obtained as an output. An output of this type is preferred, when temperature indication in digital form is desired or when the information is to be transmitted over long distances. An arrangement for realizing this is shown in the block-schematic of Fig. 3(d), where the thermistor (R_T) is used as a timing component of an oscillator or multivibrator.

Control of Temperature

It may often happen that the control of temperature in a process is also desired in addition to its measurement. In this case, a temperature-dependent signal is obtained using any one of the basic schemes described above. This signal, say X_T , is compared with a "set-point" (X_S) and the resulting error signal is used to control the heating or cooling of the process. The block diagram of Fig. 4 serves to explain the principle.

Measurement of Fluid Flow (1,8)

The use of the thermistor here is similar to that of the hot-wire anemometer with the thermistor taking the place of the Wollaston wire. The thermistor probe, energized electrically to attain

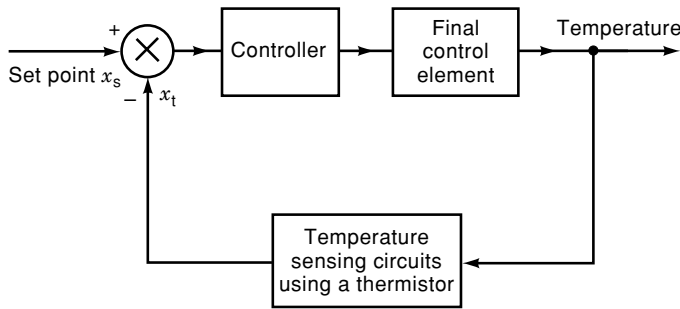


Figure 4. Thermistor in feedback path controls the temperature of a process.

a sufficiently high temperature, is placed in the fluid. Fluid flow causes a decrease in temperature of the probe, because of the heat transfer from the probe to the surrounding fluid. This decrease in temperature is accompanied by an increase in resistance of the thermistor. Additional heat input into the thermistor is needed to bring its resistance and temperature to the earlier value. To this end, the current through the thermistor is increased to a new level, which is then a measure of the fluid velocity. This mode of operation is called constant temperature or constant resistance mode. A bridge-configuration similar to Fig. 3(c) with the thermistor forming one of the arms is often preferred. The advantage of using a thermistor in place of a hot-wire lies in overall compactness, higher temperature sensitivity, and less risk of probe burn-out.

If, instead of a moving fluid, we have a static medium surrounding the thermistor, an arrangement similar to that already described can be used to measure any quantity of the medium which affects the heat transfer. In this manner, we can build thermistor-based instruments to measure variables such as thermal conductivity, gas composition, pressure of gas, and so on.

Measurement of Power in Radio Frequency Circuits

The earliest known application of a thermistor is in the measurement of RF and microwave power. The power-measuring instrument here is called *bolometer* (10), with the thermistor being known as the bolometer element. As is the case with most other applications, the thermistor is used as one arm of a simple bridge configuration shown in Fig. 5.

The thermistor suitably mounted in an RF/microwave system absorbs the high frequency power under measurement and is consequently heated. Simultaneously, it is energized by two other sources—a dc and a low-frequency ac [usually AF (audiofrequency)]. The dc current is adjustable and serves

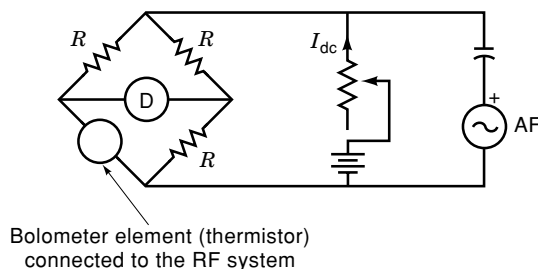


Figure 5. Thermistor helps RF power measurement through AF.

to set the thermistor resistance to a value that would be required for impedance matching on the RF side. The bridge is initially balanced with the thermistor receiving power from all the three sources, the AF power being minimal. The RF/microwave power is then turned off and without disturbing the dc, the AF power is increased to restore the bridge balance. Since the bridge is once again in balance, the thermistor resistance and hence the total input power to the bolometer element is the same under both conditions. The increase in AF power should therefore be equal to the unknown RF power. The accuracy of measurement will be enhanced if the initial AF voltage fed to the bolometer is as near zero as possible. A recent method employing two self-balancing thermistor bridges for RF power measurement is given in section 1.4.6 of Ref. 13.

Compensation and Stabilization

We are aware that the thermistor basically possesses a negative temperature coefficient of resistance. It can therefore be used to counteract the effect of the positive temperature coefficient of some other element on the response of a device or a circuit, when ambient temperature changes. Take for example a moving coil millivoltmeter, where the coil is of copper having a temperature coefficient of about +0.4%. By choosing a thermistor or a thermistor-resistor combination for the series multiplier, it is possible to make the millivoltmeter reading insensitive to ambient temperature changes.

Another application of a thermistor as a compensating element is in logarithmic amplifiers (11). In this circuit (Fig. 6), the A_1 - R_1 - D_1 combination constitutes the basic logarithmic amplifier. Its output voltage is

$$v_1 = -V_T \ln \frac{v_i}{I_0 R_1} \quad (7)$$

where I_0 is the reverse saturation current of D_1 and $V_T = kT/q$ is the voltage equivalent of thermal energy. I_0 and V_T in Eq. (7) make v_1 highly dependent on ambient temperature. If diode D_2 matches with D_1 , it would, along with the current source I_S , compensate for changes in I_0 with temperature. The voltage at the noninverting pin of A_2 is

$$v_2 = -V_T \ln \frac{v_i}{I_S R_1}$$

The output voltage of the complete logarithmic amplifier circuit becomes

$$\begin{aligned} v_o &= (1 + R_T/R_2) \cdot v_2 \\ &= -(1 + R_T/R_2) \cdot V_T \ln \frac{v_i}{I_S R_1} \end{aligned} \quad (8)$$

A suitable thermistor can be chosen for R_T so that changes in the values of R_T and V_T due to variations in ambient temperature will have equal and opposite effects on the output voltage v_o .

Amplitude Stabilization in Oscillators (11)

We often use a thermistor as a nonlinear feedback element in amplifiers and oscillators to achieve an output of constant amplitude. Let us consider the Wien-bridge oscillator of Fig. 7, built using an op-amp. Here negative feedback is provided by

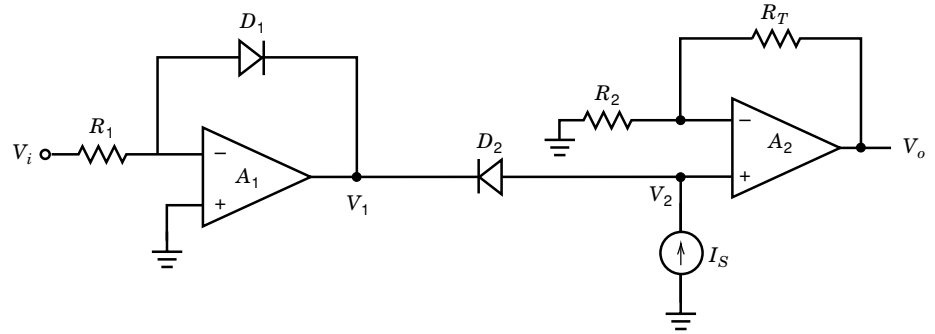


Figure 6. Thermistor R_T helps to reduce the temperature dependence of v_o .

the R_3 – R_4 potential divider connected across the output. R_4 is usually a thermistor. When there is a tendency for the amplitude of the output voltage v_o to increase, the current through the thermistor increases. This, in turn, causes a decrease in R_4 due to additional self-heating. The voltage across R_3 increases, resulting in more negative feedback, which tends to reduce the output and therefore maintain it at a constant value. We can also view the action of the thermistor here as an automatic control mechanism that alters the gain of an amplifier to achieve output amplitude stability. In principle, we can, with an appropriate arrangement, exploit the negative slope of the thermistor R – T characteristic to stabilize other electrical quantities as well.

LINEARIZATION OF THERMISTOR RESPONSE (14–19)

Whereas high sensitivity and fast response are the major plus points for a thermistor, the nonlinearity of its R – T characteristic stands as a stumbling block for its use in several applications in which we need an output varying linearly with temperature. The technique of achieving this is broadly known as “linearization.”

Shaping of R – T Characteristic

The simplest method for achieving linearization is to use a linear resistor in series or in parallel with the thermistor and obtain a terminal resistance that varies more linearly with temperature (14). Consider, for example, a typical circuit as in Fig. 8, where R_T is an Omega 44034 thermistor and R_p is a 2 k Ω resistor connected in parallel with it. The curves on the figure clearly show the effect of adding R_p to get an R – T characteristic with improved linearity over a temperature range of 30°–50°C. The sensitivity however falls considerably (from 107 Ω /K to nearly at 19.7 Ω /K, at 40°C). If we pass a constant current through the R_p – R_T combination, the resulting voltage would vary linearly with temperature. The temperature range over which this technique is applicable is quite narrow.

A Technique for Linear T -to- f Conversion

With thermistors that closely obey the two-constant law $R = A e^{B/T}$ over a specific temperature range, it is possible to obtain, using electronic circuits, an output in the form of a frequency which is proportional to temperature.

Refer to the schematic diagram of the T -to- f converter shown in Fig. 9. This circuit functions essentially as a relaxation oscillator (15). A reference voltage V_r energizes the network N containing the thermistor, and also a series R – C circuit. Let at $t = 0$, the output of the monostable multivibrator be low and the switch S open. The capacitor C starts to get charged

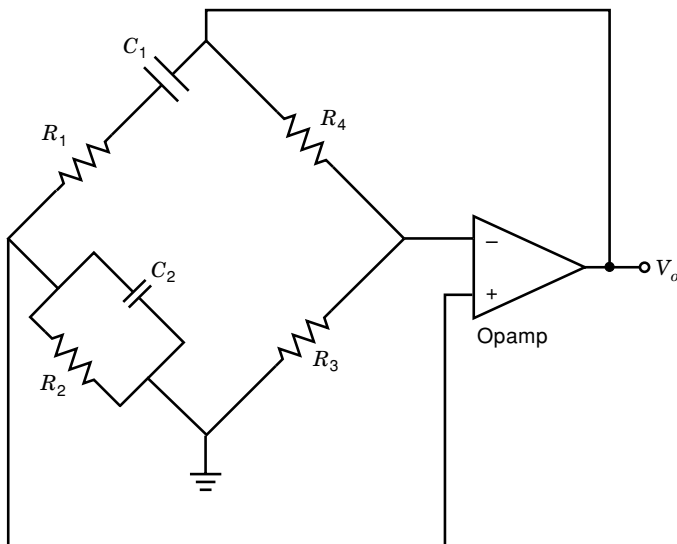


Figure 7. Thermistor (R_4) stabilizes oscillation amplitude.

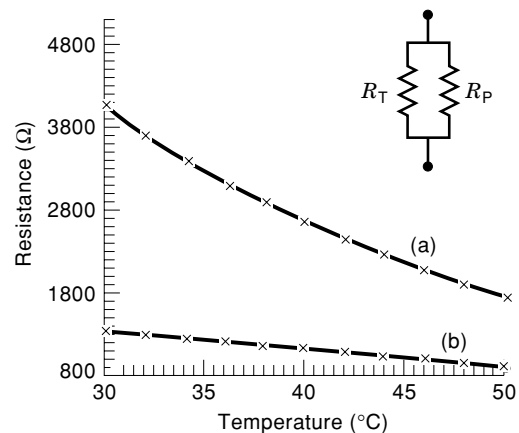
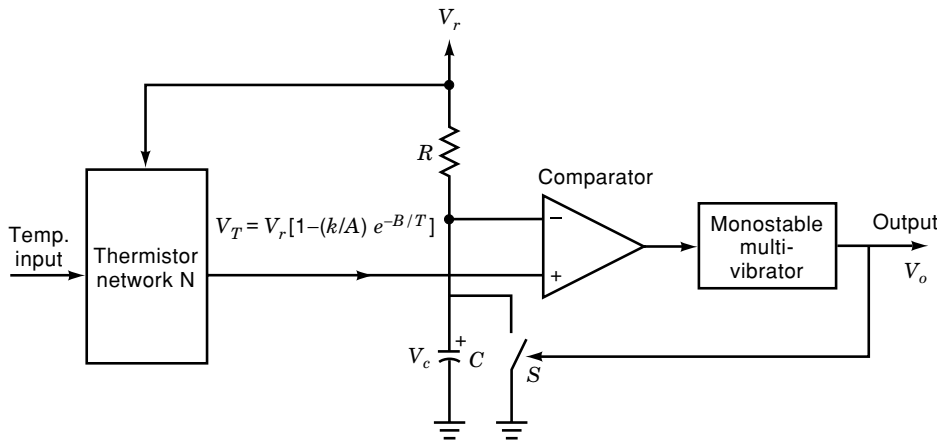


Figure 8. R – T characteristic (a) before and (b) after the connection of a 2 k Ω resistor in parallel with Omega 44034 thermistor.

Figure 9. Linear T -to- f converter.

and its voltage rises according to the relation

$$v_c = V_r(1 - e^{-t/RC}) \quad (9)$$

The network N provides a temperature-dependent voltage

$$v_T = V_r \left(1 - \frac{k}{A} e^{-B/T} \right) \quad (10)$$

where k depends on the parameter values of certain elements in N. Let v_c reach the level v_T at $t = t_0$. From Eqs. (9) and (10), we then obtain

$$t_0 = \frac{BRC}{T} - RC \ln \frac{k}{A} \quad (11)$$

At this instant, the comparator output is lowered, triggering the monoshot. The output of the monostable increases; the switch S gets closed and discharges the capacitor. The monostable remains in the high state for its period τ at the end of which it returns to low state, initiating another cycle of operation. The period of the relaxation oscillator, namely, $t_0 + \tau$ will be BRC/T , if τ is adjusted to be equal to $RC \ln(k/A)$. We thus get a frequency of oscillation

$$f = \frac{1}{t_0 + \tau} = \frac{T}{BRC}$$

which varies linearly with the absolute temperature T . The circuit achieves perfect linearization if the two-constant law is strictly valid for the thermistor used. Any nonlinearity in the output is due mainly to the departure of the thermistor behavior from the assumed law.

There are available in the literature a large number of linearizing circuits, producing an output in the form of a voltage, frequency, or time interval linearly related to temperature. In most of these circuits, linearization is achieved by expressing the output as a function of temperature in Taylor-series form and adjusting the circuit elements, to nullify the second-order term (16,17,18). The range of temperature over which these techniques will be useful is generally limited.

Wide-Range Linearization

The linearization methods discussed in the previous sections are basically hardware techniques, which would exhibit larger error as the temperature range is increased. As mentioned earlier, the major reason for the errors is the deviation of the R - T characteristic from the $Ae^{B/T}$ -law. This is because some of the assumptions such as invariance of carrier mobility with temperature, used in the derivation of that law, are not valid over a wide temperature range. One approach to reducing the linearity error would be to add suitable resistors in series and/or parallel to the physical thermistor, so as to obtain a terminal resistance which closely obeys the $Ae^{B/T}$ -law over a wide range (19). Other methods use empirical relationships involving three constants, to approximate the actual R - T characteristic of the thermistor better than the two-constant law. Under this category (1,7) we have:

$R = AT^{-c}e^{B/T}$, called the Becker-Green-Pearson (BGP) law

$R = Ae^{B/(T+\theta)}$, Bosson's law

$1/T = A + B \ln R + C (\ln R)^3$, Steinhart-Hart equation

For an Omega 44034 thermistor (7), the above three approximations would respectively give rise to a maximum fit error of 0.065 K, 0.037 K, and 0.018 K over the temperature range 0 to 100°C. This value would be -0.533 K if the two-constant law is used. It is true that the three-constant laws fit better than the conventional $Ae^{B/T}$ -law, but they do not lend themselves to easy hardware linearization. The availability of computing power in the form of microprocessors and personal computers has made software-based linearization possible using these equations. A straightforward method would be to obtain an analog voltage proportional to the thermistor resistance, convert it to digital form using an analog-to-digital converter (ADC) and compute the temperature using the chosen three-constant equation.

CERTAIN PARAMETERS OF IMPORTANCE (1,4)

Self-Heating Error

In the applications which involve temperature measurement, we have tacitly assumed that the temperature of the thermistor is the same as that of the surroundings. This assumption is valid provided the thermistor has been long enough at the

location of measurement and is not electrically energized. Passage of current through the thermistor for obtaining an electrical output is bound to heat the thermistor to a temperature higher than the surroundings. Since the measurement schemes senses only the thermistor temperature, an error arises which is termed a self-heating error. To keep it small, the thermistor current must be sufficiently low, taking into account the environmental conditions and the heat dissipation capability of the thermistor.

Dissipation Constant

To help the user estimate the error due to self-heating, manufacturers often specify for each thermistor a parameter called the dissipation constant. This constant is defined as the power required to raise the temperature of the thermistor by 1 K above its ambient, and is usually expressed in mW/K. This parameter depends on the nature of the thermistor environment and also on how the sensor is mounted. For example, a commercial tiny bead type thermistor could have a dissipation constant much less than 1 mW/K in still air. Its value for a disk type device mounted on a chassis may be as high as 60 mW/K (3,7).

Response Time

Response time of a thermistor is of importance when monitoring of rapid changes in temperature is required. Since the dynamic behavior of a thermistor is that of a first order system, the response time is specified by the term time-constant. It is the time taken by the thermistor to undergo a resistance change equal to $(1 - e^{-1})$ times the total final change in resistance that would be caused by a given step temperature change. The time-constant and the response time depend on the thermal mass, specific heat, and dissipation constant of the thermistor. Very small-size bead thermistors having time-constants less than 0.5 s are commercially available.

Limitations

Along with their numerous advantages, thermistors have a few limitations too. A rather serious one is the problem of interchangeability. Even thermistors fabricated by the same technique with strict control of the manufacturing process exhibit a spread in their R - T characteristics. Another shortcoming is lack of stability, since the resistance of the thermistor drifts with aging and continued use. Bead type thermistors are generally more stable than their disk counterparts. At high temperatures, changes in composition might occur in the device and the sensor will then have to be recalibrated for further use. This places an upper limit (about 590 K) on the operating temperature of the thermistor. A lower limit for its use is imposed by the largest resistance value beyond which measurement becomes difficult. This lower limit lies around 120 K.

POSISTORS (5,6,7)

The thermistor sensors discussed in the preceding sections are basically ones having a negative temperature coefficient of resistance. Strictly speaking, they should be called NTC thermistors, since there is another class of thermally sensitive resistors which exhibit a large positive temperature coefficient over a small temperature range. These devices are known as PTC

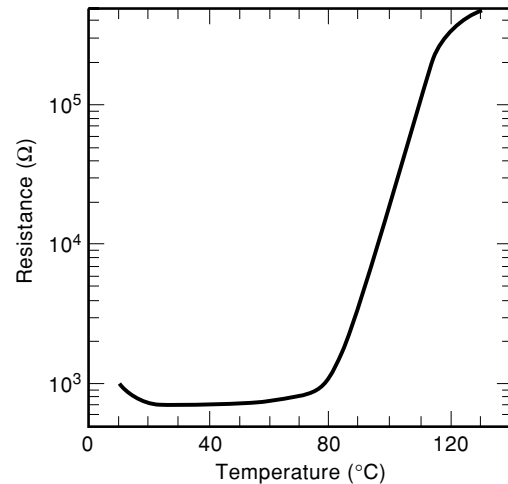


Figure 10. Typical R - T characteristic of a PTC thermistor.

thermistors or posistors. The R - T characteristic of a typical posistor is shown in Fig. 10.

It is seen that, as the temperature is increased, the device exhibits an NTC characteristic with a gradually decreasing resistance-temperature coefficient up to a certain temperature. A transition now occurs and the resistance, instead of decreasing, steeply increases with temperature. This phenomenon is noticed in devices made of certain ferroelectric materials such as barium titanate doped with strontium or lead. The PTC is traceable to the sudden decrease in the dielectric constant above the Curie point, which results in an increase in the activation energy (E_g) and an associated increase in the resistivity of the material. The steep R - T characteristic in the transition phase of the posistor makes it ideally suited for applications such as low-cost self-regulating heaters, electronic switches, and overcurrent protectors (6,9).

BIBLIOGRAPHY

1. R. W. A. Scarr and R. A. Settrington, Thermistors, their theory, manufacture and applications, *Proc. Inst. Elec. Eng.*, part B, **107**: 395-405, 1960.
2. T. G. Beckwith and N. L. Buck, *Mechanical Measurements*, Reading, MA: Addison-Wesley, 1969.
3. T. D. McGee, *Principles and Methods of Temperature Measurement*, New York: Wiley, 1988.
4. B. G. Liptak, *Instrument Engineers' Handbook—Process Measurement and Analysis*, 3rd ed., London: Butterworth-Heinemann, 1995.
5. H. B. Sachse, *Semiconducting Temperature Sensors and their Applications*, New York: Wiley, 1975.
6. *NTC and PTC Thermistors—Applications*, Munchen 80, Germany: Siemens AG, 1987.
7. *Temperature Measurement Handbook and Encyclopedia*, Stamford, CT: Omega Eng., Inc., 1985.
8. K. Okamoto et al., A digital anemometer, *IEEE Trans. Instrum. Meas.*, **IM-43**: 116-119, 1994.
9. Pamphlet, *NTC and PTC Thermistors*, Mansfield, OH: Therm-O-Disc, Inc., 1997.
10. F. E. Terman and J. M. Pettit, *Electronic Measurements*, New York: McGraw-Hill, 1952.
11. J. Millman and C. C. Halkias, *Integrated Electronics: Analog and Digital Circuits and Systems*, New York: McGraw-Hill, 1993.

12. E. O. Doebelin, *Measurement Systems Application and Design*, 3rd ed., New York: McGraw-Hill, 1983.
13. B. E. Nolingk (ed), *Instrumentation Reference Book*, 2nd ed., Oxford: Butterworth-Heinemann, 1995.
14. E. Keonjian and J. S. Schaffner, Shaping the characteristics of temperature sensitive elements, *IRE Trans. Compon. Parts*, **4**: 1954.
15. O. I. Mohamed, T. Takaoka, and K. Watanabe, A simple linear temperature-to-frequency converter, *Trans. IEICE*, **E70**: 775–778, 1987.
16. A. A. Khan and R. Sen Gupta, A linear temperature/voltage converter using thermistor in logarithmic network, *IEEE Trans. Instrum. Meas.*, **IM-33**: 2–4, 1984.
17. D. K. Stankovic, Simple thermistor temperature-to-frequency converter based on an astable multivibrator, *J. Phys. E: Sci. Instrum.*, **6**: 601–602, 1973.
18. S. Natarajan and B. Bhattacharya, Temperature-to-time converters, *IEEE Trans. Instrum. Meas.*, **IM-26**: 77–78, 1977.
19. S. Kaliyugavaradan, P. Sankaran, and V. G. K. Murti, A new compensation scheme for thermistors and its implementation for response linearization over a wide temperature range, *IEEE Trans. Instrum. Meas.*, **IM-42**: 952–956, 1993.

P. SANKARAN
Indian Institute of Technology,
Madras
S. KALIYUGAVARADAN
Madras Institute of Technology
V. G. K. MURTI
Indian Institute of Technology,
Madras

THERMOCOUPLES

Thermocouples are relatively inexpensive devices used to measure temperatures for a wide variety of applications, ranging from furnace control to equipment calibration. They can be made to fit just about any application. Unlike many temperature-measuring devices, the thermocouple is not subject to self-heating problems. A thermocouple is based on the finding of Seebeck, who showed that a small electric current will flow in a circuit composed of two dissimilar conductors when their junctions are kept at different temperatures (1). When one junction is hotter than the other, an electromotive force (emf) is produced that is proportional to the difference in temperature between the measuring junction (hot junction) and the reference junction (cold junction). This condition is known as the Seebeck emf, and the output is measured in millivolts. The pair of conductors that constitute the thermoelectric circuit is called a thermocouple. Thermocouples are the most widely used method of measuring internal temperatures in solid bodies. There are many different types of thermocouples that measure temperature over a range as low as -190°C and as high as 2000°C (2).

The measurement of temperature is thought to be a simple process, but this is a popular misconception. There is a need for controlled, reliable, and reproducible temperature-sensing devices for science, engineering, and industry. There are seven types of instruments used to measure temperature: thermocouple thermometers, radiation pyrometers, resistance thermometers, liquid-in-gas thermometers, filled-system thermometers,

optical pyrometers, and bimetal thermometers, and they all have advantages and disadvantages (3). The thermocouple is by far the most widely used device for temperature measurement because of its favorable characteristics that include good accuracy, coverage of a wide range of temperatures, fast thermal response, durability, high reliability, low cost, and versatility. This article concentrates only on the history, theory, junctions, calibration, and applications of thermocouples.

HISTORY OF THERMOCOUPLES

One person did not establish the principles or theory underlying thermoelectric effects. It was established by several scientists working over a span of many years beginning with Alessandro Volta, who concluded in 1800 that the electricity which caused Galvani's frog to twitch was due to a contact of two dissimilar metals (3). This conclusion was the forerunner of the principle of the thermocouple. Others who built on this base were Thomas Johann Seebeck (1821), Jean Charles Althanase Peltier (1834), and William Thomson—later Lord Kelvin (1848–1854). During this same period, Jean Baptiste Joseph Fourier published his basic heat-conduction equation (1821), Georg Simon Ohm discovered his equation for electric conduction (1826), James Prescott Joule found the principle of the first law of thermodynamics and the important I^2R heating effect (1840–1848), and Rudolf Julius Emanuel Clausius announced the principle of the second law of thermodynamics and introduced the concept of entropy (1850) (4,5).

Seebeck Effect

Thomas Johann Seebeck discovered the existence of thermoelectric currents while observing electromagnetic effects associated with bismuth–copper and bismuth–antimony circuits (4,5). His experiments showed that when the junctions of two dissimilar metals forming a closed circuit are exposed to different temperatures, a net thermal emf is generated which induces a continuous current. Measurements of the Seebeck effect can be made in terms of either the closed-circuit current or the open-circuit current. The Seebeck effect concerns the net conversion of thermal energy into electric energy with the appearance of an electric current. The Seebeck voltage refers to the net thermal electromotive force set up in a thermocouple under zero-current conditions. The direction and magnitude of the Seebeck voltage E_S , where E represents the thermoelectric emf, depends upon the temperature of the junctions and upon the materials making up the thermocouple. For a particular combination of materials A and B and for a small temperature difference dT , we obtain

$$dE_S = \alpha_{A,B} dT \quad (1)$$

where $\alpha_{A,B}$ is a coefficient of proportionality called the Seebeck coefficient and is also commonly called the thermoelectric power (1). The Seebeck coefficient is obtained in one of two ways:

1. As an algebraic sum ($\alpha_{A,B}$) of relative Seebeck coefficients (α_{AR}) and (α_{BR}), where for a given temperature difference and at given temperature levels, emfs of each of the substances, A and B, making up the thermocouple are

obtained with respect to an arbitrary reference material, R .

2. By numerically differentiating tabulated values of E_S versus T for a given reference temperature, T_R , according to the relation

$$E_S = \int_{T_R}^T \alpha_{A,B} dT \quad (2)$$

In either case, the Seebeck coefficient represents the net change in thermal emf caused by a unit temperature difference as in

$$\alpha_{A,B} = \lim_{\Delta T \rightarrow 0} \frac{\Delta E_S}{\Delta T} = \frac{dE_S}{dT} \quad (3)$$

If $E = aT + 0.5bT^2$ is determined by calibration, then $\alpha = a + bT$.

The Seebeck coefficient is a function of temperature level only based on the validity of the experimental relation

$$E_S = \int_{T_2}^T \alpha dT = \int_{T_1}^T \alpha dT - \int_{T_2}^{T_1} \alpha dT \quad (4)$$

where $T_1 < T_2 < T$. It follows that α is entirely independent of the reference temperature employed.

Peltier Effect

Jean Charles Althanase Peltier (1834) discovered interesting thermal effects when he introduced a small, external electric current in Seebeck's bismuth-antimony thermocouple (3,6). His experiments showed that when a small electric current is passed across the junction of two dissimilar metals in one direction, the junction is cooled and thus absorbs heat from its surroundings. When the direction of the current is reversed, the junction is heated as well as its surroundings. The Peltier effect takes place whether the current is introduced externally or is induced by the thermocouple. There are certain thermoelectric neutral points where no Peltier effect is apparent for special combinations of metals at certain temperatures.

The Thomson Effect

William Thomson—later Lord Kelvin—came to the conclusion that an electric current produces different thermal effects, depending upon the direction of its passage from hot to cold or from cold to hot, in the same metal (3,6). Thomson reasoned that if an electric current produces only the reversible Peltier heating effect, then the net Peltier voltage will equal the Seebeck voltage and will be linearly proportional to the temperature difference at the junctions of the thermocouple. Thomson also concluded that the net Peltier voltage is not the only source of emf in a thermocouple circuit but that a single conductor itself must also be a seat of emf.

The Seebeck, Peltier, and Thomson effects, together with several other phenomena, form the basis of functional thermoelectric modules.

THEORY OF THERMOCOUPLES

Electromotive Force (emf)

The emf is the energy per unit charge that is converted reversibly from chemical, mechanical, or other forms of energy into electrical energy in a conversion device such as a thermocouple. The basic principle of thermoelectric thermometry is that a thermocouple develops an emf, which is a function of the difference in temperature of its measuring (hot) junction and reference (cold) junction. If the reference junction temperature is known, the measuring junction's temperature can be measured by measuring the emf generated in the circuit. Therefore, we need an instrument capable of measuring emf.

There are three types of emf-measuring instruments in use in industry: deflection meters (millivoltmeters), digital voltmeters, and potentiometers. However, only two of them—digital voltmeters and potentiometers—are used where precision and accuracy are required for measuring thermal emfs. Digital voltmeters are high-impedance devices, and the readings are essentially independent of external circuit resistance. Potentiometers are used when the greatest accuracy is required in measuring emfs because its readings are free from uncertainties arising from changing circuit resistance (3).

Two metals, A and B, form an electric circuit with junctions that have temperatures t_1 and t_2 . In general, if the junction temperatures t_1 and t_2 are not identical, an emf will exist in such a circuit (7). The magnitude of the emf will depend on the metal used, the amount of temperature difference between t_1 and t_2 , and the actual temperature values of t_1 and t_2 . By including a suitable device to indicate any emf or flow of current that may occur in the circuit, the temperature difference $t_1 - t_2$ can be measured. The term thermoelectric power (e) as applied to such a circuit for a given pair of metals and a specified average temperature is defined as the ratio of the magnitude of the thermoelectric emf (E) to the temperature difference, $t_1 - t_2$ between the junction (see Fig. 1).

There are several basic laws that define thermoelectric circuits. These laws have been established experimentally and are generally accepted despite the lack of theoretical development.

Law of Homogeneous Metals. A thermoelectric current cannot be sustained in a circuit of a single homogeneous material, however varying in cross section, by the application of heat alone (4).

This law requires two different materials for any thermoelectric circuit (i.e., a thermocouple). No voltage V_{ij} can appear if wires A and B are chemically and physically the same, regardless of the values of T_1 and T_2 . This law provides that the position of the voltmeter (Fig. 2) does not affect the emf V_{ij} as long as both wires attached to the voltmeter are homogeneous. The voltmeter could be placed anywhere along wire A or B or at either junction.

Experiments have been reported suggesting that a nonsymmetrical temperature gradient in a homogeneous wire gives rise to a measurable thermoelectric effect. However, there is evidence that indicates that any emf observed in such a circuit arises from the effects of local inhomogeneities, and any current detected in such a circuit when the wire is heated is taken as evidence that the wire is inhomogeneous.

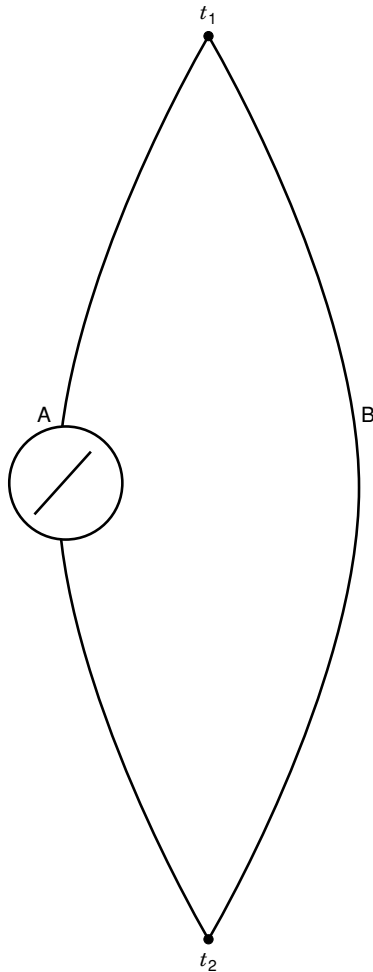


Figure 1. Thermoelectric circuit.

Law of Intermediate Metals. The algebraic sum of the thermoelectromotive forces in a circuit composed of any number of dissimilar materials is zero if all of the circuit is at a uniform temperature (7).

This law implies that a third homogeneous material can always be added in a circuit with no effect on the net emf of the circuit as long as its extremities are at the same temperature. A junction whose temperature is uniform and which makes good electric contact does not affect the emf of the thermoelectric circuit regardless of the method used in forming the junction. This is significant in that it allows for cheaper materials to be used as extension wires.

This law also implies that if the thermal emfs of any two metals with respect to a reference metal are known, then the emf of the combination of the two metals is the algebraic sum of their emfs against the reference metal.

Law of Intermediate Temperatures. If a given two-junction circuit produces an emf V_1 when junction temperatures are T_1 and T_2 and produces an emf V_2 when its junction temperatures are T_2 and T_3 , then the same circuit will produce an emf equal to $V_1 + V_2$ when its junction temperatures are T_1 and T_3 (1).

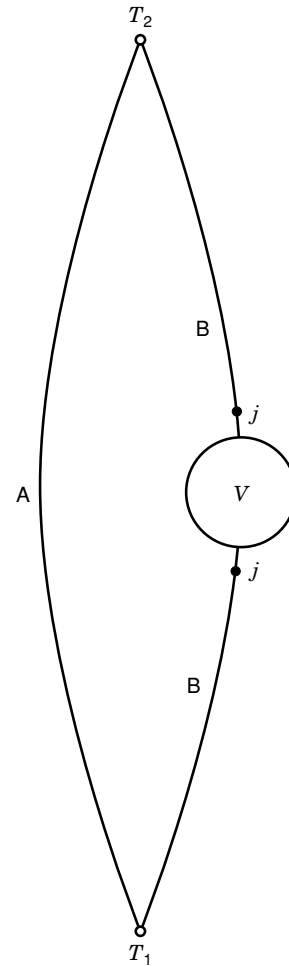


Figure 2. Thermocouple thermometer circuit. Dissimilar wires A and B are joined at temperatures T_1 and T_2 . The original current-flow circuit of Seebeck has been modified by the insertion of a high-impedance potentiometer V to emphasize the present-day thermometry technique.

This law is very important in case an engineer wanted to use a specific thermocouple circuit with a different reference junction temperature than the one for which a set of emf-temperature values is known. The temperature-emf relationship for a specific thermocouple combination is a definite physical property and thus does not depend on details of the apparatus or method used for determining this relationship. Figure 3 lists the emf curves for ISA standard thermocouples.

THERMOCOUPLE JUNCTIONS

There are numerous variations on the construction for joining the two dissimilar wires that make up a thermocouple. Thermocouples are available in four main junction types: exposed or bare-wire junction, grounded junction, ungrounded or isolated junction, and reduced diameter junction.

Exposed- or Bare-Wire Junction

In this type of junction, the sheath and insulating material are removed to expose the thermocouple wires (3). These wires are

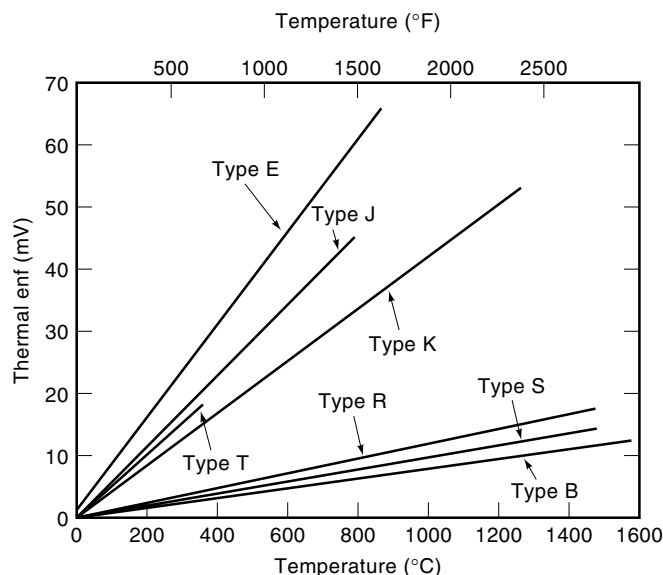


Figure 3. Thermal emf curves for ISA standard thermocouples. [Based on IPTS-68 (1974).]

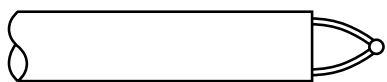


Figure 4. Exposed- or bare-wire junction

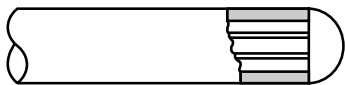


Figure 5. Grounded junction.

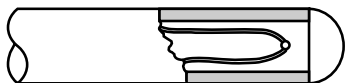


Figure 6. Ungrounded junction.

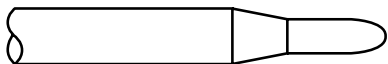


Figure 7. Reduced diameter junction.

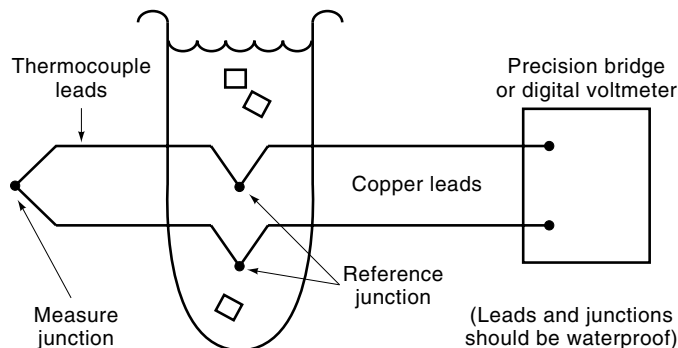


Figure 8. Thermocouple calibration system.

joined to form a measuring junction. While the thermocouple will have a fast response, the exposed ceramic is not pressure-tight, will pick up moisture, and will be subject to mechanical damage and expose the thermocouple to the environment (Fig. 4).

Grounded Junction

A closure is made by welding in an inert atmosphere so that two thermocouple wires become an integral part of the sheath weld closure (3). The wires are grounded to the sheath. This type of junction will give a slower response than an exposed wire, but the insulation is pressure-tight (Fig. 5).

Ungrounded or Isolated Junction

This type of junction is similar to the grounded junction except that the thermocouple wires are first made into a junction, which is then insulated from the sheath and the sheath enclosure (8). The closure is formed by welding without touching the thermocouple wires, and this results in an ungrounded thermocouple to the sheath material. This junction has a much slower response than the grounded junction but is still pressure-tight, protected from mechanical damage and the environment (Fig. 6).

Reduced Diameter Junction

This junction may be either grounded or insulated, and it is used where a fast response is required (3). It is more commonly used when a heavier sheath or wires are desired for strength, life, or lower resistance over the balance of the unit (Fig. 7).

CALIBRATION OF THERMOCOUPLES

Thermocouple wire is available commercially for measuring temperatures in the range of -190°C to 2000°C in matched pairs to conform to published standard tables. Each wire is calibrated separately, and then selected wires from two materials are paired such that the temperature-emf relationship for each pair does not deviate by more than the established standard tolerances. Common tolerances are usually $\pm 0.25\%$ to $\pm 0.75\%$.

The National Institute of Standards and Technology (NIST, formerly NBS) can provide temperature calibration when maximum authenticity is required for highly accurate temperature measurement applications. The temperatures covered by NIST are from -196°C to 1566°C , and a minimum length of 3 in. and a maximum of 24 in. wire is required.

The calibration of standard thermocouples consists of the determination of their emf values at a sufficient number of temperatures that they may be used to measure temperatures on a particular scale with a certain accuracy. This process may include annealing, test junction assembly, emf measurement, and construction of emf-temperature tables or equations. A diagram of a thermocouple calibration system is shown in Fig. 8 (9).

Annealing

Most base-metal thermocouples are annealed during manufacturing. Annealing is considered to be satisfactory for most thermometric purposes, so that the calibration process for base-metal thermocouples usually does not include an annealing step. For noble-metal thermocouples, annealing has been

demonstrated to be effective in promoting more uniform calibration results. NIST anneals all noble-metal thermocouples prior to calibration. The thermocouples are heated to about 1450°C in air by passage of electric current through their length while they are suspended loosely between supports. After approximately 45 min, they are annealed at 750°C for about 30 min and then cooled to room temperature.

Thermocouple Extension Wires

Thermocouple extension wires, also known as extension wires or lead wires, are electric conductors for connecting thermocouple wires to the temperature-measuring and temperature-control instrument. Extension wires are usually supplied in cable form, with positive to negative wires electrically insulated from each other. The chief reasons for using extension wires are economy and mechanical flexibility. Economy-base-metal thermoelements (\$10 per pound) are always used as extension wires for noble-metal thermocouple wires (\$700 per troy ounce). Mechanical-flexibility insulated solid or stranded wires in sizes from 14 to 20 gauge are used as extension wires. This lends mechanical sturdiness and flexibility to the thermocouple circuitry while permitting the use of larger-diameter base-metal thermocouples for improved oxidation resistance and service life, or smaller-diameter noble-metal thermocouple wire to save cost.

Test Junction Assembly

If a thermocouple is to be calibrated by comparison with a standard thermocouple, then the test wires are usually welded to the measuring junction of the reference thermocouple. By creating a single measuring-junction bead containing all of the thermocouples to be measured, an engineer can eliminate the temperature gradients between the pairs.

Cold Junction Compensation

As a differential output transducer, the voltage output of a thermocouple is dependent on the temperature at both the hot and cold junctions. The freezing point of water, 0°C (32°F), was selected as a convenient cold junction reference.

To construct an ice bath reference junction, both legs of the thermocouple are fused to copper wire to form a transition junction. The leads are then waterproofed and the transition junction is immersed in an ice bath (Fig. 9). The open-circuit voltage appears across the copper leads exiting from the ice bath.

The copper leads are used to connect to the emf readout device. This procedure avoids the generation of thermal emf at

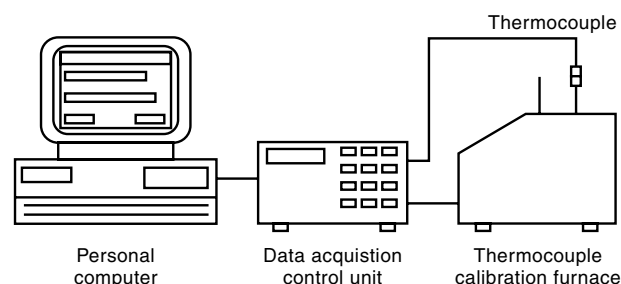


Figure 9. Ice bath circuit.

the terminals of the readout instrument. Voltages measured in this way may be directly converted into temperature by using NBS millivolt-temperature reference tables (7).

THERMOCOUPLE APPLICATIONS

The Instrument Society of America (ISA) assigned a letter designation to each of several types of thermocouples (Table 1). This allows the specification of an emf-temperature relation for each type of thermocouple without specifying its position. By specifying the emf-temperature relations by the letter designation rather than by the compositions, the ISA could ensure that manufacturers could deviate from other compositions that may be trademarked and still meet the published table values. The ISA thermocouples are accepted by NIST in the consensus temperature standard ANSI MC 96 and the useful ranges for thermocouple thermometers as established by the ASTM Committee E-20.

Representative samples of wire of each thermocouple type were studied extensively at NIST in order to develop reference tables of emf versus temperature over the useful range for each type of thermocouple. The reference tables are published along with Seebeck coefficient data in the NIST Monograph 125 issued in 1974.

Base Metals for Thermocouples

Base metals are the metal that is in greatest abundance for a given metal (10,11). However, when discussing base metals in terms of thermocouples, corrosion properties are what is most important for the material used. Base metals used for thermocouples that readily oxidize and are highly corrosive are iron and constantan. The corrosion characteristics of thermocouples are listed in Table 2.

Table 1. ISA Standard Thermocouples

ISA Designation	Approximate Composition (Positive Leg Listed First)	Useful Temperature Range (°C)
Base metal types		
Type E	(Ni + 10% Cr) vs. (Cu + 43% Ni)	-270 to 1000
Type T	Cu vs. (Cu + 43% Ni)	-270 to 400
Type J	Fe vs. (Cu + 43% Ni)	-270 to 1200
Type K	(Ni + 10% Cr) vs. (Ni + 2% Al + 2% Mn + 1% Si)	-270 to 1372
Noble metal types		
Type S	(Pt + 10% Rh) vs. Pt	-50 to 1767
Type R	(Pt + 13% Rh) vs. Pt	-50 to 1767
Type B	(Pt + 30% Rh) vs. (Pt + 6% Rh)	0 to 1820

Table 2. Corrosion Characteristics of Common Thermocouples (12–13)

Type of Thermocouple	Influence of Temperature and Gas Atmospheres
Type S, R, and B Platinum vs. platinum–rhodium	1. Resistance to oxidizing atmosphere: very good. 2. Resistance to reducing atmosphere: poor. 3. Platinum corrodes easily above 100°C. Should be used in gas-tight ceraminc protecting tube.
Type K Chromel-P vs. alumei	1. Resistance to oxidizing atmosphere: good to very good. 2. Resistance to reducing atmosphere: poor. 3. Affected by sulfur, reducing, or sulfurous gas, SO₂ and H₂S.
Type J Iron vs. constantan	1. Oxidizing and reducing atmospheres have little effect on accuracy. Best used in dry atmospheres. 2. Resistance to oxidation: good up to 400°C but poor above 700°C. 3. Resistance to reducing atmosphere: good (up to 400°C).
Type T Copper vs. constantan	1. Subject to oxidation and alteration above 400°C, due to copper; above 600°C, due to constantan wire. Contamination of copper affects calibration greatly. 2. Resistance to oxidizing atmosphere: good. 3. Resistance to reducing atmosphere: good. 4. Requires protection from acid fumes.

Noble Metals for Thermocouples

Noble metals are metals whose potential is highly positive relative to the hydrogen electrode; they have high resistance to chemical reaction, particularly to oxidation and to solution by inorganic acids (10,11). These metals are sometimes referred to as precious metals and are relatively scarce and valuable such as gold, silver, and platinum and are listed in Table 2.

Type K Thermocouples (Chromel-P Versus Alumel)

This type of thermocouple is regarded as the most versatile thermocouple because of its combination of high sensitivity, stability, oxidation resistance, and price. Type K thermocouples are recommended for use in an oxidizing or completely inert atmosphere over the temperature range of –200°C to 1260°C (–330° to 2300°F). Type K thermocouples should not be used in atmospheres that are reducing, alternately oxidizing and reducing, or vacuum for long periods of time since vaporization of chromium from the positive element may alter calibration used in thermometry applications in the –270° to 1372°C temperature range.

Type E Thermocouple (Ni–Cr Versus Cu–Ni)

Type E thermocouples have proven to be most useful in terms of lower wire conductivity and higher Seebeck coefficient. They are recommended for use over the temperature range of –200°C to 900°C (–330° to 1600°F) in oxidizing or inert atmospheres. These thermocouples are suitable for subzero temperature measurements since they are not subject to corrosion in atmospheres with high moisture contents. Type E thermocouples develop the highest emf per degree of all the commonly used types of thermocouples.

Type T Thermocouple (Copper Versus Constantan)

Type T thermocouples are resistant to corrosion in moist atmospheres and are suitable for subzero temperature measurements. They can be used in a vacuum and in oxidizing, reducing, or inert atmospheres over the temperature range of –200° to 370°C (–330° to 700°F). This type of thermocouple, along with Type E and Type K, are widely used in cryogenics, and its temperature range of use is between –270° and 400°C.

Type J Thermocouple (Iron Versus Constantan)

These thermocouples are suitable for use in vacuum and in oxidizing, reducing, or inert atmospheres over the temperature range of 0° to 760°C (32° to 1400°F). This thermocouple is not recommended for use below the ice point because rusting and embrittlement of the iron make it less desirable.

Type R and S Thermocouples (Platinum Versus Platinum–Rhodium)

Type R and S thermocouples are recommended for continuous use in oxidizing or inert atmospheres over the temperature range of 0° to 1480°C (32° to 2700°F). The continued use of these thermocouples at high temperatures causes excessive grain growth, which can result in mechanical failure of the platinum element.

Type B Thermocouple (Platinum–Rhodium Versus Platinum–Rhodium)

These thermocouples are recommended for continuous use in oxidizing or inert atmospheres over the temperature range of 870° to 1700°C (1000° to 3100°F). There are also suitable for short-term use in a vacuum. They should not be used in reducing atmospheres nor in those containing metallic or nonmetallic vapors.

Special Problems with the Use of Thermocouples

The thermocouple possesses an apparent simplicity that often deceives its users (1). The sensor appears to be a tiny detector that evaluates the temperature exactly at the location of the measuring junction. In certain commercially available thermocouple systems, the reference junction is contained within a digital voltmeter so that there is no messy ice bath to manipulate. There are several problems that are unique to thermocouples. Each problem arises from the nature of thermocouple measurements and the fact that the thermometric quantity is measured in terms of a small steady voltage. Any spurious source of voltage in the thermocouple circuit directly contributes to the temperature measurement error.

Some of the primary sources of thermocouple error areas follows:

1. Deviations from specifications in wire manufacture.
2. Use of low-impedance measuring instrumentation, leading to "loop-current" errors that arise from the flow of substantial currents within the thermocouple circuit.
3. Presence of electromagnetic interference, whether at the measuring junction or along improperly shielded extension wires.
4. Use of switching apparatus that introduced spurious and sometime variable voltages.
5. Use of extension wires that do not match the emf-temperature relation of the thermocouple wires themselves or that introduce unwanted emf's in their connections to the circuit.

SUMMARY

The use of thermocouples for temperature measurement provides a reliable and reproducible measurement for a variety of applications in many different disciplines. This article discussed the history of thermocouples, theory of thermocouples, thermocouple junctions, calibration of thermocouples, and thermocouple applications. Since the materials used to make thermocouples are relatively inexpensive and readily available, their use is widespread for scientific and industrial applications. From as low as -190°C to as high as 2000°C , thermocouples are used for just about every low-cost temperature measurement application.

BIBLIOGRAPHY

1. J. F. Schooley, *Thermometry*, New York: CRC Press, 1986, pp. 172–186.
2. H. F. Stimson, The international temperature scale of 1948, RP 1962, *J. Res. Natl. Bur. Standards*, **432**: 209–217, 1949.
3. R. J. Moffat, Thermocouple theory and practice, in *Fundamentals of Aerospace Instrumentation*, Vol. 6, Pittsburgh, PA: Instrument Society of America, 1974, pp. 111–124.
4. W. F. Roeser, Temperature, in *Thermoelectric Thermometry*, New York: Reinhold, 1941, pp. 180–205.
5. P. A. Kinzie, *Thermocouple Temperature Measurement*, New York: Wiley, 1973, Chap. 5.
6. The American Society for Testing Materials, Evolution of the International Practical Temperature Scale of 1968, STP 565, 1974.
7. H. Dean Baker, E. A. Ryder, and N. H. Baker, *Temperature Measurement in Engineering*, Vol. 1, New York: Wiley, 1953.
8. R. R. Ridgway, *Thermocouple*, U.S. Patent No. 21,521,553, Washington, D.C.: U.S. Government Printing Office, 1935.
9. Committee E-20 on Temperature Measurement, American Society for Testing Materials, *Manual on the Use of Thermocouples in Temperature Measurement*, ASTM Special Technical Publication 470B, 1981, Chaps. 2 and 5.
10. E. D. Zysk, Noble metals in thermometry, recent developments, Technical Bulletin, Englehard Industries, Vol. 5, No. 3, 1964.
11. H. E. Boyer and T. L. Gall, *Metals Handbook*, desk edition, Cleveland, OH: American Society for Metals, 1986, pp. 13-2, 13-3, 13-20–13-22.
12. R. L. Powek et al., Thermocouple reference tables based on IPTS-68, 1975, pp. 147–152.
13. A. Schulze, Metallic materials of thermocouples, *J. Inst. Fuel*, **12**: S41–S48, 1939.

Reading List

- M. D. Bethea and B. N. Rosenthal, An automated thermocouple calibration system, *IEEE Trans. Instrum. Meas.*, **41**: 702–706, 1992.
- S. Muth, Jr., Reference junctions, *Instrum. Control Syst.*, **May**: 133–134, 1947.
- T. J. Seebeck, *Evidence of the Thermal Current of the Combination Bi-Cu by Its Action on Magnetic Needle*, Berlin: Royal Academy of Science, 1822–1823, p. 265.

MARK McDOWELL
Microgravity Sciences Division

THERMOPILES

Thermopiles and thermocouples are transducers of thermal energy into electrical energy and vice versa. A thermocouple is the basic unit from which thermopiles are constructed and operates on the principles of thermoelectricity. In a closed electric circuit consisting of two dissimilar conductors joined together, electric current flows when the two junctions are maintained at different temperatures. The pair of conductors that make up the circuit are called thermoelements and the electric circuit is called a thermocouple. The open circuit voltage generated in this way is widely used for the accurate measurement of temperature and is schematically illustrated in Fig. 1. For a given pair of conductors, thermoelectric voltage is found to be a function of the temperature difference between the two junctions. In order to measure temperature using a thermocouple, one of the junctions, called the reference junction, is maintained at a fixed reference temperature. In this way, the thermoelectric voltage generated depends only on the temperature of the second junction which is known as the measurement junction. The temperature-voltage relationship for most commonly used thermocouple materials are available in the form of tables, curves, and mathematical equations. For most metals and alloys, thermoelectric voltages generated are very small—of the order of several tens of microvolts per degree Celsius.

If several thermocouples made of the same pair of conductors are connected in series with the alternate junctions at the hot and cold temperatures, respectively, the total output voltage increases by a factor, N , of the number of thermocouples in the circuit. Such a circuit consisting of several

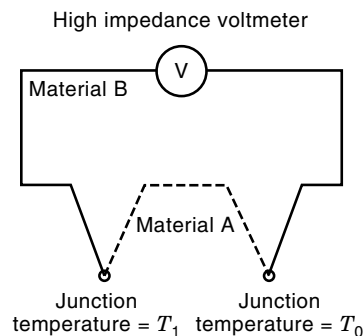


Figure 1. Schematic diagram of a thermocouple circuit consisting of two dissimilar materials A and B with the two junctions at different temperatures T_0 and T_1 . The open circuit thermoemf is measured using a high impedance voltmeter.

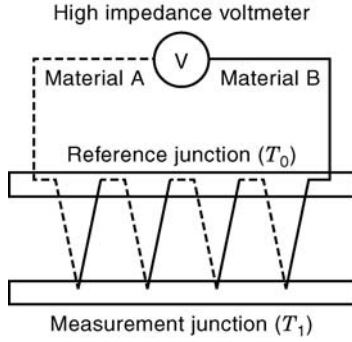


Figure 2. Schematic diagram of a thermopile consisting of four thermocouples connected in series. The alternate junctions of the thermopile are connected to the measurement junction at a temperature T_1 and the reference junction at a temperature T_0 .

thermocouples connected in series is called a thermopile and is shown in Fig. 2. This is a simple way of increasing the output voltage for a given temperature difference and is often used in many practical applications. Thermopiles find widespread use in various applications that include the accurate measurement of temperature, cooling and heating applications, generation of electricity, and infrared radiation detection. Thermoelectric thermometry is widely employed in scientific and industrial temperature measurement and control. When used in direct thermal contact with the object of interest, thermocouples can measure temperature over a large range from as low as 1 K (-272°C) to as high as 2000°C . This temperature range is further extended when thermopiles are used as the transducer elements in radiation pyrometry. Thermoelectric power generation and thermoelectric cooling are used to a much lesser extent but still dominate some niche applications such as power generation in spacecraft. Recent advances in silicon micromachining technology and thin-film deposition techniques are finding new applications for thermopiles as infrared detectors, accelerometers, and flow sensors.

We begin with a very brief review of thermoelectricity based on a standard thermocouple circuit as shown in Fig. 1. The bulk of this article will, however, concentrate on the various practical applications of thermopiles. A detailed account of thermoelectricity can be found in several excellent books such as Barnard (1) and Pollock (2).

BASICS OF THERMOELECTRICITY

Seebeck Effect

The Seebeck effect can be qualitatively understood by assuming a conductor as being made up of a rigid lattice of atoms in which a gas of free electrons moves under the influence of applied force fields. In the presence of a thermal gradient in such a conductor, electrons at the hot end diffuse toward the cold end. Scattering of electrons by the lattice causes a transfer of some of their energy to the lattice, resulting in the process of thermal conduction. The diffusion of high energy electrons also leads to the build-up of excessive electrons toward the cold end of the conductor and therefore to a build-up of an electric potential which is opposed to the thermal gradient. In this way, a dynamic equilibrium is established between the high energy electrons, being driven toward the cold end by the thermal gradient,

and the low energy electrons, being driven toward the hot end by the potential gradient. This electric potential arises whenever there exists a thermal gradient in a conductor and is known as the thermoelectric emf, or electromotive force. It is important to understand that thermoelectric electromotive force (emf) is not a junction potential but instead arises from the interaction of electrons with the lattice within a material in the presence of a thermal gradient.

In order to measure the Seebeck potential, electrical connections to the two ends of the specimen under test are made using a material different from the specimen. Otherwise the symmetry in the circuit will cause no net emf to be detected. The necessary use of two different materials results in the well known configuration of a thermocouple—in which two different conductors are joined together to form the hot and the cold junctions and the open circuit voltage of the couple is measured using a voltmeter. For homogeneous conductors, the thermoelectric emf depends only upon the temperatures of the two junctions and not upon the detailed shapes of the samples or the detailed forms of the temperature distributions along them. The rate of change of thermoelectric emf $V(T)$ with temperature T is known as the absolute thermopower $S(T)$ of the conductor, that is,

$$\frac{\Delta V(T)}{\Delta T} = S(T) \quad (1)$$

For a thermocouple as shown in Fig. 1, the open circuit voltage $V_{AB}(T_0, T_1)$ between the free ends of the thermoelements at temperature T_0 is given by

$$V_{AB}(T_0, T_1) = V_A(T_0, T_1) - V_B(T_0, T_1) = \int_{T_0}^{T_1} [S_A(T) - S_B(T)] dT \quad (2)$$

If material B is chosen to be a superconductor and is in its superconducting state (i.e., below the transition temperature), it will make no contribution to V_{AB} and the absolute thermopower of material A , S_A , can be experimentally measured.

In addition to the Seebeck effect, thermoelectricity has a manifestation in two other forms, known as the Peltier and Thomson effects.

Peltier Effect

If an electric current flows across a junction between two dissimilar materials there is a net evolution or absorption of heat in the junction region depending upon the direction of the flow of current. A Peltier coefficient Π_{AB} may be defined as the rate of absorption or evolution of heat per unit current flowing from material A to material B . If dQ_{AB}/dt is the net rate of evolution or absorption of heat at the junction at temperature T , then

$$\frac{dQ_{AB}(T)}{dt} = \Pi_{AB} I \quad (3)$$

where I is the current flowing from A to B . Π_{AB} is a function of the junction temperature of the two conductors. For metals, direct measurement of the Peltier coefficient is difficult to perform because of the small amount of heat evolution/absorption associated with the Peltier effect, as well as the difficulty in decoupling the effect from the associated Joule heating and

Thomson effect. Instead, it is usually determined using the experimental values of S_{AB} and the Kelvin relations.

Thomson Effect

The Thomson effect relates to the evolution or absorption of heat on the passage of an electric current through a single conductor in the presence of a thermal gradient. Thomson heat is proportional to the product of the current and the thermal gradient. The constant of proportionality is called the Thomson coefficient μ which is defined as the heat generated per second per unit current flow per unit temperature gradient when current flows through a conductor in the presence of a temperature gradient. It is a reversible heat, in the sense that the conductor changes from a Thomson heat absorber to a heat generator when either the direction of the current or the thermal gradient is reversed but not both at the same time. Using arguments based upon equilibrium thermodynamics, Thomson, later called Lord Kelvin, derived the relationship between the Seebeck coefficient, the Peltier coefficient, and the Thomson coefficient. These relations are known as the Kelvin relations and for a material at a temperature T can be written as:

$$\frac{\Pi}{T} = S \quad (4)$$

$$\frac{\mu}{T} = \frac{dS}{dT} \quad (5)$$

These relations allow any two of the thermoelectric effects to be quantitatively calculated if one of either μ or S can be determined over a given range of temperature.

The term “thermoelectric effects” as commonly used refers to all the reversible phenomena which occur at the junctions of dissimilar materials and throughout regions of a material in which finite thermal gradients are present. This is important because passage of current is simultaneously associated with the irreversible phenomena of the evolution of heat (“Joule heating”) and the conduction of heat. From a thermodynamic perspective and from Eq. (4), the Seebeck coefficient, S , can be interpreted as the mean transport entropy per unit charge, while Thomson coefficient μ , defined as amount of heat evolved or absorbed within a material per unit current flowing in the presence of a temperature gradient in the material, can be interpreted as the specific heat per unit charge. Thermodynamic perspective of thermoelectric effects is extremely insightful and these arguments were instrumental in the derivation of the Kelvin relations above. However, Lord Kelvin derived these relationships based on the questionable assumptions of reversible thermodynamics, which apparently gave correct results but had to await the developments of irreversible thermodynamics by Onsager in 1931 before these relationships were finally validated. In order to quantitatively predict the thermopower of a material, a detailed understanding of the electronic behavior of the material is necessary. Nevertheless, the various contributions to thermopower can be understood by remembering that the Seebeck coefficient is related to the electrochemical potential Θ' as:

$$S \propto \frac{1}{e} \frac{\Delta\Theta'}{\Delta T} \quad (6)$$

where e is the charge of the carrier, and T is the temperature.

The electrochemical potential,

$$\Theta' = \Theta + e\phi \quad (7)$$

includes both the chemical potential (commonly known as Fermi level) Θ , and the electrical potential, ϕ . In a homogeneous material, the chemical potential depends only on the temperature, whereas the quantity $e\Delta\phi/\Delta T$ depends on the detailed environment at any given point inside the material. The nature of this electrical environment is determined by the detailed way in which the electrons are scattered by the lattice vibrations, impurities, and other imperfections. In metals and alloys

$$\frac{\partial\Theta}{\partial T} \ll \left| e \frac{\Delta\phi}{\Delta T} \right| \quad (8)$$

and the thermopower is primarily determined by the scattering effects. On the other hand, in insulators, the opposite is true and the thermopower is dominated by chemical potential. The magnitude of the electronic thermopower for bulk metals and alloys at high temperatures (above Debye temperature ≈ 300 K for most metals and alloys), as given by the free-electron model, Barnard (1), is

$$S = \frac{-\pi^2 k^2 T}{3e} \left[\frac{1}{\Omega} \frac{\partial\Omega}{\partial E} + \frac{1}{l} \frac{\partial l}{\partial E} \right]_{E=E_F} \quad (9)$$

where k is the Boltzmann constant, e the electronic charge, Ω is the area of the Fermi surface, l is the electron mean-free-path, and E_F is the Fermi energy. Equation (9) implies that thermopower is very sensitive to changes in the electron scattering mechanisms, presence of impurities, strain, and even pressure. It must be emphasized that, although solid state physics gives a broad understanding of the phenomena of thermoelectricity, it has been very difficult to accurately predict the magnitude and occasionally even the sign of the thermopower of metals and alloys. Therefore, for most of the practical thermocouple systems, the dependence of thermoelectric emf on temperature is an empirical relationship.

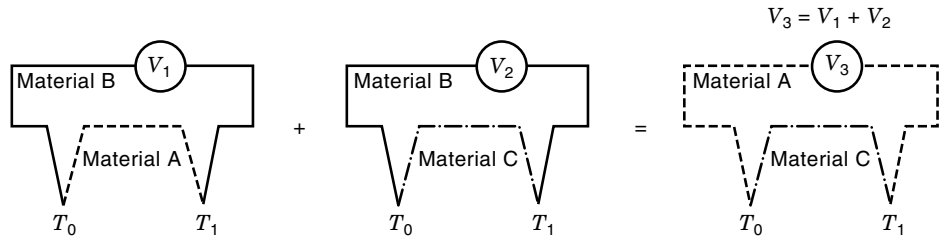
LAWS OF THERMOCOUPLE CIRCUITS

Based on the thermoelectric effects discussed and using the simple thermocouple circuit shown in Fig. 1, a number of empirical statements can be deduced. These laws are simple restatements of the detailed principles of thermoelectricity and form the basis for the practical construction and applications of thermocouple circuits.

Law of Homogeneous Materials

No thermoelectric voltage can be measured in a thermocouple circuit consisting of chemically and physically identical homogeneous materials no matter what the shape or cross-section of the conductor or the temperature difference between the two junctions. In reality though one must remember that, in the presence of a thermal gradient, thermoelectric voltages might be observed in single material circuits which are very likely to include inhomogeneities such as stressed or oxidized sections of the same material. Thermoelectrically speaking, these sections can be considered as different materials.

Figure 3. Schematic illustration of the law of different materials at different temperatures. This law is useful in generating the thermocouple calibration chart, the thermoelements of which have been calibrated against a common reference material.



Law of Different Materials at a Single Temperature

No thermoelectric voltage can be measured in a thermoelectric circuit consisting of many dissimilar materials and junctions if it is maintained throughout at a constant temperature. This law simply affirms the fact that in the absence of thermal gradients no thermoelectric phenomena can be observed.

Law of Different Materials at Different Temperatures

For the same junction temperatures, the algebraic sum of the thermoelectric voltages in two circuits composed respectively of materials A–B and B–C is the same as the thermoelectric emf of a single circuit composed of materials A–C. This law is schematically shown in Fig. 3. The main application of this law is in the generation of the temperature–thermoemf relationship for a given thermocouple combination if the temperature–thermoemf relationship of the constituent materials are available with respect to a common reference material such as platinum.

Law of Intermediate Temperatures

For a given two junction thermocouple, if the thermoelectric emf measured is V_1 when the two junctions are at temperatures T_0 and T_1 and V_2 when the two junctions are at temperatures T_1 and T_2 , then the thermoelectric emf V_3 produced by the same thermocouple when its junctions are at temperatures T_0 and T_2 is equal to $V_1 + V_2$. This law is particularly important if a thermocouple is intended for use at a different reference junction temperature than that used in its original calibration. This law is schematically illustrated in Fig. 4.

THERMOPILES AND TEMPERATURE SENSORS

As we have already discussed, a thermocouple essentially transforms thermal energy into electrical energy and vice versa. As temperature sensors, thermocouples have the advantages of not requiring any power supply for their operation and can be used over a very large temperature range. However, they have the disadvantages of small output voltages, susceptibility to noise pickup by the thermocouple wires,

and relatively high drift. For homogeneous, macroscopic metals and alloys at high temperatures, the laws of thermoelectricity are essentially independent of the physical dimensions of the thermocouple. This means that most thermocouples are constructed from wires as small as is practical or most cost effective. However, small thermocouples are more readily suited for spot measurements of temperature and therefore provide a small physical cross-section to interact with the sensing stimulus.

A thermopile consisting of several thermocouples connected in series not only offers a simple way of increasing the sensing cross-section of the thermocouple but as discussed earlier also offers the possibility of obtaining a larger electrical signal for a given temperature difference. As transducers of thermal energy into electrical energy the two main, Seebeck effect based, applications of thermopiles are as temperature and radiant flux sensors and as generators of electrical power. As transducers of electrical energy into thermal energy using Peltier effect, the main application of thermopiles is as cooling devices or refrigerators. The use of several thermocouples connected in series results in higher sensitivity in sensor applications and higher power in electricity generation applications. In cooling and heating applications, the many thermocouple junctions of thermopiles not only result in more power but can also be set-up into practical configurations for uniform heating and cooling of large surface areas.

The arrangement of connecting several thermocouples in series such that the alternate junctions are at two different temperatures amplifies the thermoelectric voltage of a single thermocouple by the number of thermocouples, N , connected in series.

$$V_{\text{Thermopile}} = N \cdot V_{\text{Thermocouple}} \quad (10)$$

However, this also results in an increase in the noise due to the N -fold increase in the resistance of the device. In an open circuit voltage measurement configuration, the primary source of internal noise in a thermocouple is from the random fluctuation of the charge carriers in the material and is called the Johnson noise. The root mean square (rms) open circuit Johnson noise voltage, V_n , in a 1 Hz bandwidth for a material of resistance R

Figure 4. Schematic illustration of the law of intermediate temperatures. This law is mainly used for operating thermocouples at reference temperatures which are different from the standard reference temperature.

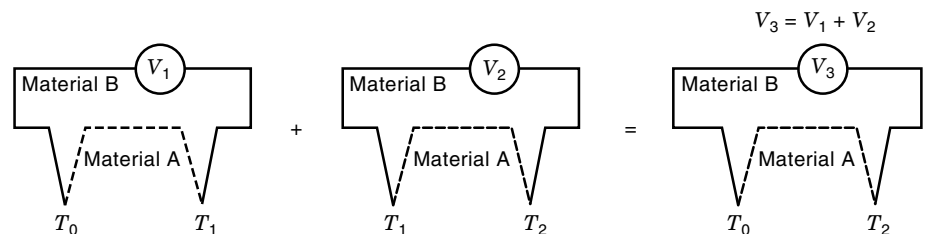


Table 1. Seven Standard Types of Thermocouples Used in the Measurement of Temperature

Thermocouple Type	Nominal Composition (% weight)	Useful Temperature Range (°C)	Comments
Type B	(Pt + 30% Rh) vs. (Pt + 6% Rh)	0–1704	Recommended for continuous use in oxidizing and inert atmospheres. Limited vacuum use is possible.
Type E	(Ni + 10% Cr) vs. (Cu + 43% Ni)	–250–871	Primarily for oxidizing atmospheres. Does not corrode at sub-zero temperatures
Type J	Fe vs. (Cu + 43% Ni)	–210–1200	Suitable for vacuum use or where free oxygen is deficient since iron rapidly oxidizes above 538°C (1000°F).
Type K	(Ni + 10% Cr) vs. (Ni + 2% Al)	–250–1260	Suitable for continuous use above 538°C in oxidizing atmospheres.
Type R	(Pt + 13% Rh) vs. Pt	–50–1482	Not suitable for reducing atmospheres and for continuous high temperature applications.
Type S	(Pt + 10% Rh) vs. Pt	–50–1482	Not suitable for reducing atmospheres and for continuous high temperature applications.
Type T	Cu vs. (Cu + 43% Ni)	–250–400	High resistance to corrosion from atmospheric moisture or moisture condensation. Can be used in either vacuum, oxidizing, or reducing atmosphere.

which is at a temperature T is given by

$$V_n = \sqrt{4kRT} \quad (11)$$

where k is the Boltzmann constant. It can be seen that, while the signal from a thermopile increases by a factor of N , the open circuit noise voltage in a thermopile increases by a factor of $\sqrt{N}$, therefore resulting in an effective signal to noise ratio increase of $\sqrt{N}$.

Choice of Materials

Based on the material of construction, thermopiles can be broadly classified into two major groups as: (1) metal-alloy thermopiles and (2) semiconductor thermopiles. Metal-alloy thermopiles are mainly used in temperature measurement applications whereas semiconductor thermoelements find applications in power generation, refrigeration, and radiant flux measurements. This is due to the fact that although metals and alloys have a small Seebeck coefficient, typically of the order of a few tens of microvolts per kelvin, they can be very reliably and inexpensively reproduced. Semiconductors on the other hand exhibit superior properties for thermoelectric energy conversion applications with larger Seebeck coefficients, typically of the order of hundreds of microvolts per kelvin. Semiconductor thermoelements are more expensive and can be less easily formed into various convenient forms for temperature measurement applications.

Most commercially available thermocouples and thermopiles used for temperature measurement are made up of metals and alloys. As stated above metals and alloys exhibit very reproducible thermoelectric properties and can be very easily and inexpensively formed into convenient forms such as thin wires suitable for temperature measurements. Additionally, modern potentiometers can measure very low voltages very accurately and reliably. These developments have made it possible to routinely measure temperatures accurately to a fraction of a degree Celsius using standard metallic thermocouples. Although it is possible to measure temperature using any combinations of metals and alloys, only seven combinations of different alloys are commonly used for temperature measurement in the range of 20 K to 2000 K. For each of the seven

thermocouple combinations there exist internationally agreed reference tables for thermal emf versus temperature as well as a letter designation. Table 1 lists the type designation of the alloy compositions, their useful temperature range, and the typical application environment for each of the thermocouple types. For a more detailed discussion on the use of thermocouples for temperature measurement refer to the American Society for Testing and Materials (ASTM) manual (3).

Thermoelectric Figure of Merit

The efficiency of energy conversion of thermoelectric devices is not only related to the absolute thermopower of the materials but also to their thermal and electrical conductivity. This can be qualitatively understood, for example, by considering a thermocouple as an electricity generator. In order to achieve high energy conversion efficiencies, the conductive heat losses from the transducing junction through the thermocouple legs need to be minimized (i.e., low thermal conductivity materials are required) and the electrical power output Z , for a given material is given by (V^2R) needs to be maximized (i.e., large thermopower and low electrical resistivity materials are required). In fact, thermoelectric figure of merit, Z , for a given material is given by

$$Z = \frac{S^2\sigma}{\kappa} \quad (12)$$

where S is the absolute thermopower of the material, and σ and κ are its electrical conductivity and thermal conductivity, respectively. The thermoelectric figure of merit has the dimensions of inverse of temperature. A more detailed derivation of the thermoelectric figure of merit is based on the maximization of the coefficient of performance of a thermoelectric power generator/refrigerator or the normalized detectivity D of an infrared detector. For a more comprehensive treatment of the figure of merit of thermoelectric materials, see references by Rowe and Bhandari (4), Kaye and Welsh (5), or Baltes et al. (6).

Materials with a large value of Z over a large temperature range are required in order to achieve a high energy conversion efficiency. For a thermocouple, the figure of merit can be extended to take into consideration the different

Table 2. Thermal and Electrical Properties, Including the Seebeck Coefficient and the Thermoelectric Figure of Merit, of Some Metals, Alloys and Semiconductors

Material	$\sigma(\Omega^{-1} \text{ m}^{-1})$	$S(\mu\text{V/K})$	$\kappa(\text{Wm}^{-1}\text{K}^{-1})$	$Z(\text{K}^{-1})$
<i>Positive Thermoelements</i>				
Cu	6.0×10^7	+1.83	401	5.01×10^{-7}
Au	4.2×10^7	+1.94	318	4.97×10^{-7}
Ag	6.3×10^7	+1.50	429	3.35×10^{-7}
Sb	2.4×10^6	+40.0	8.00	4.80×10^{-4}
Sb_2Te_3	2.0×10^5	+130	2.82	1.20×10^{-3}
* $\text{Bi}_{0.5}\text{Sb}_{1.5}\text{Te}_3$	5.9×10^4	+230	1.05	2.97×10^{-3}
*Polysilicon ($p = 3 \times 10^{19}$)	1.3×10^4	+135	31.0	7.97×10^{-6}
* $\text{Si}_{0.7}\text{Ge}_{0.3}$ ($p = 10^{19}$)	8.8×10^4	+121	5.50	2.33×10^{-4}
<i>Negative Thermoelements</i>				
Al	3.6×10^7	-1.70	236	4.41×10^{-7}
Ni	1.3×10^7	-19.0	91.0	5.16×10^{-5}
Bi	8.3×10^5	-60.0	22.0	1.36×10^{-4}
* Bi_2Te_3	1.2×10^5	-210	2.34	2.30×10^{-3}
$\text{Bi}_{0.87}\text{Sb}_{0.13}$	1.4×10^5	-100	3.10	4.55×10^{-4}
Si ($n = 10^{19}$)	2.0×10^4	-200	145	5.50×10^{-6}
*Polysilicon ($n = 3 \times 10^{19}$)	1.1×10^5	-121	29.4	5.58×10^{-5}
*GaAs ($n = 10^{17}$)	6.7×10^3	-390	39.0	2.74×10^{-5}

* Refers to values for thin film materials.

material properties of the two materials and is given by

$$Z_{AB} = \frac{S_{AB}^2}{\left(\sqrt{\frac{\kappa_A}{\sigma_A}} + \sqrt{\frac{\kappa_B}{\sigma_B}} \right)^2} \quad (13)$$

where the subscripts *A* and *B* refer to the properties of the two materials constituting the thermocouple respectively. Table 2 lists the thermopower and the figure of merit for some of the commonly used bulk and thin film thermocouple materials respectively.

Standard Construction

The main application of thermocouples and thermopiles is as temperature sensors. As has been discussed earlier, in their most popular use a thermopile is constructed of several thermocouples connected in series. Thermopiles are used for the direct measurement of temperature by placing the measurement junction in intimate thermal contact with the object of interest and the reference junction at a fixed known temperature. The detailed construction of a thermopile depends on the specific application for which it is being used. For most temperature measurement applications, the thermopile configuration that is most practically implemented is a modified form of the thermopile shown in Fig. 2. The measurement junction of the thermopile made of metals A and B, is located at the temperature, T_1 , to be measured. The input and output wires of the reference junction are connected to an extension wire C and is maintained at a temperature T_0 . The extension wires C finally connect the thermopile to a high impedance voltmeter at room temperature T_R . This configuration occurs invariably in all situations where the temperature to be measured is located remotely as compared to the voltmeter. Summing the various thermoelectric voltages for the circuit we get

$$V = V_C(T_0, T_R) + N\{V_A(T_0, T_1) - V_B(T_0, T_1)\} - V_C(T_0, T_R) \quad (14)$$

If the two wires made of material C are homogeneous and have

the same thermoelectric properties, then the total thermoelectric emf measured is the same as obtained using the thermopile made only of materials A–B. Developments in electronic circuits have led to accurate tracking of the variations in reference junction temperature using thermistors. This is often used to compensate for any errors in the measured temperature caused due to drifts in the temperature of the reference junction.

Thermocouples can also be connected in parallel as shown in Fig. 5. In such a case the parallel thermopile configuration provides a temperature averaging effect. However, extreme care must be taken to prevent the possibility of closed loop currents within the thermocouple pairs which can arise easily due to the variations in the calibration curves of the individual thermocouples, differences in the individual resistance of the thermocouples, and variations in the Seebeck coefficient of the individual thermoelements in the temperature range of measurements. Therefore, in spite of the possibility of temperature averaging, the parallel arrangement of thermocouples is not commonly used.

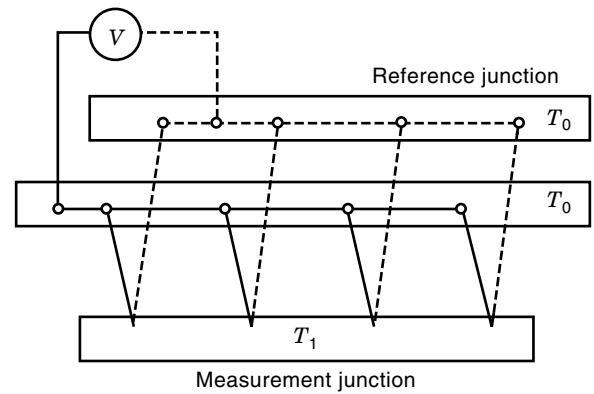


Figure 5. Schematic drawing of a thermopile consisting of four thermocouples connected in electrical parallel configuration. Used with a certain amount of caution, this configuration can be used for temperature averaging.

Reference Junction

As we have already discussed, a thermopile essentially measures the difference in temperature between the two junctions across which it spans. To measure temperature using a thermopile; one of its junctions is maintained at a fixed reference temperature and the temperature of the measurement junction is deduced using the calibration chart of the thermocouple constituting the thermopile. Reference junction serves two functions, namely, (1) as a junction at a standard reference temperature of 0°C for producing the thermocouple voltage-temperature calibration chart or (2) as a junction at a known temperature, other than the reference temperature of 0°C, which is either fixed or variable. The standard reference temperature used in thermocouple calibration is the ice point, or 0°C—the equilibrium temperature between ice and water at 1 atmosphere pressure. Historically, this reference temperature has evolved from the ease of availability of ice and water in the laboratory and its uniqueness in the phase diagram. For a detailed treatment of reference junctions for thermocouple calibration refer to Quinn (7).

The fixed reference temperature of 0°C is good for calibrating thermocouples and thermopiles but is very cumbersome and most often impractical to implement in typical industrial situations. In order to maintain the reference junctions at a constant temperature of 0°C, Peltier coolers, discussed later, are sometimes used. However, it is very difficult to maintain the temperature of the actual cold junctions at this temperature due to the variability in thermal contact and the difficulty of surrounding the reference junction sufficiently to ensure the temperature to the desired accuracy. Therefore, it is desirable to use temperatures other than 0°C for the reference junction. In its most widespread use, the reference junction is normally kept at ambient conditions and its temperature is continuously monitored using a thermistor. In this way, any variations in the temperature of the reference junction are known and electronically compensated.

In using thermopiles to measure temperature, in addition to the errors associated with reference junction, and deviations from standard thermoemf due to thermocouple materials and extension wires, users should also be aware of the possible error sources associated with measurement junctions. These errors can be especially significant when thermopiles are used in the measurement of surface temperatures, since their very presence perturbs the heat transfer characteristics of the surface and changes the temperature distribution. The thermopile wires provide additional heat transfer paths and can effectively lower the actual temperature of the measurement surface. Another major source of error arises from the thermal contact resistance between the surface and the thermopile junction. This thermal contact resistance causes a temperature gradient to be set up between the surface and the thermopile junction and therefore prevents the junction from attaining the surface temperature. The relationship between the measured and true temperature can be expressed as:

$$\zeta = \frac{T_m - T_t}{T_m - T_a} \quad (15)$$

where ζ is called the installation factor, T_m is the measured temperature, T_t is true surface temperature, and T_a is the temperature of the ambient. In addition, errors can also arise in

surfaces where temperature gradients exist due to the errors associated with the exact positions of the thermopile junctions relative to the surface. The response time of the thermopiles can also cause errors in the measurement of transient temperature signals.

OTHER APPLICATIONS OF THERMOPILES

Thermopiles are extensively used as the basic transduction elements in many thermal sensors (8). Thermal sensors operate by first converting the input signals (the measurand) into a temperature signal in the sensor element which is subsequently transduced into an electrical signal. The most well-known application of thermopiles as thermal sensors has been for detecting infrared radiation. When used as radiation sensors, the measurement junction of the thermopile is placed in intimate contact with a radiation absorbing structure while the reference junction is attached to a heat sink. Absorption of incident radiation causes a preferential rise in the temperature of the measurement junction and therefore in a thermoemf. Recent advances in silicon micromachining technology have made possible several novel thermopile-based thermal sensors which include accelerometers, calorimeters, and mass flow sensors. A comprehensive review of silicon micromachined sensors is presented by Middelhoek (9), and for an excellent review of thermoelectric microsensors and systems refer to Baltes et al. (6). In addition, thermopiles are also used in heating and cooling applications and in some niche power generation applications. These applications of thermopiles are discussed in detail in the following sections.

Radiation Detectors

The most commonly known physical manifestation of energy is as electromagnetic radiation. Electromagnetic radiation spans a very large spectrum from gamma rays at the high energy end to radio waves at the low energy end. Lying between the visible and microwave parts of the electromagnetic spectrum, infrared radiation is widely used for: noncontact temperature measurements, intrusion alarm systems, remote sensing, astronomy, and heat emission based target detection and tracking. These applications are all possible because heated objects provide an excellent source for infrared radiation which is emitted by virtue of their temperature. The quantity of importance, most often, is the radiant power being emitted by a blackbody or monochromatic source. Excellent reviews on infrared radiation and infrared detectors are available in Smith, Jones, and Chasmar (10) and Keyes (11).

The sensing element of a thermopile infrared detector consists of a very small heat capacity radiation absorbing structure which is thermally isolated from the main sensor body which acts as a heat sink. The measurement junction of the thermopile is thermally attached to the sensing element while, the reference junction is attached to the main body of the sensor. Upon absorption of the incident radiant flux, the sensing element heats up relative to the main body of the sensor and the thermopile generates an open circuit voltage which is calibrated to read the power density. Traditionally, thermopiles were fabricated as bulk material devices by welding together fine wires and thin black radiation absorbers. However, developments in micromachining technology have made it possible to fabricate thermally isolated structures with small heat

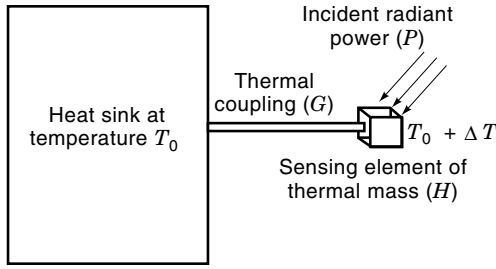


Figure 6. Schematic representation of a sensing element of thermal mass (H), connected to the heat sink at temperature (T_0) via a thermal conductance (G). Absorption of the incident radiation of power (P) causes the temperature of the sensing element to rise to $T + \Delta T$.

capacity while advances in microfabrication techniques have made it possible to fabricate very high density thermopile structures using thin films. These developments have led to the design and construction of a variety of thin-film thermopile radiation detectors covering a wide range of impedance, sensitivity, and time constants (12,13,14,15).

A simple model of a thin-film thermopile is shown in Fig. 6. The detector element of thermal mass H is coupled to the heat sink at a constant temperature T_0 via a thermal conductance G . When a radiant power P is incident on the sensor element, its temperature T is found by solving the equation:

$$\epsilon P = H \frac{\partial T}{\partial t} + G(T - T_0) \quad (16)$$

where ϵ is the emissivity of the sensor element and t is time. For a sinusoidal power input, $P = P_0 \exp(i\omega t)$, the steady state solution of Eq. (16) is

$$\Delta T = T - T_0 = \frac{AP_0}{\sqrt{G^2 + \omega^2 H^2}} \quad (17)$$

A characteristic thermal time constant, τ , for the detector can be defined as

$$\tau = H/G \quad (18)$$

From Eq. (17) it can be seen that a high sensitivity, which corresponds to a large ΔT , can be achieved by making G as small as possible and $\omega H \ll G$. In other words, both the heat capacity of the detector element and its thermal coupling to the surroundings should be as small as possible. A small heat capacity for the detector element implies a detector element of as small and light weight a construction as is practically possible whereas the smallest value for G is achieved when the thermal coupling of the sensing element to the heat sink is only through radiative exchange. The design challenge of a thermopile radiation detector, therefore, is the optimization of the interaction of the sensing element (measurement junction) with incident radiation while simultaneously reducing its thermal mass and all other thermal contacts to the surroundings.

Silicon micromachining technology, by the use of precise lithographic patterning and crystallographic etching techniques, has made it possible to fabricate free-standing membranes and structures with very low heat capacity and good thermal isolation (16,17). The measurement junction of such

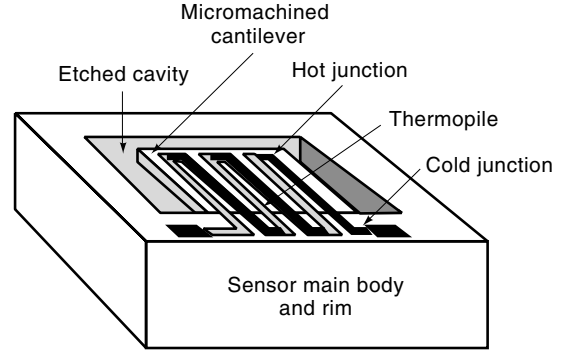


Figure 7. Schematic diagram of a micromachined thermopile infrared detector. The hot junctions of the thermopile (the sensing element) are fabricated on a free-standing cantilever while the cold junctions are thermally attached to the silicon rim which acts as the heat sink. These devices are fabricated using silicon microfabrication techniques along with anisotropic etching of silicon.

a thermopile is located along the contours of the highest temperatures on the free-standing structure whereas the reference junction is located along the thick rim of the substrate. Figure 7 schematically illustrates a typical micromachined cantilever structure based thermopile. This technology is naturally oriented toward the use of semiconductors and metal thin films as the thermoelements. Fortunately, semiconductors, as we have discussed earlier, are found to have a high thermoelectric figure of merit and make an excellent choice for this technology. To further improve the responsivity and spectral response, these infrared detectors are often coated with absorber materials such as gold black or with infrared absorbing thin films (18,19). Thin-film thermopiles with a responsivity of several tens of volts per watts and a time constant of tens of milliseconds are commercially available. Figure 8 shows a commercial thin film Bi/Sb thermopile infrared detector from Dexter Research Center and

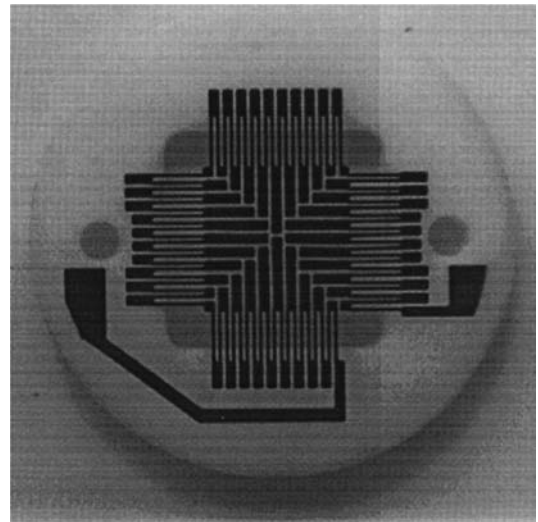


Figure 8. Optical micrograph of a commercial thin film Bi/Sb thermopile infrared detector from Dexter Research Center. The device consists of a 2 mm $\times$ 2 mm active area and 40 thermocouple junctions connected in series. The detector has a responsivity of 19.2 V/W and a time constant of 40 ms. (Courtesy: Dexter Research Inc., Michigan.)

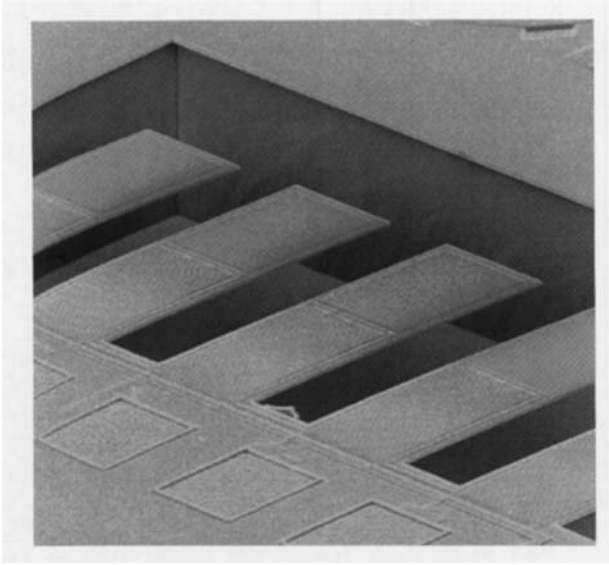


Figure 9. SEM (scanning electron microscope) micrograph of a cantilever beam infrared sensor array. The length of the oxide/nitride cantilever beam is $300\text{ }\mu\text{m}$, the width is $200\text{ }\mu\text{m}$ and the depth of the etch groove is about $120\text{ }\mu\text{m}$. The thermopile is made up of *n*-polysilicon/*p*-polysilicon as the thermoelements. (Courtesy Physical Electronics Laboratory, ETH, Zurich, Ref. 14)

Figure 9 shows a typical CMOS (complementary metal-oxide-semiconductor) based polysilicon thermopile. Further reduction in the thermal mass of such sensors has been achieved by completely removing the supporting membranes under the thermoelements. These modifications have reduced the time constants of the infrared detectors to a few microseconds. Such a free-standing copper-constantan microthermopile is shown in Fig. 10. Current efforts in this field are focused on the further optimization of the thermo-mechanical design of the sensing elements (20), utilizing thin film semi-metals and semiconductors as thermoelements and the integration of the sensor with on-chip antenna structures (21). Efforts are also focused on fabricating these devices as planar imaging arrays with integrated CMOS circuitry for signal conditioning and readout (22).

Integrated Thermopiles

The demand for intelligent sensor systems, which has included an increasing number of sensors as well as an increasingly complex analog and digital circuitry, has been the motivation behind the development of integrated microsensors. In particular, thermal sensors are at the heart of several such microsystems. Thermal sensors first convert the measurement signals into a thermal signal, which is then transduced into an electrical signal primarily using thermoelectric or thermoresistive techniques. Thermoresistive elements (thermistors) compete in nearly every application along with thermopiles and are also widely used in sensor applications. This section will limit the discussion only to the use of integrated thermopiles in various microsensor applications. Integrated thermopiles offer the possibility of batch fabricating microsensors and systems at large volumes and low cost with high reliability and thus form a subject of great current interest. Integrated thermopiles fabricated using micromachining technology are used, as the transduction elements, in several thermal microsensors ap-

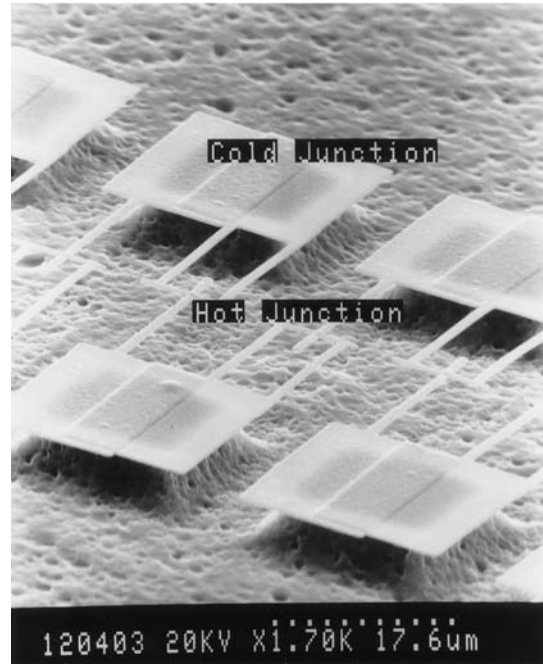


Figure 10. SEM micrograph of a copper-constantan free-standing microthermopile. The free-standing wires are $1\text{ }\mu\text{m}$ wide and the length of the free-standing wires is $35\text{ }\mu\text{m}$. The cold junctions (large pads) are thermally attached to the substrate. The small thermal mass of the sensing element results in a fast response time of $\sim 20\text{ }\mu\text{s}$.

plications which can be broadly classified as: (1) thermomechanical sensors (mass flow sensors, pressure sensors, and accelerometers), (2) thermal converters (electrical power sensors), and (3) thermochemical sensors (calorimeters and reaction enthalpy sensors).

Mass Flow Sensors, Pressure Sensors, and Accelerometers

In a mass flow sensor, a thermally isolated resistor is heated using a constant power. Thermopile sensors are placed along the upstream and downstream of the flow with the resistor heater in between them. The heat transfer coefficient, which is the ratio of the heat flow (from the heater surface to the flowing fluid) and the temperature difference (between the heater surface and the flowing fluid), among other parameters, is a function of the flow velocity. It is larger for the upstream flow than for the downstream flow and therefore induces a temperature difference between the upstream and the downstream thermopiles which is used to detect the flow velocity as well as the flow direction along the line of the sensors (23,24). The heat transfer coefficient is a very sensitive function of the heater and sensor geometry, the nature of fluid flow (laminar, turbulent, and so on), the heat capacity and thermal conductivity of the fluid itself. A simpler way of detecting flow is to measure the increase in power required to maintain the temperature of a heated resistor at a constant value in the presence of fluid flow. This is the principle on which hot wire anemometer work and the power, P , dissipated in the wire is related to the flow velocity, v , as

$$P/\Delta T \propto v^{1/2} \quad (19)$$

where ΔT is the temperature difference between the wire and

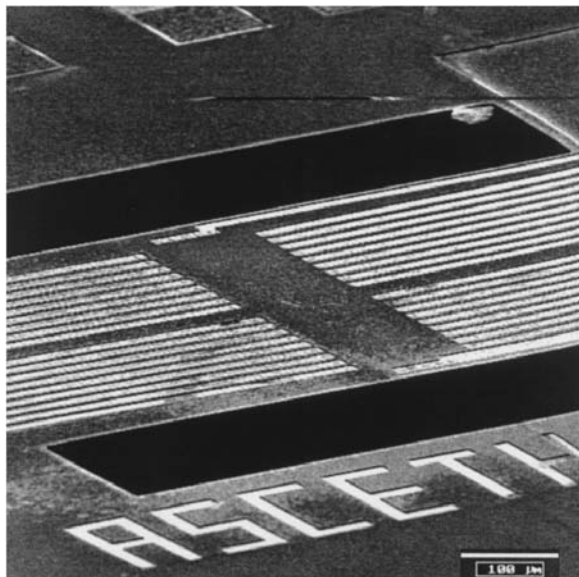


Figure 11. SEM picture of a bridge type flow sensor fabricated using bipolar technology. An *n*-doped monocrystalline silicon/aluminum thermopile with 36 thermocouples is located on both sides of a 1 mm by 0.5 mm bridge with a diffused heating resistor in the center of the bridge. (Courtesy Physical Electronics Laboratory, ETH, Zurich, Ref. 23)

the fluid. Figure 11 shows a CMOS thermopile based flow sensor.

Thermal conductance of a gas is a function of the gas pressure (i.e., the dissipation of thermal energy from a heater to a heat sink across a small gap filled with the gas depends on the gas pressure). For a constant power dissipation in the heater, this pressure-dependent thermal conductance of the gas causes a variation in the heater temperature and provides an accurate way of measuring the gas pressure. The temperature of the resistor heater is measured using an integrated silicon-based thermopile. For the typical micromachined sensor geometries, these pressure sensors are best suited for vacuum measurements in the range of 10 mPa to 10 kPa. Such integrated thermopile-based vacuum sensors have been fabricated using both bulk and surface micromachining techniques, see (Refs. 6) and (25).

A silicon accelerometer is based on the detection of movement of a suspended mass under the influence of acceleration. A thermopile based accelerometer detects the change in heat flow between the suspended mass and a heat source under the influence of acceleration. A typical sensor consists of a thermally isolated bridge with thermopiles on top of a heat source and a proof mass etched out of bulk silicon suspended in close proximity over the bridge (26).

Thermal Converters and Thermal Conductivity Sensors

In a thermal converter, electric power is dissipated into an ohmic resistor and converted into thermal power which is then measured using a thermopile-based temperature sensor. The operation of these sensors is based on the square-law relationship between the voltage and power. These devices have been used in wide-frequency-band ac power measurements and in particular for the measurement of microwave ac power

(true root mean square voltage value) (6,25). In addition to ac power measurements, integrated thermopiles have also been used for the measurement of thermal conductivity of thin films and materials. Such sensors are based on the fabrication of micromachined freestanding structures of appropriate geometries in which the contribution to thermal effects of specific layers can be effectively isolated and quantified. These freestanding structures are heated by dissipating known quantities of electrical power and integrated thermopiles are used for the accurate temperature measurement in such structures (6,25)

Integrated Thermopile-Based Calorimeters

In microcalorimetry applications, the measurement junction of a thermopile is placed on a thermally isolated bridge or membrane structure and is coated with a catalytic layer for the desired chemical reaction. Under the influence of the surface catalyst, the chemical reaction takes place on the membrane surface and the heat of the reaction is detected by the change in temperature of the measurement junction of the thermopile with respect to the reference junction (27). This is an example of a chemical sensor based on microcalorimetry. If instead of a catalyst, a coating that converts absorbed microwaves into heat is used, the microcalorimeter can be used as a microwave sensor. Another example would be to coat the thermally isolated structure with a ferromagnetic coating with high hysteresis. Ac magnetic fields would cause heat generation in such a sensor structure, which can be measured using an integrated thermopile (8).

Thermoelectric Refrigeration

The passage of an electric current through a thermoelectric circuit can be used to cool one of its junctions by selectively pumping away heat to the other junction using Peltier effect. Since thermoelectric refrigerators contain no moving parts, they operate quietly, require very low maintenance and have a long life. In addition, the coefficient of performance of thermoelectric refrigerators is independent of the size of the system to be cooled which makes them ideally suited for low capacity cooling applications. However, the widespread use of thermoelectric refrigerators has been hindered by the higher costs and lower cooling efficiency as compared with the more commonly used compression-cycle refrigerators. Consequently, thermoelectric refrigerators are employed in certain niche application such as in spacecraft, artificial satellites, scientific equipment, cooled stages for laboratory applications where compactness, quiet performance, operation in vacuum environments, and high reliability are of major concern. Small cooling systems with powers of less than 10 W are commonly used as cold traps in vacuum systems, cooled stages inside vacuum chambers, as cooling jackets for cooled infrared detectors, as active heat sinks for cooling of main processor chips in personal computers, and for controlling the temperature of thermocouple reference junctions. Reviews on thermoelectric refrigeration can be found in Rowe and Bhandari (4) and Ioffe (28).

A thermoelectric refrigerator module is normally constructed such that all the cold junctions of the thermopile are on one face and the hot junctions on another face. Heat is pumped to the hot junction at a rate proportional to the current in the circuit and to the number of thermocouples, *N*, constituting

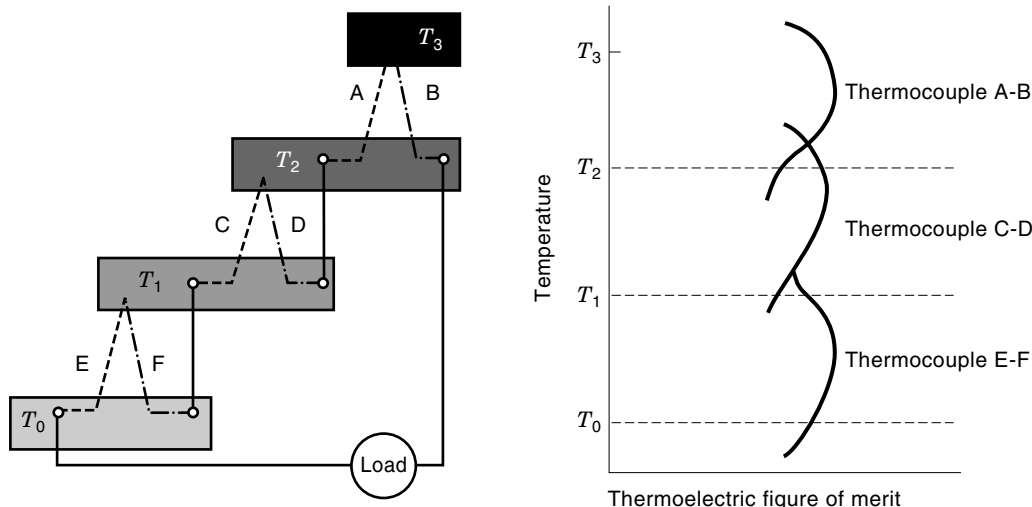


Figure 12. Schematic diagram illustrating the use of different thermocouples in a temperature sandwich arrangement. Each thermocouple operates at the best average Z value and the overall figure of merit is given by the envelope as shown on the right hand side.

the thermopile. The most widely used materials in the construction of thermoelectric refrigerators are alloys of bismuth, antimony, selenium, and tellurium. Single layer module sizes range from 2.5 mm to 50 mm square and from 2.5 mm to 7.5 mm thick, with cooling capacities ranging from a few watts to about 100 W. The maximum temperature difference that a single thermoelectric module can produce for a cooling power, $P_c = 0$ W, is around 65°C .

Thermoelectric materials have a different value for the thermoelectric figure of merit, Z , at different temperatures. If several thermocouple pairs are connected in series in a cascade arrangement at different temperatures as shown in Fig. 12, such that each thermopile operates at its best Z value, the overall coefficient of performance of the refrigerator can be increased. Since the lower stages remove heat dissipated by the stages above them and from the load, multistage modules are constructed in a pyramidal shape as shown in Fig. 13. Multistage modules of up to six layers can generate several hundreds of watts of cooling capability and provide temperatures as low as -100°C with the heat sink at 27°C . In most small refrigeration systems with cooling powers in the range of up to 50 W, the

pumped heat is dissipated into the surroundings by natural air convection or by using cooling liquids. In general, thermoelectric refrigeration systems are most economical cooling solutions for small cooling powers up to 50 W. In the 50 W and above cooling power range, thermoelectric systems are often more expensive and always less efficient than compression cycle systems. For these cooling powers, a thermoelectric system is only used if some other characteristic is of greater importance. Review of industrial thermoelectric cooling applications are presented in (Refs. 29) and (30).

Thermoelectric Generators

Thermopiles can also be employed to generate electrical power if a temperature difference is maintained across its two faces. These thermoelectric generators are primarily used as remote, maintenance free, unattended power supplies in space stations and space vehicles, unmanned light houses, navigational buoys, remote mines, and drilling platforms. Another area where thermoelectric generators find application is in medical applications, especially as miniature batteries delivering under $300 \mu\text{W}$ of energy in cardiac pacemakers and other stimulator applications (4).

The main components of a thermoelectric generator are: (1) a heat source, (2) a thermopile, (3) a heat sink, and (4) an output control system. Most often, the physical arrangement of a thermopile generator is determined by the nature of its heat source. Thermoelectric generators are normally classified by the kind of fuel source they use as isotopic or nonisotopic. Nonisotopic generators are mainly powered by fossil fuels and to a limited extent by solar radiation. In isotopic generators, radioactive isotopes are used as the fuel sources. Radioactive isotopes emit high energy density over long periods of time. For example, plutonium (P-238) has a half life of 87 years and so the thermal output from the fuel capsule only decreases by 4% in 10 years. Thus, radioactive sources are very attractive fuel sources for thermoelectric generators. Currently all radioactive thermoelectric generators (RTGs) use either plutonium-238 or strontium-90 as the fuels (4,5). Depending on the kind of

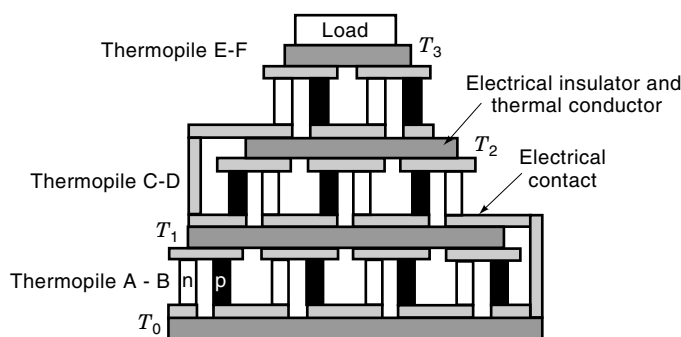


Figure 13. A three-stage schematic of the cascade arrangement of thermopiles. This arrangement affords flexibility in the geometrical requirements of each stage for optimal performance.

radiation emitted by each of these sources, elaborate care has to be taken to shield and confine the emission while maintaining the compactness and light weight of the entire system. In general, plutonium-238 requires minimal shielding and is best suited for space and medical applications where cost considerations are not important, while strontium-90 at one fortieth the cost of plutonium is employed in terrestrial applications where bulkiness of shielding is not an issue. However, due to safety and security reasons, RTGs have not been used for medical applications since the late 1980s. On the positive side, though, the development of these generators has had a significant impact on the development of high Z thin film materials.

Aside from the heat source, production of power using thermoelectric generators apparently seems very straightforward. The practical implementation, however, involves a very complex arrangement of coupling the heat source to the thermopile junctions while electrically isolating the junctions at the hot face of the generator module at the high operating temperatures. In addition, problems associated with material degradation and poisoning when continuously operated at high temperatures reduce the conversion efficiencies and performance over time. In general, the overall conversion of a thermoelectric device depends upon the temperature of its operation and the proportion of available heat which passes through the thermoelements.

Conventionally, thermoelectric power is generated at relatively high values of current and low voltage, with the load voltage from the thermoelectric module being about half the open circuit voltage of the thermopile. Solar powered thermoelectric generators using selective absorber coatings and with thermoelements fabricated from fine grained Si-Ge alloys have been predicted to operate with an efficiency of better than 12% between room temperature and 1000 K. A portable flame-powered generator, weighing about 5 kg, using leaded petrol as fuel, has been developed with a power output of 45 W at 6 Vdc and an overall efficiency of 2%. On the other hand, batteries powered by plutonium-238 and a bismuth-telluride thermopile module can supply several tenths of a volt for more than 10 years.

BIBLIOGRAPHY

1. R. D. Barnard, *Thermoelectricity in Metals and Alloys*, London: Taylor & Francis, 1972.
2. D. D. Pollock, *Thermocouples—Theory and Properties*, Boca Raton, FL: CRC Press, 1991.
3. Manual on the use of thermocouples in temperature measurement, STP-470A, *American Society for Testing and Materials (ASTM)*, 1974.
4. D. M. Rowe and C. M. Bhandari, *Modern Thermoelectrics*, Reston, VA: Reston Publishing, 1983.
5. J. Kaye and J. A. Welsh, *Direct Conversion of Heat to Electricity*, New York: Wiley, 1960.
6. H. Baltes, D. Moser, and F. Völklein, Thermoelectric microsensors and microsystems, in *Sensors: A comprehensive survey*, W. Göpel, J. Hesse, and J. N. Zemel (eds.), Vol. 7, Weinheim, VCH, 1994.
7. T. J. Quinn, *Temperature*, London: Academic Press, 1983.
8. A. W. Van Herwaarden and P. M. Sarro, Thermal sensors based on the Seebeck effect, *Sensors and Actuators*, **10**: 321–346, 1986.
9. S. Middelhoek, *Silicon Sensors*, London: Academic Press, 1989.
10. R. A. Smith, F. E. Jones, and R. P. Chasmar, *The Detection and Measurement of Infrared Radiation*, London: Oxford University Press, 1968.
11. R. J. Keyes (ed.), *Optical and infrared detectors, Topics in Applied Physics*, 19, Berlin: Springer, 1977.
12. J. Müller, U. Dillner, and R. Güttich, Thermoelectric and Bolometric Infrared Microsensors, in *Sensors Update Vol. 3*, H. Baltes, W. Göpel, and J. Hesse, (eds.), Weinheim, VCH, 1997.
13. G. R. Lahiji and K. D. Wise, A batch fabricated silicon thermopile infrared detector, *IEEE Trans. Electron Devices*, **ED-29**: 14–22, 1982.
14. R. Lenggenhager, H. Bates, and J. Peer, et al., Thermoelectric infrared sensors by CMOS technology, *IEEE Electron Devices Lett.*, **13**: 454–456, 1992.
15. T. A. S. Srinivas, P. J. Timans, and H. Ahmed, A high-performance infrared detector using MOS technology, *Sensors Materials*, **8** (6): 317–326, 1996.
16. K. E. Petersen, Silicon as a mechanical materials, *Proc. IEEE*, **70**: 420–457, 1982.
17. D. B. Lee, Anisotropic etching of silicon, *J. Appl. Phys.*, **40**: 4569–4574, 1969.
18. S. Bauer et al., Thin metal films as absorbers for infrared sensors, *Sensors Actuators A*, **37-38**: 497–501, 1993.
19. W. Lang, K. Kuhl, and H. Sandmaier, Absorbing layers for thermal infrared detectors, *Proc. Int. Conf. Solid State Sensors and Actuators*, San Francisco, USA, 635–638, 1991.
20. F. Völklein and H. Baltes, Optimization tool for the performance parameters of thermoelectric microsensors, *Sensors Actuators A*, **36**: 65–71, 1993.
21. N. Chong and H. Ahmed, Antenna-coupled polycrystalline silicon air-bridge thermal detector for mid-infrared radiation, *Appl. Phys. Lett.*, **71** (12): 1607–1609, 1997.
22. A. D. Oliver, W. G. Baer, and K. D. Wise, Bulk-micromachined 1024-element uncooled infrared imager, *Proc. 8th Int. Conf. Solid-State Sensors Actuators, Eurosensors IX*, Stockholm, Sweden, 636–639, 1995.
23. D. Moser and H. Baltes, High sensitivity CMOS gas flow sensor based on a thin dielectric membrane, *Sensors Actuators A*, **37-38** (2): 33–37, 1993.
24. U. Dillner, et al., Low power consumption thermal gas-flow sensor based on thermopiles of highly effective thermoelectric materials, *Sensors and Actuators A*, **60**, 1–4, 1997.
25. H. Baltes, O. Paul, and D. Jaeggi, Thermal CMOS sensors—an overview, in *Sensors Update Vol. 1*, 121–142, H. Baltes, W. Göpel, and J. Hesse (eds.), Weinheim, VCH, 1997.
26. R. Hiratsuka et al., A novel accelerometer based on a silicon thermopile, *Proc. Int. Conf. Solid State Sensors Actuators*, San Francisco, USA, 420–423, 1991.
27. J. M. Koehler et al., Thermoelectrical microcalorimetry and microanalysis, on free-standing membranes, *13th Int. Symp. Microchemical Techniques*, Montreux, 1994.
28. A. F. Ioffe, *Semiconductor Thermoelements and Thermoelectric Cooling*, London: Infosearch, 1957.
29. J. G. Stockholm and P. M. Schlicklin, Industrial thermoelectric cooling and electricity generation between 200 K and 500 K, in *First Eur. Conf. Thermoelectrics*, D. M. Rowe, (ed.), 235–263, London, Peter Peregrinus, 1988.
30. H. J. Goldsmid, Electronic refrigeration, in *First Eur. Conf. Thermoelectrics*, D. M. Rowe (ed.), 342–368, London, Peter Peregrinus, 1988.

SRINIVAS TADIGADAPA
Integrated Sensing Systems
Incorporated

ACCELEROMETERS

Acceleration is an important parameter for general-purpose absolute motion measurements and vibration and shock sensing. Accelerometers are commercially available in a wide variety of ranges and types to meet diverse application requirements. They are manufactured to be small in size, light in weight, and rugged and robust to operate in harsh environments. They can be configured as active or passive sensors. An active accelerometer (e.g., piezoelectric) gives an output without the need for an external power supply, while a passive accelerometer only changes its electric properties (e.g., capacitance) and requires an external electrical power. In applications, the choice of active- or passive-type accelerometers is important, since active sensors cannot measure static or dc mode operations. For true static measurements, passive sensors must be selected.

Accelerometers can be classified in a number of ways, such as deflection or null-balance types, mechanical or electrical types, and dynamic or kinematic types. The majority of industrial accelerometers can be classified as either deflection type or null-balance type. Those used in vibration and shock measurements are usually the deflection types, whereas those used for measurements of motions of vehicles, aircraft, and so on for navigation purposes may be either type. In general, null-balance types are used when extreme accuracy is needed.

A large number of practical accelerometers are the deflection types; the general configuration is shown in Fig. 1. There are many different deflection-type accelerometers. Although principles of operation are similar, they differ in minor details, such as the spring elements used, types of damping provided, and types of relative motion transducers employed. These types of accelerometers behave as second-order systems; the detailed mathematical analysis will be given in the following sections.

Dynamic accelerometers have an operation that is based on measuring the force required to constrain a seismic mass to track the motion of the accelerated base, such as a spring-constrained-slug accelerometer. Although applicable to all, the mathematical treatment of the dynamic response of an accelerometer as a second-order system is given in detail in

the section dedicated to seismic accelerometers. Another type is the kinematic accelerometer, which is based on timing the passage of an unconstrained proof mass from spaced points marked on the accelerated base, and is used for highly specific applications such as interspace spacecraft and gravimetry-type measurements.

For practical purposes, accelerometers can also be classified as mechanical or electrical types, depending on whether the restoring force or other measuring mechanism is based on mechanical properties (for example, the law of motion, distortion of a spring, or fluid dynamics) or on electrical or magnetic forces.

TYPES OF ACCELEROMETERS

Seismic Accelerometers

These accelerometers make use of a seismic mass that is suspended by a spring or a lever inside a rigid frame. The schematic diagram of a typical seismic accelerometer is shown in Fig. 1. The frame carrying the seismic mass is connected firmly to the vibrating source whose characteristics are to be measured. As the system vibrates, the mass tends to remain fixed in its position so that the motion can be registered as a relative displacement between the mass and the frame. This displacement is sensed by an appropriate transducer and the output signal is processed further. Nevertheless, the seismic mass does not remain absolutely steady, but for selected frequencies it can satisfactorily act as a reference position.

By proper selection of mass, spring, and damper combinations, the seismic instruments may be used for either acceleration or displacement measurements. In general, a large mass and soft spring are suitable for vibration and displacement measurements, while relatively small mass and a stiff spring are used in accelerometers. However, the term *seismic* is commonly applied to instruments that sense very low levels of vibration in the ground or in structures. They tend to have low natural frequencies.

The following equation may be written by using Newton's second law of motion to describe the response of seismic arrangements similar to shown in Fig. 1:

$$m d^2 x_2 / dt^2 + c dx_2 / dt + k x_2 = c dx_1 / dt + k x_1 + m g \cos(\theta) \quad (1)$$

where x_1 is the displacement of the vibration frame, x_2 is the displacement of the seismic mass, c is the velocity constant, θ is the angle between the sense axis and gravity, and k is the spring constant.

Subtracting $m d^2 x_1 / dt^2$ from both sides of the equation and rearranging gives

$$m d^2 z / dt^2 + c dz / dt + k z = m g \cos(\theta) - m d^2 x_1 / dt^2 \quad (2)$$

where $z = x_2 - x_1$ is the relative motion between the mass and the base.

In Eq. (1), it is assumed that the damping force on the seismic mass is proportional to velocity only. If a harmonic vibratory motion is impressed on the instrument such that

$$x_1 = x_0 \sin \omega_1 t \quad (3)$$

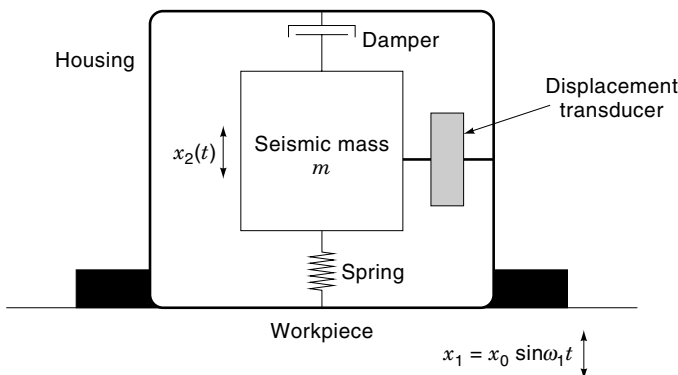


Figure 1. A typical deflection-type seismic accelerometer. In this basic accelerometer, the seismic mass is suspended by a spring or cantilever inside a rigid frame. The frame is connected to the vibrating structure; as vibrations take place the mass tends to remain fixed so that relative displacements can be picked up. They are manufactured in many different types and sizes with diverse characteristics.

where ω_1 is the frequency of vibration (rad/s). Writing

$$m d^2 x_1 / dt^2 = m x_0 \omega_1^2 \sin \omega_1 t$$

modifies Eq. (2) as

$$-m d^2 z / dt^2 + c dz / dt + kz = mg \cos(\theta) + m a_1 \sin \omega_1 t \quad (4)$$

where $a_1 = m x_0 \omega_1^2$.

Equation (4) will have transient and steady-state solutions. The steady-state solution of the differential equation (4) may be determined as

$$z = [mg \cos(\theta) / k] + [m a_1 \sin \omega_1 t / (k - m \omega_1^2 + j c \omega_1)] \quad (5)$$

Rearranging Eq. (5) results in

$$z = [mg \cos(\theta) / \omega_n] + \{a_1 \sin(\omega_1 - \phi) / [\omega_n^2 (1 - r^2)^2 + (2\zeta r)^2]^{1/2}\} \quad (6)$$

where $\omega_n = \sqrt{k/m}$ is the natural frequency of the seismic mass, $\zeta = (c/2) \sqrt{k/m}$ is the damping ratio, which also can be written in terms of critical damping ratio as $\zeta = c/c_c$, where $c_c = 2\sqrt{km}$, $\phi = \tan^{-1}[c\omega_1/(k - m\omega_1^2)]$ is the phase angle, and $r = \omega_1/\omega_n$ is the frequency ratio.

A plot of Eq. (6), $(x_1 - x_2)_0/x_0$ against frequency ratio ω_1/ω_n is illustrated in Fig. 2. This figure shows that the output amplitude is equal to the input amplitude when $c/c_c = 0.7$ and $\omega_1/\omega_n > 2$. The output becomes essentially a linear function of the input at high frequency ratios. For satisfactory system performance, the instrument constant c/c_c and ω_n should carefully be calculated or obtained from calibrations. In this way the anticipated accuracy of measurement may be predicted for frequencies of interest. A comprehensive treatment of the analysis is by McConnell (1); interested readers should refer to this text for further details.

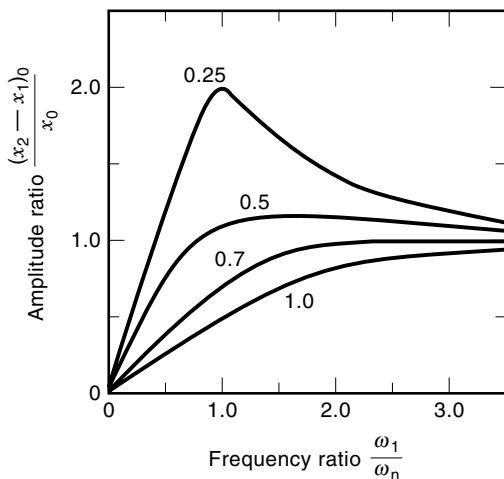


Figure 2. A typical displacement of a seismic instrument. The amplitude becomes large at low damping ratios. The instrument constants should be selected such that in measurements the frequency of vibration is to be much higher than the natural frequency, for example, greater than 2. Optimum results may be obtained when the value of the instrument constant c/c_c is about 0.7.

If the seismic instrument has a low natural frequency and a displacement sensor is used to measure the relative motion z , then the output is proportional to the displacement of the transducer case. If the velocity sensor is used to measure the relative motion, the signal is proportional to the velocity of the transducer. This is valid for frequencies significantly above the natural frequency of the transducer. Velocity coil output produces a device commonly known as a geophone. It is not an accelerometer in the strict sense but it is similarly used. It excels at measuring low to medium frequency vibrations, as it offers exceptionally low self-generated noise output and very low output impedance. However, if the instrument has a high natural frequency and the displacement sensor is used, the measured output is proportional to the acceleration

$$kz = m d^2 x_1 / dt^2 \quad (7)$$

This equation is true since displacement x_2 becomes negligible in comparison with x_1 .

In these instruments the input acceleration a_0 can be calculated simply by measuring $(x_1 - x_2)_0$, the static deflection relative to the case. Generally, in acceleration measurements, unsatisfactory performance is observed at frequency ratios above 0.4. Thus, in such applications, the frequency of acceleration must be kept well below the natural frequency of the instrument. This can be done by constructing the instrument to have a low natural frequency by selecting soft springs and large masses.

Seismic instruments are constructed in a variety of ways. Figure 3 illustrates the use of a voltage divider potentiometer for sensing of the relative displacement between the frame and the seismic mass. In the majority of potentiometric instruments, the device is filled with a viscous liquid that interacts continuously with the frame and the seismic mass to provide damping. These accelerometers have a low frequency of operation (less than 100 Hz) and are mainly intended for slowly varying acceleration and low-frequency vibrations. A typical family of such instruments offers many different models, covering the range of $\pm 1 g$ to $\pm 50 g$ full scale. The natural frequency ranges from 12 Hz to 89 Hz, and the damping ratio ζ can be kept between 0.5 to 0.8 by using a temperature-compensated liquid-damping arrangement. Potentiometer resistance may be

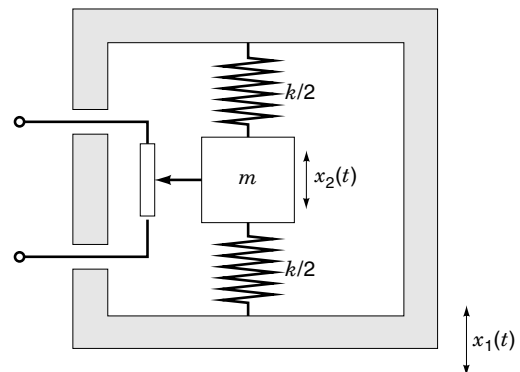


Figure 3. A potentiometer accelerometer. The relative displacement of the seismic mass is sensed by a potentiometer arrangement. The potentiometer adds extra weight, making these accelerometers relatively heavier. Suitable liquids filling the frame may be used as damping elements. These accelerometers are used in low-frequency applications.

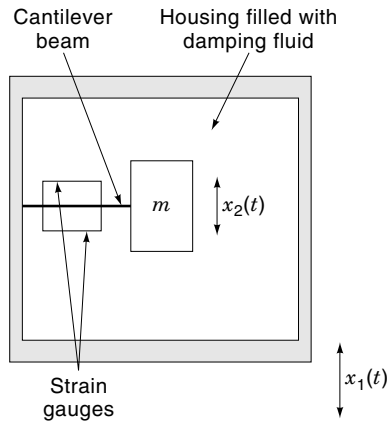


Figure 4. A strain-gauge seismic instrument. The displacement of the proof mass is sensed by piezoresistive strain gauges. The natural frequency of the system is low due to need of a long level beam to accommodate strain gauges. The signal is processed by bridge circuits.

selected in the range of $1,000 \Omega$ to $10,000 \Omega$, with a corresponding resolution of 0.45% to 0.25% of full scale. The cross-axis sensitivity is less than $\pm 1\%$. The overall accuracy is $\pm 1\%$ of full scale or less at room temperatures.

Linear variable differential transformers (LVDTs) offer another convenient means of measurement of the relative displacement between the seismic mass and the accelerometer housing. These devices have higher natural frequencies than potentiometer devices, up to 300 Hz. Since the LVDT has lower resistance to motion, it offers much better resolution. A typical family of liquid-damped differential-transformer accelerometers exhibits the following characteristics. The full scale ranges from $\pm 2 g$ to $\pm 700 g$, the natural frequency from 35 Hz to 620 Hz, the nonlinearity 1% of full scale, the full-scale output is about 1 V with an LVDT excitation of 10 V at 2,000 Hz, the damping ratio ranges from 0.6 to 0.7, the residual voltage at the null position is less than 1%, and the hysteresis is less than 1% full scale. The size is 50 mm^3 , with a mass of about 120 g.

Electrical resistance strain gauges are also used for displacement sensing of the seismic mass as shown in Fig. 4. In this case, the seismic mass is mounted on a cantilever beam rather than on springs. Resistance strain gauges are bonded on each side of the beam to sense the strain in the beam resulting from the vibrational displacement of the mass. Damping for the system is provided by a viscous liquid that entirely fills the housing. The output of the strain gauges is connected to an appropriate bridge circuit. The natural frequency of such a system is about 300 Hz. The low natural frequency is due to the need for a sufficiently large cantilever beam to accommodate the mounting of the strain gauges. Other types of seismic instruments with piezoelectric transducers using seismic masses are discussed in detail in the section dedicated to piezoelectric-type accelerometers.

Seismic vibration instruments are affected seriously by temperature changes. Devices employing variable resistance displacement sensors will require correction factors to account for resistance change due to temperature. The damping of the instrument may also be affected by changes in the viscosity of the fluid due to temperature. For instance, the viscosity of silicone oil, often used in these instruments, is strongly

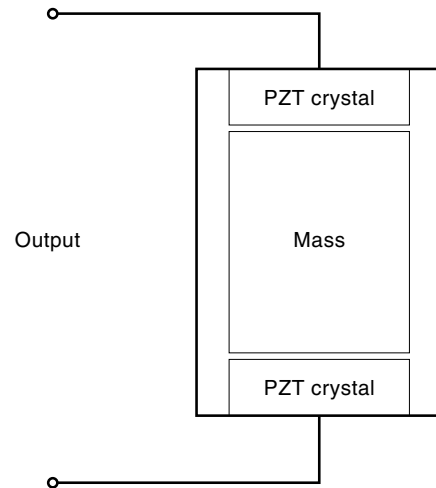


Figure 5. A compression-type piezoelectric accelerometer. The crystals are under compression at all times either by a mass or mass and spring arrangement. Acceleration causes a deformation of the crystal, thus producing a proportional electrical signal. They are small in size and widely used. They demonstrate poor performance at low frequencies.

dependent on temperature. One way of eliminating the temperature effect is to use an electrical resistance heater in the fluid to maintain the temperature at a constant value regardless of surrounding temperatures.

Piezoelectric Accelerometers

Piezoelectric accelerometers are used commonly for general-purpose acceleration, shock, and vibration measurements. They are basically motion transducers with large output signals and comparatively small size. They are available with very high natural frequencies and are therefore suitable for high-frequency applications and shock measurements.

These devices utilize a mass in direct contact with the piezoelectric component or crystal as shown in Fig. 5. When a varying motion is applied to the accelerometer, the crystal experiences a varying force excitation ($F = ma$), causing a proportional electric charge q to be developed across it.

$$q = d_{ij}F = d_{ij}ma \quad (8)$$

where q is the charge developed and d_{ij} is the material's piezoelectric coefficient.

As this equation shows, the output from the piezoelectric material is dependent on its mechanical properties, d_{ij} . Two commonly used piezoelectric crystals are lead-zirconate titanate ceramic (PZT) and crystalline quartz. They are both self-generating materials and produce a large electric charge for their size. The piezoelectric strain constant of PZT is about 150 times that of quartz. As a result, PZTs are much more sensitive and smaller in size than quartz counterparts. In the accelerometers, the mechanical spring constants for the piezoelectric components are high, and the inertial masses attached to them are small. Therefore, these accelerometers are useful for high-frequency applications. Figure 6 illustrates a typical frequency response of a PZT device. Typically, the roll-off starts near 100 Hz. These active devices have no dc response. Since

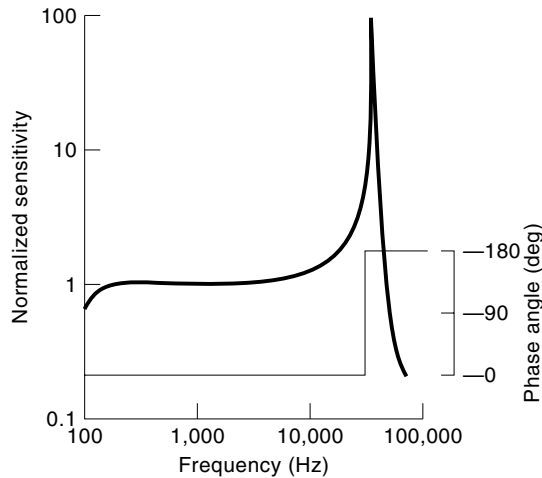


Figure 6. The frequency response of a typical piezoelectric accelerometer. Measurements are normally confined to the linear portion of the response curve. The upper frequency of the accelerometer is limited by the resonance of the PZT crystal. The phase angle is constant up to the resonance frequency.

piezoelectric accelerometers have comparatively low mechanical impedances, their effects on the motion of most structures is negligible. They are also manufactured to be rugged and have outputs that are stable with respect to time and environment.

Mathematically, their transfer function approximates to a third-order system as

$$e_o(s)/a(s) = (K_q/C\omega_n^2) \tau s / [(\tau s + 1)(s^2/\omega_n^2 + 2\zeta s/\omega_n + 1)] \quad (9)$$

where K_q is the piezoelectric constant related to charge (C·cm), τ is the time constant of the crystal, and s is the Laplace variable. It is worth noting that the crystal itself does not have a time constant τ , but the time constant is observed when the accelerometer is connected into an electric circuit, for example, and RC circuit.

The low-frequency response is limited by the piezoelectric characteristic $\tau s/(\tau s + 1)$, while the high-frequency response is related to mechanical response. The damping factor ζ is very small, usually less than 0.01 or near zero. Accurate low-frequency response requires large τ , which is usually achieved by use of high-impedance voltage amplifiers. At very low frequencies thermal effects can have severe influences on the operation characteristics.

In piezoelectric accelerometers, two basic design configurations are used: compression types and shear-stress types. In compression-type accelerometers, the crystal is held in compression by a preload element; therefore the vibration varies the stress in compressed mode. In a shear-stress accelerometer, vibration simply deforms the crystal in shear mode. The compression accelerometer has a relatively good mass-to-sensitivity ratio and hence exhibits better performance. But, since the housing acts as an integral part of the spring-mass system, it may produce spurious interfaces in the accelerometer output if excited around its proper natural frequency.

Microelectronic circuits have allowed the design of piezoelectric accelerometers with charge amplifiers and other signal-conditioning circuits built into the instrument housing. This arrangement allows greater sensitivity and high-frequency

response and smaller size accelerometers, thus lowering the initial and implementation costs.

Piezoelectric accelerometers are available in a very wide range of specifications and are offered by a large number of manufacturers. For example, the specifications of a shock accelerometer may have 0.004 pC/g in sensitivity and a natural frequency of up to 250,000 Hz, while a unit designed for low-level seismic measurements might have 1,000 pC/g in sensitivity and only 7,000 Hz natural frequency. They are manufactured as small as 3×3 mm² in dimension with about 0.5 g in mass, including cables. They have excellent temperature ranges and some of them are designed to survive the intensive radiation environment of nuclear reactors. However, piezoelectric accelerometers tend to have larger cross-axis sensitivity than other types, about 2% to 4%. In some cases, large cross-axis sensitivity may be minimized during installations by the correct orientation of the device. These accelerometers may be mounted with threaded studs, with cement or wax adhesives, or with magnetic holders.

Electromechanical Force-Balance (Servo) Accelerometers

Electromechanical accelerometers, essentially servo or null-balance types, rely on the principle of feedback. In these instruments, an acceleration-sensitive mass is kept very close to a neutral position or zero displacement point by sensing the displacement and feeding back the effect of this displacement. A proportional magnetic force is generated to oppose the motion of the mass displaced from the neutral position, thus restoring this position just as a mechanical spring in a conventional accelerometer would do. The advantages of this approach are better linearity and elimination of hysteresis effects as compared with the mechanical springs. Also, in some cases, electrical damping can be provided, which is much less sensitive to temperature variations.

In high-vibration environments, force balance accelerometers benefit from two unique capabilities: velocity storage allows them to operate at saturation a small percentage of the time without actually losing information, and dynamic range change permits the useful range to be greater at high frequencies than near dc.

One very important feature of null-balance type instruments is the capability of testing the static and dynamic performances of the devices by introducing electrically excited test forces into the system. This remote self-checking feature can be quite convenient in complex and expensive tests in which it is extremely critical that the system operates correctly before the test commences. These instruments are also useful in acceleration control systems, since the reference value of acceleration can be introduced by means of a proportional current from an external source. They are usually used for general-purpose motion measurements and monitoring low-frequency vibrations. They are specifically applied in measurements requiring better accuracy than that achieved by those accelerometers based on mechanical springs such as the force-to-displacement transducer.

There are a number of different electromechanical accelerometers: coil-and-magnetic types, induction types, etc.

Coil-and-Magnetic Accelerometers. These accelerometers are based on Ampere's law, that is, "a current-carrying conductor disposed within a magnetic field experiences a force proportional to the current, the length of the conductor within the

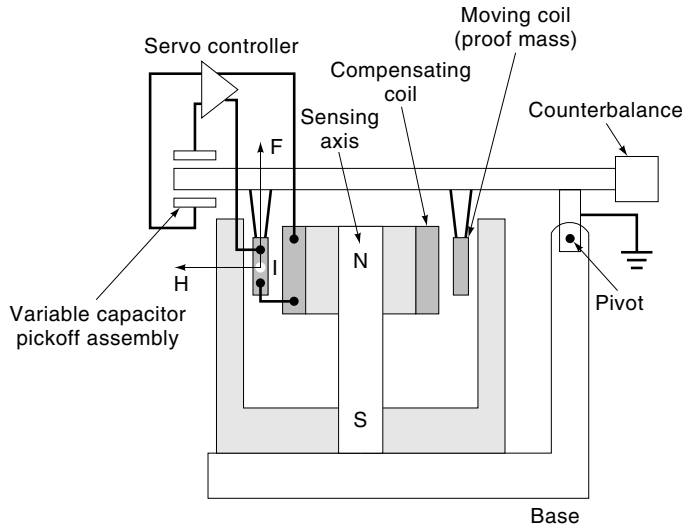


Figure 7. A basic coil and permanent magnet accelerometer. The coil is supported by an arm with minimum friction bearings to form a proof mass in a magnetic field. Displacement of the coil due to acceleration induces an electric potential in the coil to be sensed and processed. A servo system maintains the coil in a null position.

field, the magnetic field density, and the sine of the angle between the conductor and the field."

Figure 7 illustrates one form of accelerometer making use of this principle. The coil is located within the cylindrical gap defined by a permanent magnet and a cylindrical soft iron flux return path. It is mounted by means of an arm situated on a minimum friction bearing or flexure so as to constitute an acceleration-sensitive seismic mass. A pickoff mechanism senses the displacement of the coil under acceleration and causes the coil to be supplied with a direct current via a suitable servo controller to restore or maintain a null condition.

Assuming a downward acceleration with the field being radial (90°), by using Ampere's law the force experienced by the coil may be written as

$$F = ma = i l B \quad (10)$$

or the current

$$i = ma / l B \quad (11)$$

where B is the effective flux density and l is the total effective length of the conductor in the magnetic field.

Current in the restoring circuit is linearly proportional to acceleration, provided (1) armature reaction effects are negligible and fully neutralized by a compensating coil in opposition to the moving coil, and (2) the gain of the servo system is large enough to prevent displacement of the coil from the region in which the magnetic field is constant.

In these accelerometers, the magnetic structure must be shielded adequately to make the system insensitive to external disturbances or the earth's magnetic field. Also, in the presence of acceleration there will be a temperature rise due to $i^2 R$ losses. The effect of these $i^2 R$ losses on the performance are determined by the thermal design and heat-transfer properties of the accelerometers. In many applications, special care must be

exercised in choosing the appropriate accelerometer such that the temperature rises caused by unexpected accelerations cannot affect the scale factors or the bias conditions excessively. In others, digital signal conditioning can be used to produce a constant temperature rise after an initial transient at the turn on.

A simplified version of another servo accelerometer is given in Fig. 8. The acceleration a of the instrument case causes an inertial force f on the sensitive mass m , tending to make it pivot in its bearings or flexure mount. The rotation θ from the neutral position is sensed by an inductive pickup coil and amplified, demodulated, and filtered to produce a current i_a directly proportional to the motion from the null position. This current is passed through a precision stable resistor R to produce the output voltage signal and is applied to a coil suspended in a magnetic field. The current through the coil produces magnetic torque on the coil, which takes action to return the mass to the neutral position. The current required to produce magnetic torque that just balances the inertial torque due to acceleration is directly proportional to the acceleration a . Therefore the output voltage e_o becomes a measure of acceleration a . Since a nonzero displacement θ is necessary to produce the current i_a , the mass is not exactly returned the null position but becomes very close to zero because of the high-gain amplifier. Analysis of the block diagram reveals that

$$e_o / R = (m r a - e_o K_c / R) (K_p K_a / K_s) / (s^2 / \omega_{nl}^2 + 2 \zeta_1 s / \omega_{nl} + 1) \quad (12)$$

where K_c , K_p , K_a , and K_s are constants. Rearranging this expression gives

$$m r R K_p K_a a / K_s = (s^2 / \omega_{nl}^2 + 2 \zeta_1 s / \omega_{nl} + 1 + K_c K_p K_a a / K_s) e_o \quad (13)$$

By designing the amplifier gain K_a to be large enough so that $K_c K_p K_a a / K_s \gg 1.0$, then

$$e_o / a(s) = K / (s^2 / \omega_{nl}^2 + 2 \zeta_1 s / \omega_{nl} + 1 + K_c K_p K_a a / K_s) e_o \quad (14)$$

where

$$K \cong m r R / K_c \quad \text{V/(m/s}^2\text{)} \quad (15)$$

$$\omega_n \cong \omega_{nl} \sqrt{K_c K_p K_a / K_s} \quad \text{rad/s} \quad (16)$$

$$\zeta \cong \zeta_1 / \sqrt{K_c K_p K_a / K_s} \quad (17)$$

Equation (15) shows that the sensitivity depends on the values of m , r , R , and K_c , all of which can be made constant. In this case, a high-gain feedback is useful in shifting the requirements for accuracy and stability from mechanical components to a selected few parameters for which the requirements can be met easily. As in all feedback systems, the gain cannot be made arbitrarily high because of dynamic instability; however, a sufficiently high gain can be achieved to obtain good performances. At very low frequencies, less than a few hertz, high gain can be used with no loss of stability, and modern integrated circuit (IC) amplifiers have static gains over one million. An excellent comprehensive treatment of this topic is given by Doebelin (2).

Induction Accelerometers. The cross-product relationship of current, magnetic field, and force gives the basis for induction-type electromagnetic accelerometers, which are essentially

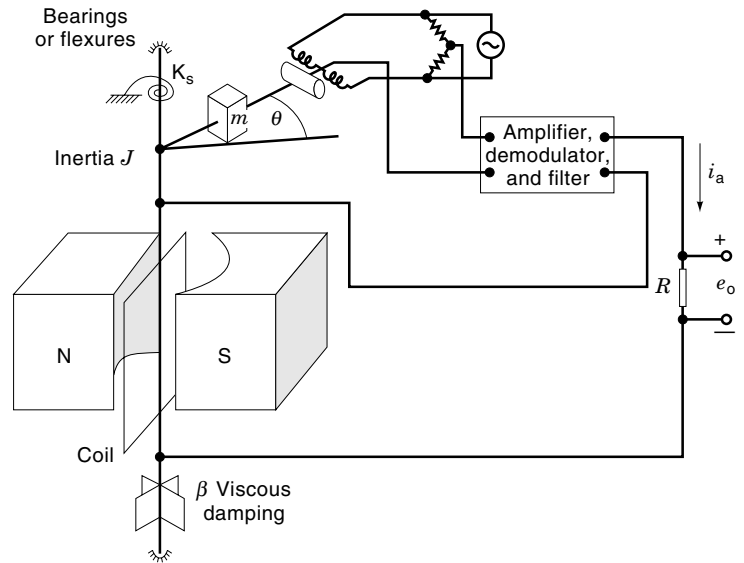


Figure 8. A simplified version of a rotational-type servo accelerometer. Acceleration of the instrument case causes an inertial force on the sensitive mass, tending to make it pivot in its bearings or flexure mount. The rotation from neutral is sensed by an inductive sensing apparatus and amplified and demodulated, and then filtered to produce a current directly proportional to the motion from the null position. The block diagram representation is useful in analysis.

generators rather than motors. One type of instrument, the cup-and-magnet design, includes a pendulous element with a pickoff mechanism and a servo controller driving a tachometer coupling. A permanent magnet and a flux return ring, closely spaced with respect to an electrically conductive cylinder, are attached to the pendulus element. A rate-proportional drag force is obtained by electromagnetic induction effects between the magnet and conductor. The pickoff mechanism senses pendulum deflection under acceleration and causes the servo controller to turn the rotor in a sense to drag the pendulus element toward the null position. Under steady-state conditions motor speed is a measure of the acceleration acting on the instrument. Stable servo operation is achieved by employing a time-lead network to compensate the inertial time lag of the motor and magnet combination. The accuracy of the servo-type accelerometers is ultimately limited by consistency and stability of scale factors of coupling and cup-and-magnet devices as a function of time and temperature. Since the angular rate is proportional to acceleration, angular position represents a change in velocity. This is a useful feature, particularly in navigation applications.

Another accelerometer based on induction design uses the eddy-current induction torque generation. The force-generating mechanism of an induction accelerometer consists of a stable magnetic field, usually supplied by a permanent magnet, which penetrates orthogonally through a uniform conduction sheet. The movement of the conducting sheet relative to the magnetic field in response to an acceleration results in

a generated electromotive potential in each circuit in the conductor. This action is in accordance with the law of Faraday's principle. In induction-type accelerometers, the induced eddy currents are confined to the conductor sheet, making the system essentially a drag coupling.

A typical commercial instrument based on the servo-accelerometer principle might have a micromachined quartz flexure suspension, differential capacitance angle pickoff, air squeeze film plus servo lead compensation for system damping. Of the various available models, a 30 *g* range unit has a threshold and resolution of 1 μ *g*, a frequency response that is flat to within 0.05% at 10 Hz and 2% at 100 Hz, a natural frequency of 1,500 Hz, a damping ratio from 0.3 to 0.8, and transverse or cross-axis sensitivity of 0.1%. If, for example, the output current is about 1.3 mA/*g*, a 250 Ω readout resistor would give about ± 10 V full scale at 30 *g*. These accelerometers are good for precision work and used in many applications such as aircraft and missile control systems, measurement of tilt angles for borehole navigation, and axle angular bending in aircraft weight and balance systems.

Piezoresistive Accelerometers

Piezoresistive accelerometers are essentially semiconductor strain gauges with large gauge factors. High gauge factors are obtained since the material resistivity is dependent primarily on the stress, not only on dimensions. This effect can

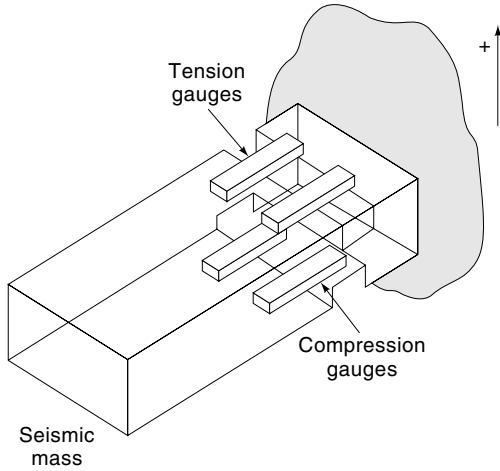


Figure 9. Bonding of piezoelectric and piezoresistive elements on to inertial system. As the inertial member vibrates, deformation of the tension and compression gauges causes resistance to change. The change in resistance is picked up and processed further. Accelerometers based on PZTs are particularly useful in medium- to high-frequency applications.

be greatly enhanced by appropriate doping of semiconductors such as silicon. The increased sensitivity is critical for vibration measurements since it allows miniaturization of the accelerometer. Most piezoresistive accelerometers use two or four active gauges arranged in a Wheatstone bridge. Extra precision resistors are used, as part of the circuit, in series with the input to control the sensitivity, balancing and offsetting temperature effects. The mechanical construction of a piezoresistive accelerometer is shown in Fig. 9.

In some applications, overload stops are necessary to protect the gauges from high-amplitude inputs. These instruments are useful for acquiring vibration information at low frequencies, for example, below 1 Hz. In fact, the piezoresistive sensors are inherently true static acceleration measurement devices. Typical characteristics of piezoresistive accelerometers may be 100 mV/g as the sensitivity, 0 Hz to 750 Hz as the frequency range, 2,500 Hz as the resonance frequency, 25 g as the amplitude range, 2,000 g as the shock rating, and 0°C to 95°C as the temperature range, with a total mass of about 25 g.

Differential-Capacitance Accelerometers

Differential-capacitance accelerometers are based on the principle of the change of capacitance in proportion to applied acceleration. They come in different shapes and sizes. In one type, the seismic mass of the accelerometer is made as the movable element of an electrical oscillator as shown in Fig. 10. The seismic mass is supported by a resilient parallel-motion beam arrangement from the base. The system is set to have a certain defined nominal frequency when undisturbed. If the instrument is accelerated the frequency varies about and below the nominal value depending on the direction of acceleration.

The seismic mass carries an electrode located in opposition to a number of base-fixed electrodes that defined variable capacitors. The base-fixed electrodes are resistance coupled in the feedback path of a wideband, phase-inverting amplifier. The gain of the amplifier is predetermined to ensure maintenance of oscillations over the range of variation of the capacitance

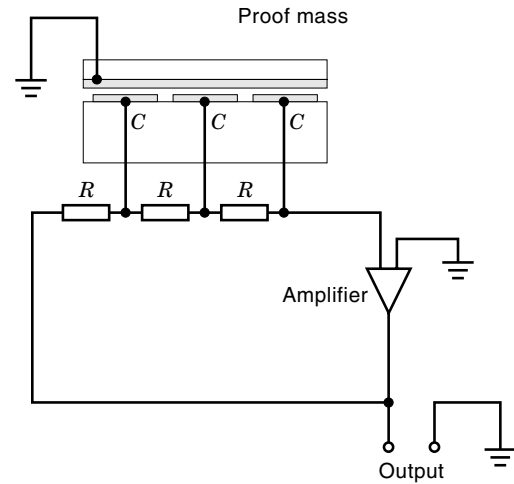


Figure 10. A typical differential-capacitance accelerometer. The proof mass is constrained in its null position by a spring. Under acceleration, variable frequencies are obtained in the electrical circuit. In a slightly different version the proof mass may be constrained by an electrostatic feedback force, thus resulting in a convenient mechanical simplicity.

determined by the applied acceleration. The value of the capacitance C for each of the variable capacitors is given by

$$C = \epsilon k S / h \quad (18)$$

where k is the dielectric constant, ϵ is the capacitivity of free space, S is the area of electrode, and h is the variable gap.

Denoting the magnitude of the gap for zero acceleration as h_0 , the value of h in the presence of acceleration a may be written

$$h = h_0 + ma/K \quad (19)$$

where m is the value of the proof mass and K is the spring constant. Thus,

$$C = \epsilon k S / (h_0 + ma/K) \quad (20)$$

For example, the frequency of oscillation of the resistance-capacitance type circuit is given by the expression

$$f = \sqrt{6}/2\pi RC \quad (21)$$

Substituting this value of C in Eq. (20) gives

$$f = (h_0 + ma/K) \sqrt{6}/2\pi R \epsilon k S \quad (22)$$

Denoting the constant quantity $\sqrt{6}/2\pi R \epsilon k S$ as B and rewriting Eq. (22) gives

$$f = B h_0 + B m a / K \quad (23)$$

The first term on the right-hand side expresses the fixed bias frequency f_0 and the second term denotes the change in frequency resulting from acceleration, so that the expression may

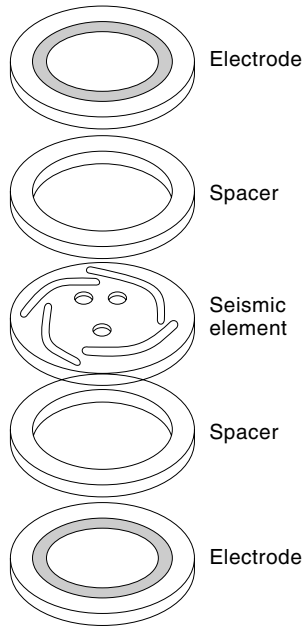


Figure 11. A diaphragm-type capacitive accelerometer. The seismic element is cushioned between the electrodes. Motion of the mass between the electrodes causes air movement passing through the holes, which provides a squeeze film damping. In some cases oil may be used as the damping element.

be written as

$$f = f_0 + f_a \quad (24)$$

If the output frequency is compared with an independent source of a constant frequency of f_0 whereby f_a may be determined.

A commonly used example of a capacitive-type accelerometer is based on a thin diaphragm with spiral flexures that provide the spring, proof mass, and moving plate necessary for the differential capacitor, as shown in Fig. 11. Plate motion between the electrodes pumps air parallel to the plate surface and through holes in the plate to provide squeeze film damping. Since air viscosity is less temperature sensitive than oil, the desired damping ratio of 0.7 hardly changes more than 15%. A family of such instruments are easily available with full-scale ranges from ± 0.2 g (4 Hz flat response) to $\pm 1,000$ g (3,000 Hz), a cross-axis sensitivity less than 1%, and a full-scale output of ± 1.5 V. The size of a typical device is about 25 mm^3 with a mass of 50 g.

Strain-Gauge Accelerometers

Strain-gauge accelerometers are based on resistance properties of electrical conductors. If a conductor is stretched or compressed, its resistance alters because of dimensional changes and the changes in the fundamental property of material called piezoresistance. This indicates that the resistivity ρ of the conductor depends on the mechanical strain applied onto it. The dependence is expressed as the gauge factor

$$(dR/R)/(dL/L) = 1 + 2\nu + (d\rho/\rho)/(dL/L) \quad (25)$$

where 1 indicates the resistance change due to length, 2ν in-

dicates resistance change due to area, and $(d\rho/\rho)/(dL/L)$ indicates the resistance change due to piezoresistivity.

In acceleration measurements, the resistance strain gauges can be selected from different types, such as unbonded metal-wire gauges, bonded metal-wire gauges, bonded metal-foil gauges, vacuum-deposited thin-metal-film gauges, bonded semiconductor gauges, and diffused semiconductor gauges. However, bonded and unbonded metal-wire gauges usually find wider applications in accelerometers. Occasionally, bonded semiconductor gauges, known as piezoresistive transducers, are used, but they suffer from high temperature sensitivities, nonlinearities, and some mounting difficulties. Nevertheless, in recent years, they have found new applications with the development of micromachine transducer technology, which is discussed in detail in the microaccelerometer section.

Unbonded-strain-gauge accelerometers use the strain wires as the spring element and as the motion transducer, using similar arrangements as in Fig. 9. They are useful for general-purpose motion and vibration measurements from low to medium frequencies. They are available in wide ranges and characteristics: typically ± 5 g to ± 200 g full scale, a natural frequency of 17 Hz to 800 Hz, 10 V excitation voltage ac or dc, full-scale output ± 20 mV to ± 50 mV, a resolution less than 0.1%, an inaccuracy less than 1% full scale, and a cross-axis sensitivity less than 2%. The damping ratio (using silicone oil damping) is 0.6 to 0.8 at room temperature. These instruments are small and light, usually with a mass less than 25 g.

Bonded-strain-gauge accelerometers generally use a mass supported by a thin flexure beam. The strain gauges are cemented onto the beam to achieve maximum sensitivity, temperature compensation, and sensitivity to both cross-axis and angular accelerations. Their characteristics are similar to unbonded-strain-gauge accelerometers but have greater sizes and weights. Often silicone oil is used for damping. Semiconductor strain gauges are widely used as strain sensors in cantilever-beam and mass types of accelerometers. They allow high outputs (0.2 V to 0.5 V full scale). Typically, a ± 25 g acceleration unit has a flat response from 0 Hz to 750 Hz, a damping ratio of 0.7, a mass of 28 g, and an operational temperature of -18°C to $\pm 93^\circ\text{C}$. A triaxial $\pm 20,000$ g model has a flat response from 0 kHz to 15 kHz, a damping ratio of 0.01, and a compensation temperature range of 0°C to 45°C and is $13 \times 10 \times 13 \text{ mm}^3$ in size and 10 g in mass.

Inertial Types: Cantilever and Suspended-Mass Accelerometers

There are a number of different inertial-type accelerometers, most of which are in development stages or used under very special circumstances, such as gyropendulum, reaction-rotor, vibrating-string, and centrifugal-force-balance designs. In many types, the force required to constrain the mass in the presence of the acceleration is supplied by an inertial system.

The vibrating-string instrument, Fig. 12, makes use of a proof mass supported longitudinally by a pair of tensioned, transversely vibrating strings with uniform cross section and equal lengths and masses. The frequency of vibration of the strings is set to several thousand cycles per second. The proof mass is supported radially in such a way that the acceleration normal to the strings does not affect the string tension. In the presence of acceleration along the sensing axis, a differential tension exists on the two strings, thus altering the frequency of vibration. From the second law of motion the frequencies may

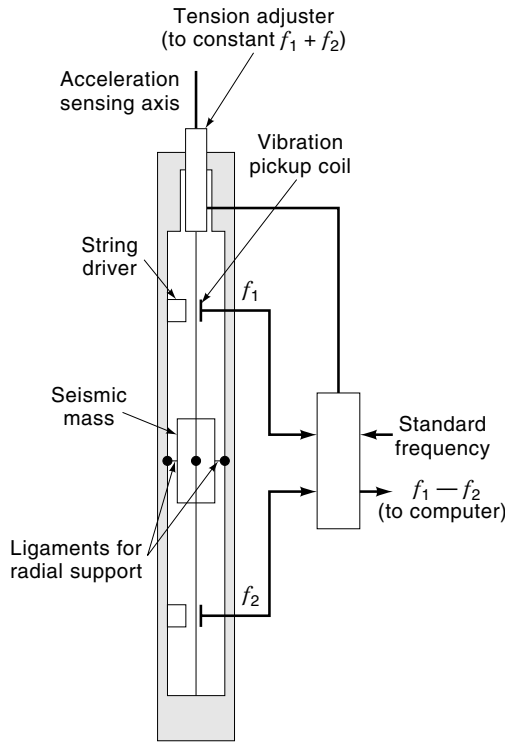


Figure 12. A vibrating-string accelerometer. A proof mass is attached to two strings of equal mass and length and it is supported radially by suitable bearings. The vibration frequencies of strings are dependent on the tension imposed by the acceleration of the system in the direction of the sensing axis.

be written as

$$f_1^2 = T_1/4m_s l, \quad f_2^2 = T_2/4m_s l \quad (26)$$

where T is the tension and m_s and l are the masses and lengths of strings, respectively.

The quantity $T_1 - T_2$ is proportional to ma where a is the acceleration along the axis of the strings. An expression for the difference of the frequency-squared terms may be written as

$$f_1^2 - f_2^2 = (T_1 - T_2)/4m_s l = ma/4m_s l \quad (27)$$

hence

$$f_1 - f_2 = ma/[(f_1 + f_2)4m_s l] \quad (28)$$

The sum of frequencies ($f_1 + f_2$) can be held constant by servoing the tension in the strings with reference to the frequency of a standard oscillator. Then, the difference between the frequencies becomes linearly proportional to acceleration. In some versions, the beamlike property of the vibratory elements is used by gripping them at nodal points corresponding to the fundamental mode of the vibration of the beam and at the respective centers of percussion of the common proof mass. The output frequency is proportional to acceleration, and the velocity is proportional to phase, thus offering an important advantage. The velocity change can be measured by something almost as simple as counting zero crossings. Improved versions of these devices lead to cantilever-type accelerometers, discussed next.

In this accelerometer, a small cantilever beam mounted on the block is placed against the vibrating surface, and an appropriate mechanism is provided for varying the beam length. The beam length is adjusted such that its natural frequency is equal to the frequency of the vibrating surface, and hence the resonance condition is obtained. Recently, slight variations of cantilever-beam arrangements are finding new applications in microaccelerometers.

In a different type of suspended-mass configuration, a pendulum is used that is pivoted to a shaft rotating about a vertical axis. Pickoff mechanisms are provided for the pendulum and the shaft speed. The system is servo controlled to maintain it at null position. Gravitational acceleration is balanced by the centrifugal acceleration. The shaft speed is proportional to the square root of the local value of the acceleration.

All inertial force accelerometers just described are absolute measure instruments. That is, their scale factors are predetermined solely by establishing mass, length, and time quantities, as distinguished from voltage and spring stiffness.

Electrostatic-Force-Feedback Accelerometers

Electrostatic accelerometers are based on Coulomb's law between two charged electrodes. They measure the voltage in terms of force required to sustain a movable electrode of a known area, mass, and separation from an affixed electrode. The field between the electrodes is given by

$$E = Q/\epsilon k S \quad (29)$$

where E is the intensity or potential gradient (dV/dx); Q is the charge, S is the area of the conductor, and k is the dielectric constant of the space outside the conductor.

By using this expression it can be shown that the force per unit area of the charged conductor (in N/m^2) is given by

$$F/S = Q^2/2\epsilon k S^2 = \epsilon k E^2/2 \quad (30)$$

In an electrostatic-force-feedback accelerometer (similar in structure as in Fig. 9) an electrode of mass m and area S is mounted on a light pivoted arm for moving relative to the fixed electrodes. The nominal gap h between the pivoted and fixed electrodes is maintained by means of a force-balancing servo system capable of varying in the electrode potential in response to signals from a pickoff mechanism that senses relative changes in the gaps.

Considering one movable electrode and one stationary electrode and assuming that the movable electrode is maintained at a bias potential V_1 and the stationary one at a potential V_2 . The electrical intensity E in the gap can be expressed as

$$E_1 = (V_1 - V_2)/h \quad (31)$$

so that the force of attraction may be found as

$$F_1 = \epsilon k E^2 S/2h^2 = \epsilon k (V_1 - V_2)^2 S/2h^2 \quad (32)$$

In the presence of acceleration, if V_2 is adjusted to restrain the movable electrode to the null position, the expression relating acceleration and electrical potential may be given by

$$a = F_1/m = \epsilon k (V_1 - V_2)^2 S/2h^2 m \quad (33)$$

The device so far described can measure acceleration in one direction only, and the output is quadratic, that is,

$$(V_1 - V_2) = D\sqrt{a} \quad (34)$$

where D is the constant of proportionality.

The output may be linearized in a number of ways, one of them being the quarter-square method. If the servo controller applies a potential $-V_2$ to the other fixed electrode, the force of attraction between this electrode and the movable electrode becomes

$$a = F_1/m = \epsilon k(V_1 + V_2)^2 S/2h^2 m \quad (35)$$

and the force-balance equation of the movable electrode when the instrument experiences a downward acceleration a now is

$$ma = F_1 - F_2 = [(V_1 + V_2)^2 - (V_1 - V_2)^2] \epsilon k S/2h^2 m$$

or

$$ma = \epsilon k S(4V_1 V_2)/2h^2 m \quad (36)$$

Hence, if the bias potential V_1 is held constant and the gain of the control loop is high so that variations in the gap are negligible, the acceleration becomes a linear function of the controller output voltage V_2 as

$$a = V_2[(\epsilon k S 2V_1)/h^2 m] \quad (37)$$

The principal difficulty in mechanizing the electrostatic force accelerometer is the relatively high electric field intensity required to obtain an adequate force. Also, extremely good bearings are necessary. Damping can be provided electrically or by viscosity of the gaseous atmosphere in the interelectrode space if the gap h is sufficiently small. The scheme works best in micromachined instruments. Nonlinearity in the voltage breakdown phenomenon permits larger gradients in very small gaps.

The main advantages of electrostatic accelerometers are extreme mechanical simplicity, low power requirements, absence of inherent sources of hysteresis errors, zero temperature coefficients, and ease of shielding from stray fields.

Microaccelerometers

By the end of the 1970s it became apparent that the essentially planar integrated-circuit (IC) technology processing could be modified to fabricate three-dimensional electromechanical structures, called micromachining. Accelerometers and pressure sensors were among the first IC sensors. The first accelerometer was developed in 1979. Since then the technology has been progressing steadily, and now an extremely diverse range of accelerometers is readily available. Most sensors use bulk micromachining rather than surface micromachining techniques. In bulk micromachining the flexures, resonant beams, and all other critical components of the accelerometer are made from bulk silicon in order to exploit the full mechanical properties of a single silicon crystal. With proper design and film process, bulk micromachining yields an extremely stable and robust accelerometer.

The selective etching of multiple layers of deposited thin films, or surface micromachining, allows movable microstructures to be fabricated on silicon wafers. With surface micromachining, layers of structure material are disposed and patterned

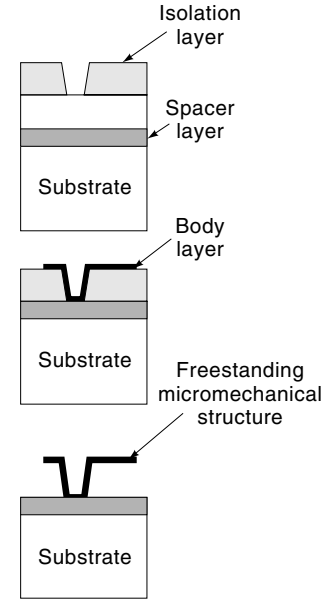


Figure 13. Steps of surface micromachining. The acceleration-sensitive three-dimensional structure is formed on a substrate and a sacrificial element. The sacrificial element is etched to leave a free-standing structure. The spacing between the structure and substrate is about $2 \mu\text{m}$.

as shown in Fig. 13. These structures are formed by polysilicon and a sacrificial material such as silicon dioxide. The sacrificial material acts as an intermediate spacer layer and is etched away to produce a freestanding structure. Surface machining technology also allows smaller and more complex structures to be built in multiple layers on a single substrate.

The operational principles of microaccelerometers are very similar to capacitive force-balance or vibrating-beam accelerometers, discussed earlier. Manufacturing techniques may change from one manufacturer to another. However, in general, vibrating-beam accelerometers are preferred because of better air-gap properties and improved bias performance characteristics.

Vibrating-beam accelerometers, also termed resonant beam force transducers, are made in such a way that an acceleration along a positive input axis places the vibrating beam in tension. Thus the resonant frequency of the vibrating beam increases or decreases with the applied acceleration. A common structure, mechanically coupled beam also known as a double-ended tuning fork (DETF) is shown in Fig. 14.

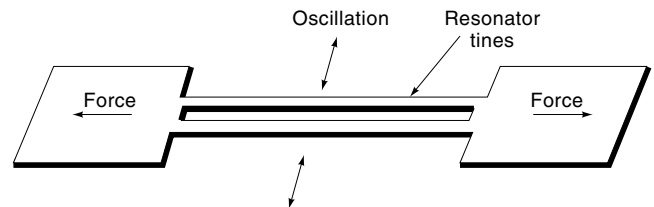


Figure 14. A double-ended tuning fork (DETF) acceleration transducer. Two beams are vibrated 180° out of phase to eliminate reaction forces at the beam ends. The resonant frequency of the beam is altered by acceleration. The signal-processing circuits are also integrated in the same chip.

In DETF, an electronic oscillator capacitively couples energy into two vibrating beams to keep them oscillating at their resonant frequency. The beams vibrate 180° out of phase to cancel reaction forces at the ends. The dynamic cancellation effect of the DETF design prevents energy from being lost through the ends of the beam. Hence, the dynamically balanced DETF resonator has a high Q factor, which leads to a stable oscillator circuit. The acceleration signal is output from the oscillator as a frequency-modulated square wave that can be used for a digital interface.

The vibrating beam accelerometer is similar in philosophy to the vibrating string accelerometer. Frequency output provides an easy interface with digital systems, and measurement of phase provides an easy integration to velocity. Static stiffness eliminates the tension and makes the device much smaller. A recent trend is to manufacture vibrating beam accelerometers as micromachined devices. With differential frequency arrangements, many common mode errors can be eliminated, including clock errors.

The frequency of resonance of the system must be much higher than any input acceleration, and this limits the measurable range. In one military micromachined accelerometer the following characteristics are given: a range of $\pm 1200\text{ g}$, a sensitivity of 1.11 Hz/g , a bandwidth of $2,500\text{ Hz}$, and unloaded DETF frequency of $9,952\text{ Hz}$, the frequency at $+1,200\text{ g}$ is $11,221\text{ Hz}$, the frequency at $-1,200\text{ g}$ is $8,544\text{ Hz}$, and the temperature sensitivity is $5\text{ mg/}^\circ\text{C}$. The accelerometer size is 6 mm diameter by 4.3 mm length, with a mass of about 9 g , and it has a turn-on time less than 60 s . The accelerometer is powered with $+9$ to $+16\text{ V dc}$ and the nominal output is $9,000\text{ Hz}$, square wave.

Surface micromachining has also been used to manufacture application-specific accelerometers, such as for air-bag applications in the automotive industry. In one type, a three-layer differential capacitor is created by alternate layers of polysilicon and phosphosilicate glass (PSG) on a 0.38 mm thick, 100 mm long wafer. A silicon wafer serves as the substrate for the mechanical structure. The trampoline-shaped middle layer is suspended by four supporting arms. This movable structure is the seismic mass for the accelerometer. The upper and lower polysilicon layers are fixed plates for the differential capacitors. The glass is sacrificially etched by hydrofluoric acid (HF).

CALIBRATIONS AND SENSITIVITY

Calibrations of accelerometers are necessary in acceleration, vibration, and shock sensing. The calibration methods can broadly be classified to be *static* or *dynamic* calibrations. The static calibration is conducted at one or several levels of constant accelerations. For example, if a tilting table calibration method is selected, the vertical component of the free fall is made use of without a choice of magnitude, but with the advantage of being readily available, and accurate to a few parts in 10^7 . On the other hand, if a centrifuge is selected, it produces a constant acceleration as a function of speed of rotation, and the magnitudes can be chosen in a wide range from 0 g to well over $50,000\text{ g}$, but with typical uncertainty of 1% in measuring the actual radius from the axis of rotation to the effective center of mass. The dynamic calibration is usually done by using an electrodynamic shaker.

The electrodynamic shaker is designed to oscillate in a sinusoidal motion with variable frequencies and amplitudes. They are stabilized at selected levels of calibration. This is an absolute method that consists of measuring the displacement with laser interferometry and a precise frequency meter for accurate frequency measurements. The shaker must be driven by a power amplifier thus giving a sinusoidal output with minimal distortion. The National Institute of Science and Technology uses this method as a reference standard. Precision accelerometers, mostly of the piezoelectric types, are calibrated by the absolute method and then used as the working standards. A preferred method is the back-to-back calibration, where the test specimen is directly mounted on the working standard, which in turn is mounted on an electrodynamic shaker.

Sensitivity

A vibrational structure may have been subjected to different forms of vibrations, such as compressional, torsional, or transverse. A combination of all these vibrations may also take place simultaneously, which makes analysis and measurement difficult and complex. It was discussed earlier that the differential equations governing the vibrational motion of a structure depends on the number of degrees of freedom, which are described as a function of space coordinates $f(x, y, z, t)$. For example, the transverse vibrations of structures may be a fourth-order equation differential equation.

Fortunately, most common acceleration and vibration measurements are simple in nature, being either compressional or torsional. They can easily be expressed as second-order differential equations, as explained in the frequency response section. However, during measurements, most accelerometers are affected by transverse vibrations and their sensitivity can play a major role in the accuracy of the measurements.

The transverse, also known as cross-axis, sensitivity of an accelerometer is its response to acceleration in a plane perpendicular to the main accelerometer axis as shown in Fig. 15.

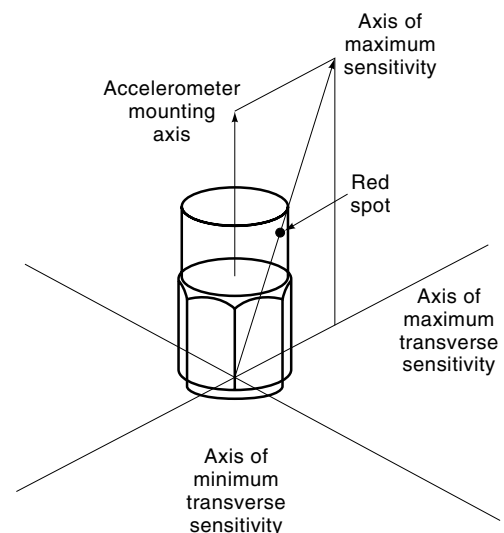


Figure 15. A vectorial illustration of cross-axis sensitivity. Accelerometers may sense vibrations not only in the direction of the main axis but also perpendicular to the main axis. These cross-axis responses are minimized in many accelerometers to a value less than 1% . Sometimes this sensitivity may be used to determine the correct orientation of the device.

The cross-axis sensitivity is normally expressed as a percentage of the main-axis sensitivity and should be as low as possible. There is no single value of cross-axis sensitivity, but it varies depending on the direction. The direction of minimum sensitivity is usually supplied by the manufacturers.

The measurement of the maximum cross-axis sensitivity is part of the individual calibration procedure and should always be less than 3% to 4%. If high levels of transverse vibration are present, this may result in erroneous overall results. In this case, separate arrangements should be made to establish the level and frequency contents of the cross-axis vibrations. The cross-axis sensitivity of typical accelerometers mentioned in the relevant sections were (2% to 3% for piezoelectric types and less than 1% for most others). In force-feedback accelerometers the transverse sensitivity is proportional to the input axis misalignment; therefore, it may be calibrated as a function of the temperature to within several microradians.

APPLICATIONS

This section is concerned with applications of accelerometers to measure physical properties such as acceleration, vibration and shock, and motions associated with inertial navigation. A full understanding of accelerometer dynamics is necessary in relation to characteristics of acceleration, vibration, and shock. The vibrations can be periodic, stationary random, nonstationary random, or transient.

Periodic Vibrations. In periodic vibrations, the motion of an object repeats itself in an oscillatory manner. This can be represented by a sinusoidal waveform

$$x(t) = X_{\text{peak}} \sin(\omega t) \quad (38)$$

where $x(t)$ is the time-dependent displacement, $\omega = 2\pi ft$ is the angular frequency, and X_{peak} is the maximum displacement from a reference point.

The velocity of the object is the time rate of change of displacement

$$u(t) = dx/dt = \omega X_{\text{peak}} \cos(\omega t) = U_{\text{peak}} \sin(\omega t + \pi/2) \quad (39)$$

where $u(t)$ is the time-dependent velocity, and $U_{\text{peak}} = \omega X_{\text{peak}}$ is the maximum velocity.

The acceleration of the object is the time rate change of velocity

$$a(t) = du/dt = d^2u/dt^2 = -\omega^2 X_{\text{peak}} \sin(\omega t) = A_{\text{peak}} \sin(\omega t + \pi) \quad (40)$$

where $a(t)$ is the time-dependent acceleration, and $A_{\text{peak}} = \omega^2 X_{\text{peak}} = \omega U_{\text{peak}}$ is the maximum acceleration.

From the preceding equations it can be seen that the basic form and the period of vibration remains the same in acceleration, velocity, and displacement. But velocity leads displacement by a phase angle of 90° and acceleration leads velocity by another 90° . The amplitudes of the three quantities are related as a function of frequency, shown in Fig. 16.

In nature, vibrations can be periodic but not necessarily sinusoidal. If they are periodic but nonsinusoidal, they can be

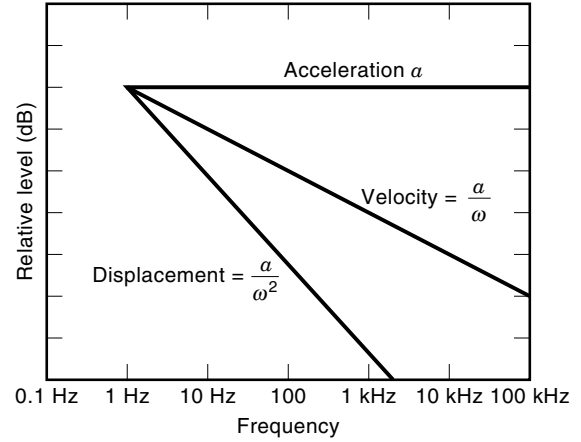


Figure 16. The logarithmic relationship between acceleration, velocity, and displacement. Velocity at a particular frequency can be obtained by dividing acceleration by a factor proportional to frequency. For displacement acceleration must be divided by a factor proportional to square of the frequency. Phase angles need to be determined separately, but they can be neglected in time-average measurements.

expressed as a combination of a number of pure sinusoidal curves, describes by Fourier analysis as

$$x(t) = X_0 + X_1 \sin(\omega_1 t + \phi_1) + X_2 \sin(\omega_2 t + \phi_2) + \dots + X_n \sin(\omega_n t + \phi_n) \quad (41)$$

where $\omega_1, \omega_2, \dots, \omega_n$ are frequencies (rad/s), $X_0, X_1, \dots, X_n$ are maximum amplitudes of respective frequencies, and $\phi_1, \phi_2, \dots, \phi_n$ are phase angles.

The number of terms may be infinite, and the higher the number of elements better the approximation. These elements constitute the *frequency spectrum*. The vibrations can be represented in the time domain or frequency domain, both of which are extremely useful in the analysis. As an example, in Fig. 17, the time response of the seismic mass of an accelerometer is given against a rectangular pattern of excitation of the base.

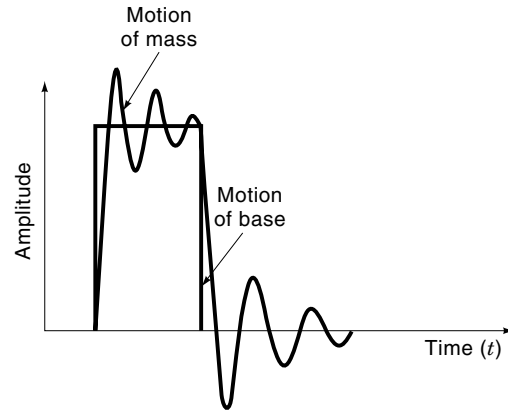


Figure 17. The time response of a shock excitation of a single-degree-of-freedom system. As the duration of the shock pulse increases sustained oscillations get shorter in time but larger in amplitude. The maximum system response may be as high as twice the magnitude of the shock pulse.

Stationary Random Vibrations. Random vibrations occur often in nature, where they constitute irregular cycles of motion that never repeat themselves exactly. Theoretically, an infinitely long time record is necessary to obtain a complete description of these vibrations. However, statistical methods and probability theory can be used for the analysis by taking representative samples. Mathematical tools such as probability distributions, probability densities, frequency spectra, cross-correlations, auto correlations, digital Fourier transforms (DFTs), fast Fourier transforms (FFTs), auto-spectral-analysis, rms values, and digital filter analysis are some of the techniques that can be employed. Interested readers should refer to references for further information.

Transients and Shocks. Often, short duration and sudden occurrence vibrations need to be measured. Shock and transient vibrations may be described in terms of force, acceleration, velocity, or displacement. As in the case of random transients and shocks, statistical methods and Fourier transforms are used in the analysis.

Nonstationary Random Vibrations. In this case, the statistical properties of vibrations vary in time. Methods such as time averaging and other statistical techniques can be employed.

The majority of accelerometers described here can be viewed and analyzed as seismic instruments consisting of a mass, a spring, and a damper arrangement as shown in Fig. 1. Taking only the mass-spring system, if the system behaves linearly in a time invariant manner, the basic second-order differential equation for the motion of the mass alone under the influence of a force can be written as

$$f(t) = m d^2 x / dt^2 + c dx / dt + kx \quad (42)$$

where $f(t)$ is the force, m the mass, c the velocity constant, and k the spring constant.

Nevertheless, in seismic accelerometers the base of the arrangement is in motion too. Therefore, Eq. (42) may be generalized by taking the effect motion of the base into account. Then this equation may be modified as

$$m d^2 z / dt^2 + c dz / dt + kz = mg \cos(\theta) - m d^2 x_1 / dt^2 \quad (43)$$

where $z = x_2 - x_1$ is the relative motion between the mass and the base, x_1 the displacement of the base, x_2 the displacement of the mass, and θ the angle between the sense axis and gravity.

In order to lay a background for further analysis, taking the simple case, the complete solution to Eq. (42) can be obtained by applying the superposition principle. The superposition principle states that if there are simultaneously superimposed actions on a body, the total effect can be obtained by summing the effects of each individual action.

Equation (42) describes essentially a second-order system that can be expressed through Laplace transform as

$$X(s)/F(s) = 1/ms^2 + cs + k \quad (44)$$

or

$$X(s)/F(s) = K/(s^2/\omega_n^2 + 2\zeta s/\omega_n + 1) \quad (45)$$

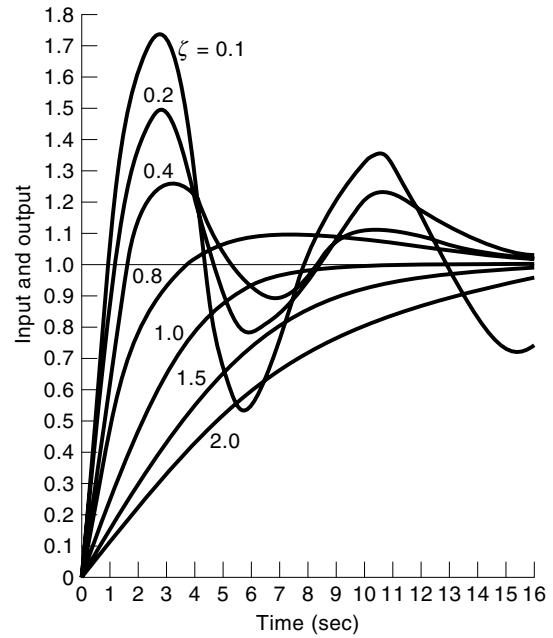


Figure 18. Unit-step-time responses of a second-order system with various damping ratios. The maximum overshoot, delay, rise, settling times, and frequency of oscillations depend on the damping ratio. A smaller damping ratio gives a faster response but larger overshoot. In many applications a damping ratio of 0.707 is preferred.

where s is the Laplace operator, $K = l/k$ is the static sensitivity, $\omega_n = \sqrt{k/m}$ is the undamped critical frequency (rad/s), and $\zeta = (c/2)\sqrt{k/m}$ is the damping ratio.

As can be seen in the performance of accelerometers, the important parameters are the static sensitivity, the natural frequency, and the damping ratio, which are functions of mass, velocity, and spring constants, respectively. Accelerometers are designed to have different characteristics by suitable selection of these parameters.

Once the response is expressed in the form of Eqs. (44) and (45), analysis can be taken further either in the time domain or in the frequency domain. The time response of a typical second-order system for a unit-step input is given in Fig. 18. The Bode plot gain phase responses are depicted in Fig. 19. Detailed discussions about frequency response, damping, damping ratio, and linearity are made in relevant sections, and further information can be obtained in the references.

Systems in which a single structure moves in more than one direction are termed multi-degree-of-freedom systems. In this case, the accelerations become functions of dimensions as d^2x/dt^2 , d^2y/dt^2 , and d^2z/dt^2 . Hence, in multichannel vibration tests multiple transducers must be used to create uniaxial, biaxial, or triaxial sensing points for measurements. Mathematically, a linear multi-degree-of-freedom system can be described by a set of coupled second-order linear differential equations, and when the frequency response is plotted it normally shows one resonance peak per degree of freedom.

Frequently, acceleration and vibration measurements of thin plates or small masses are required. Attaching an accelerometer with a comparable mass onto a thin plate or a small test piece can cause *mass loading*. Since acceleration is dependent on the mass, the vibration characteristics of the loaded test piece may be altered, thus yielding wrong measurements.

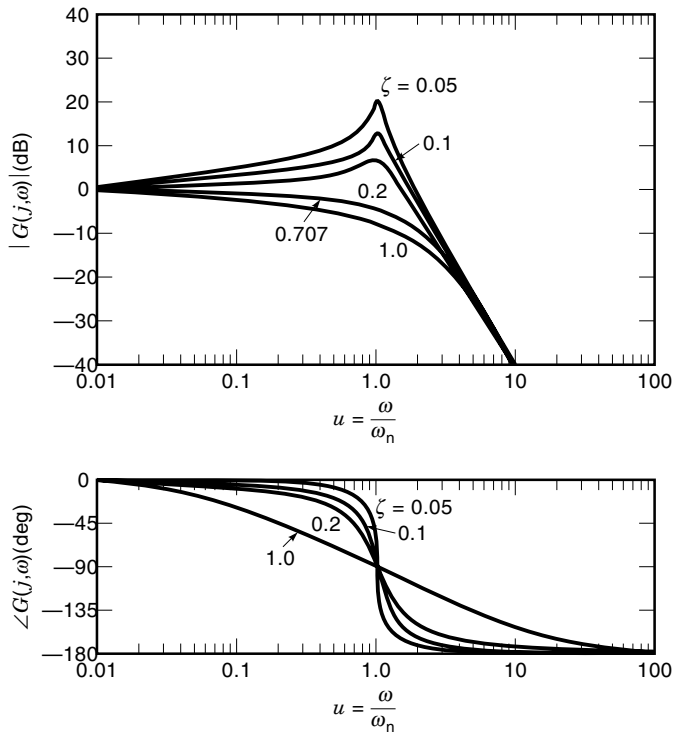


Figure 19. Bode plot of the gain and phase angle of the transfer function $G(j\omega)$ of a second-order system against frequency. Curves are functions of frequencies as well as damping ratios. These plots can be obtained theoretically or by practical tests conducted in frequency range.

In such cases, a correct interpretation of the results of the measuring instruments must be made. Some experimental techniques are also available for the correction of the test results in the form performing repetitive tests conducted by sequentially adding small known masses and by observing the differences.

In general, accelerometers are preferred over other displacement and velocity sensors due to the following reasons:

1. They have a wide frequency range from zero to very high values. Steady accelerations can easily be measured.
2. Acceleration is more frequently needed since destructive forces are often related to acceleration rather than to velocity or displacement.
3. Measurement of transients and shocks can readily be made, which is much easier than displacement of velocity sensing.
4. Displacement and velocity can be obtained by simple integration of acceleration by electronic circuitry. Integration is preferred over differentiation.

Selection, Full-Scale Range, and Overload Capability

Ultimate care must be exercised for the selection of correct accelerometer to meet the requirements of a particular application. At first glance there may seem to be a confusingly large selection of accelerometers available. But they can be classified in two main groups. The first group is the general-purpose accelerometers offered in various sensitivities, frequencies, and full scale and overload ranges, with different mechanical and

electrical connection options. The second group of accelerometers have characteristics targeted toward a particular application. A list of manufacturers of accelerometers is supplied in Table 1.

In deciding about the application type, for example, general purpose or special, and the accelerometer to be employed, the following characteristics need to be considered: the transient response or cross-axis sensitivity; frequency range, sensitivity, mass and dynamic range; cross-axis response; and environmental conditions, temperature, cable noise, stability of bias, scale factor, and misalignment. Some useful hints about these characteristics will be given below.

The Frequency Range. Acceleration measurements are normally confined to using the linear portion of the response curve. The response is limited at low frequencies as well as at high frequencies by the natural resonances. As a rule of thumb the upper-frequency limit for the measurement can be set to one-third of the accelerometer's resonance frequency such that the vibrations measured will be less than 1 dB in linearity. It should be noted that an accelerometer's useful frequency range is significantly higher, that is, to one-half or two-thirds of its resonant frequency. The measurement frequencies may be set to higher values in applications in which lower linearity (say 3 dB) may be acceptable as in the case of monitoring of internal conditions of machines since the repeatability is more important than the linearity. The lower measuring frequency limit is determined by two factors. The first is the low-frequency cutoff of the associated preamplifiers. The second is the effect of ambient temperature fluctuations to which the accelerometer may be sensitive.

The Sensitivity, Mass, and Dynamic Range. Ideally, a higher transducer sensitivity is better, but compromises may have to be made for sensitivity versus frequency, range, overload capacity, size, and so on.

Accelerometer mass becomes important when using small and light test objects. The accelerometer should not load the structural member, since additional mass can significantly change the levels and frequency presence at measuring points and invalidate the results. As a general rule, the accelerometer mass should not be greater than one-tenth the effective mass of the part or the structure that it is mounted onto for measurements.

The dynamic range of the accelerometer should match the high or low acceleration levels of the measured objects. General-purpose accelerometers can be linear up to 5,000 g to 10,000 g , which is well into the range of most mechanical shocks. Special accelerometers can measure up to 100,000 g .

An important point in the practical application of accelerometers is that if mechanical damping is a problem, air damping is preferable to oil damping, since oil damping is extremely sensitive to viscosity changes. If the elements are stable against temperature, electronic damping may be sufficient.

The Transient Response. Shocks are characterized as sudden releases of energy in the form of short-duration pulses exhibiting various shapes and rise times. They have high magnitudes and wide frequency contents. In the applications where transient and shock measurements are involved, the overall linearity of the measuring system may be limited to high and low frequencies by phenomena known as zero shift and ringing,

Table 1. List of Manufacturers

Allied Signal, Inc. 101 Colombia Rd. Dept. CAC Morristown, NJ 07962 Tel.: 602-496-1000 or 800-707-4555 Fax: 602-496-1001	Instrumented Sensor Technologies 4701 A Moor Street Okemos, MI 48864 Tel.: 517-349-8487 Fax: 517-349-8469	Rutherford Controls 2697 International Pkwy Building #3, Suite 122 Virginia Beach, VA 23452 Tel.: 800-899-5625 Fax: 804-427-9549
Bokam Engineering, Inc. 9552 Smoke Tree Avenue Fountain Valley, CA 92708 Tel.: 714-962-3121 Fax: 714-962-5002	Jewel Electrical Instruments 124 Joliette Street Manchester, NH 03102 Tel.: 603-669-6400 or 800-227-5955 Fax: 603-669-6962	Sensotech, Inc. 1202 Chesapeake Ave. Columbus, OH 43212 Tel.: 614-486-7723 or 800-867-3890 Fax: 614-486-0506
CEC Vibration Division of Sontortronics 196 University Parkway Pomona, CA 91768 Tel.: 909-468-1345 or 800-468-1345 Fax: 909-468-1346	Kistler Instrument Co. 75 John Glenn Dr. Amherst, NY 14228-2171 Tel.: 800-755-5745	SETRA 45 Nagog Park Acton, MA 01720 Tel.: 508-263-1400 or 800-257-3872 Fax: 508-264-0292
Dytran Instrument, Inc. Dynamic Transducers and Systems 21592 Marilla Street, Chatsworth, CA 91311 Tel.: 800-899-7818 Fax: 800-899-7088	Lucas Control Products, Inc. 1000 Lucas Way Hampton, VA 23666 Tel.: 800-745-8008 Fax: 800-745-8004	Silicon Microstructures, Inc. 46725 Fremond Blvd. Fremond, CA 94358 Tel.: 510-490-5010 Fax: 510-490-1119
ENDEVCO 30700 Rancho Viejo Road San Juan Capistrano, CA 92675 Tel.: 800-289-8204 Fax: 714-661-7231	Metrix Instrument Co. 1711 Townhurst Houston, TX 77043 Fax: 713-461-8223	SKF Condition Monitoring 4141 Ruffin Road San Diego, CA 92123 Tel.: 800-959-1366 Fax: 619-496-3531
Entran Devices, Inc. 10-T Washington Ave. Fairfield, NJ 07004 Tel.: 800-635-0650	Patriot Sensors and Controls Corporation 650 Easy Street Simi Valley, CA 93065 Tel.: 805-581-3985 or 800-581-0701 Fax: 805-583-1526	Summit Instruments, Inc. 2236 N. Cleveland-Massillon Rd. Akron, Ohio 44333-1255 Tel.: 800-291-3730 Fax: 216-659-3286
First Inertia Switch G-10386 N. Holly Rd. Dept. 10, P.O. Box 704 Grand Blanc, MI 48439 Tel.: 810-695-8333 or 800-543-0081 Fax: 810-695-0589	PCB Piezoelectronics, Inc. 3425 Walden Avenue Depew, NY 14043 Tel.: 716-684-0001 Fax: 716-684-0987	Wilcoxon Research 21-T Firstfield Road Gaithersburg, MD 20878 Tel.: 800-842-7367 Fax: 301-330-8873
	PMC/BETA 9 Tek Circle Natick, MA 01760 Tel.: 617-237-6020 Fax: 508-651-9762	

respectively. The zero shift is caused by both the phase nonlinearity in the preamplifiers and the accelerometer not returning to steady-state operation conditions after being subjected to high shocks. Ringing is caused by high-frequency components of the excitation near-resonance frequency preventing the accelerometer to return back to its steady-state operation condition. To avoid measuring errors due to these effects the operational frequency of the measuring system should be limited to the linear range.

Full-Scale Range and Overload Capability. Most accelerometers are able to measure acceleration in both positive and negative directions. They are also designed to be able to accommodate overload capacity. Appropriate discussions are made on full scale range and overload capacity of accelerometers in the appropriate sections. Manufacturers also supply information on these two characteristics.

Environmental Conditions. In selection and implementation of accelerometers, environmental conditions such as temperature ranges, temperature transients, cable noise, magnetic field effects, humidity, and acoustic noise need to be considered.

Manufacturers supply information on environmental conditions of use.

SIGNAL CONDITIONING

Common signal conditioners are appropriate for interfacing accelerometers to computers or other instruments for further signal processing. Caution needs to be exercised to provide the appropriate electric load to self-generating accelerometers. Generally, the generated raw signals are amplified and filtered suitably by the circuits within the accelerometer casing supplied by manufacturers. Nevertheless, piezoelectric and piezoresistive transducers require special signal conditioners with certain characteristics, which is discussed in the following section. Examples of signal conditioning circuits will also be given for the microaccelerometers.

Signal Conditioning Piezoelectric Accelerometers. The piezoelectric accelerometer supplies a very small energy to the signal conditioner. It has a high capacitive source impedance. The equivalent circuit of a piezoelectric accelerometer can be

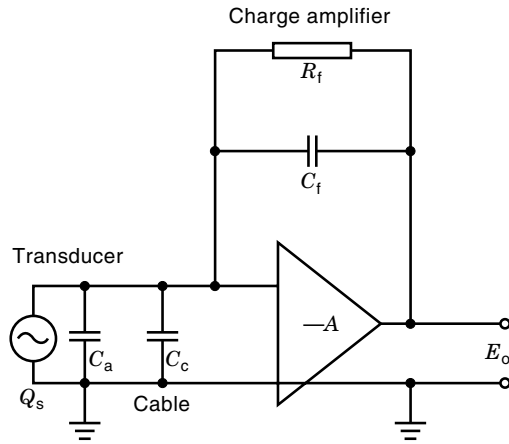


Figure 20. A typical charge amplifier. The transducer charge, which is proportional to acceleration, is first converted to voltage form to be amplified. The output voltage is a function of the input charge. The response of the amplifier can be approximated by a first-order system. In the PZT transducer the preamplifier is usually integrated within the same casing.

regarded as an active capacitor that charges itself when loaded mechanically. The configuration of external signal conditioning elements are dependent on the equivalent circuit selected. The charge amplifier design of the conditioning circuit is the most common approach, since the system gain and low-frequency responses are well defined. The performance of the circuit is independent of cable length and capacitance of the accelerometer.

The charge amplifier consists of a charge converter output voltage, which occurs as a result of the charge input signal returning through the feedback capacitor to maintain the input voltage at the input level close to zero, as shown in Fig. 20. An important point about charge amplifiers is that their sensitivities can be standardized. They basically convert the input charge to voltage first and then amplify this voltage. With the help of basic operational-type feedback, the amplifier input is maintained at essentially 0 V; therefore it looks like a short circuit to the input. The charge converter output voltage that occurs as a result of a charge input signal is returned through the feedback capacitor to maintain the voltage at the input level near zero. Thus, the charge input is stored in the feedback capacitor, producing a voltage across it that is equal to the value of the charge input divided by the capacitance of the feedback capacitor. The complete transfer function of the circuit describing the relationship between the output voltage and the input acceleration magnitude may be determined by the following complex transform:

$$E_o/a_0 = S_a j R_f C_f \omega \{1 + j R_f C_f [1 + (C_a + C_c)/(1 + G) + C_f] \omega\} \quad (46)$$

where E_o is the charge converter output (V), a_0 the magnitude of acceleration (m/s^2), S_a the accelerometer sensitivity (mV/g), C_a the accelerometer capacitance (F), C_c the cable capacitance (F), C_f the feedback capacitance (F), R_f the feedback loop resistance, and G the amplifier open-loop gain.

In most applications, since C_f is selected to be large compared with $(C_a + C_c)/(1 + G)$, the system gain becomes independent of the cable length. In this case the denominator of

the equation can be simplified to give a first-order system with roll-off at

$$f_{-3\text{db}} = \frac{1}{2} \pi R_f C_f \quad (47)$$

with a slope of 10 dB per decade. For practical purposes, the low-frequency response of this system is a function of well-defined electronic components and does not vary by cable length. This is an important feature when measuring low-frequency vibrations.

Many accelerometers are manufactured with preamplifiers and other signal-conditioning circuits integrated with the transducer enclosed in the same casing. Some accelerometer preamplifiers include integrators to convert the acceleration proportional outputs to either velocity or displacement proportional signals. To attenuate noise and vibration signals that lie outside the frequency range of interest most preamplifiers are equipped with a range of high-pass and low-pass filters. This avoids interference from electrical noise or signals inside the linear portion of the accelerometer frequency range. Nevertheless, it is worth mentioning that these devices usually have two time constants, external and internal. The mixture of these two time constants can lead to problems particularly at low frequencies. The internal time constant is usually fixed by the manufacturer through design and construction. Special care must be observed to account for the effect of external time constants in many applications by mainly observing impedance matching.

Signal Conditioning of Piezoresistive Transducers. Piezoresistive transducers generally have high-amplitude outputs, low-output impedances, and low intrinsic noise. Most of these transducers are designed for constant-voltage excitations. They are usually calibrated for constant-current excitations to make them independent of external influences. Many piezoresistive transducers are configured as full-bridge devices. Some have four active piezoresistive arms, together with two fixed precision resistors to permit shunt calibration in the signal conditioner as shown in Fig. 21.

Microaccelerometers. In microaccelerometers signal-conditioning circuitry is integrated within the same chip with the sensor as shown in Fig. 22. A typical example of the signal-conditioning circuitry is given in Fig. 23 in block diagram

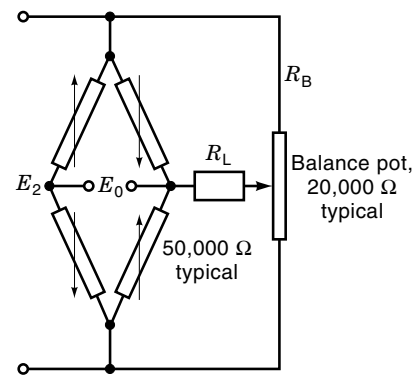


Figure 21. A bridge circuit for piezoresistive and strain gauge accelerometers. The strain gauges form the four arms of the bridge. The two extra resistors are used for balancing and for fine-adjustment purposes. This type of arrangement reduces temperature effects.

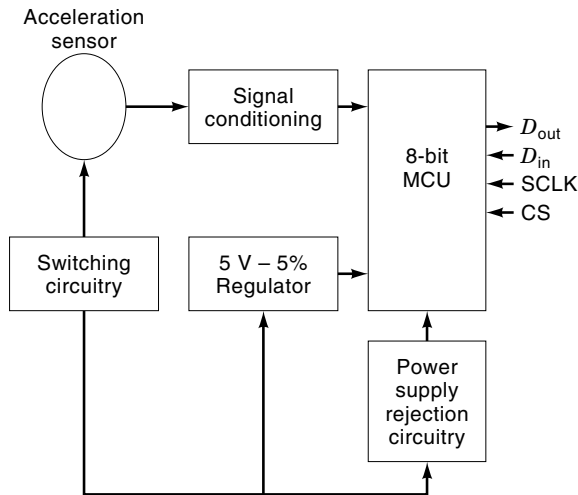


Figure 22. A block diagram of an accelerometer combined with microcontroller unit (MCU). The signal-conditioning, switching, and power-supply circuits are integrated to form a microaccelerometer. The device can directly be interfaced with a digital signal processor or a computer by using the data out D_{out} , data in D_{in} , chip select CS, and clock synchronization SCLK pins. In some cases analog-to-digital converters and memory are also integrated.

form. In this type of accelerometer, the electronic system is essentially a crystal-controlled oscillator circuit and the output signal of the oscillator is a frequency-modulated acceleration signal. Some circuits provide a buffered square-wave output that can directly be interfaced digitally. In this case the need for analog-to-digital conversion is eliminated, thus removing one of the major sources of errors. In other types of accelerometers signal conditioning circuits such as analog-to-digital converters are retained within the chip.

Force Feedback Accelerometers. These often must be digitized for use in digital systems. If they are used in inertial navigation systems there may be specific problems introduced by the accuracy requirements. The dynamic range may exceed

24 bits, and the system must operate in real time. Accurate integration may be required to get velocity changes as an output. A common solution is to use voltage-to-frequency or current-to-frequency converters to convert the analog signals to a train of velocity-weighted pulses. These converters cost as much and add as much to the error budget as the accelerometer.

Here, it is worth mentioning that global positioning systems (GPSs) are becoming add-ons to many position-sensing mechanisms. Because of antenna dynamics, shadowing, multipath effects, and the need for redundancy in critical systems such as aircraft, many of these systems will require inertial aiding tied in with accelerometers and gyros. With the development of micromachining, small and cost-effective GPS-assisted inertial systems will be available in the near future. These developments will require extensive signal processing with a high degree of accuracy. Dynamic ranges of the order of one million to one (e.g., 30 to 32 bits) need to be dealt with. A challenge awaits the signal-processing practitioner in achieving these accuracy requirements.

BIBLIOGRAPHY

1. K. G. McConnell, *Vibration Testing: Theory and Practice*, New York: Wiley, 1995.
2. E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.

Reading List

- E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
- R. Frank, *Understanding Smart Sensors*, Boston: Artech House, 1996.
- Machine Vibration: Dynamics and Control*, London: Springer, 1992–1996.
- K. G. McConnell, *Vibration Testing: Theory and Practice*, New York: Wiley, 1995.

HALIT EREN
Curtin University of Technology

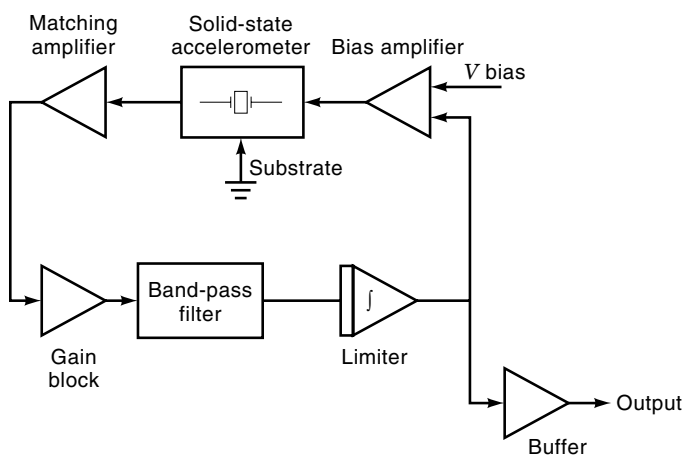


Figure 23. A block diagram of a signal-conditioning circuit of a microaccelerometer. The output signal of the oscillator is a frequency-modulated acceleration signal. The circuit provides a buffered square-wave frequency output that can be read directly into a digital device.

TACHOMETERS

Tachometers are devices for measuring the rotation speed of machine shafts, relative to a referential, which is generally the machine's stator. Tachometers get their name from the Greek words *takhys*, meaning "quick," and *metron*, meaning "measure." Therefore, etymologically, a tachometer is a device capable of measuring a speed. However, in the technical context, linear velocity meters generally are not referred to as tachometers, and the word is reserved for shaft rotation speed meters.

Moving vehicles can have translation movements and rotation movements relative to a referential. When guidance and navigation problems are considered, these movements must be sensed and their speed measured. The devices for measuring these rotational movement speeds of the vehicle relative to a fixed referential of inertia are generally called rotation rate meters or gyrometers and are not referred to as tachometers.

A rotation speed can be obtained from an angular position sensor by means of a time derivation. However, derivation

increases measurement errors and noise that may exist in the position measurement. The speed can also be obtained from an angular acceleration sensor by means of a time integration. However, integration has an initial value undetermination and may also cause the integrator to wind up due to any offset that may exist in the acceleration measurement. Therefore, even when a mechanical system has angular position or acceleration sensors, the speed of rotation is generally measured independently with some kind of tachometer.

Tachometers may be classified in several ways. One possible classification is based on the type of the tachometer's output quantity, which may be

1. A voltage value (amplitude or rms [root mean square] value): tachogenerators
2. A visual indication by a pointer mechanical position: drag-cup and fly-ball tachometers
3. A frequency value (or time value): inductive, Hall effect, optical pulse, strain gauge, variable capacitance, and switch tachometers; or stroboscopes

In this article, the description of the several types of tachometers follows an order suggested by this classification. Note that some tachometers may fit into more than one class, depending on the way their output is used. For instance, tachogenerators and inductive pulse tachometers may fit into class 1 or 3; stroboscopes may be classified as class 2 or 3.

Another classification is based on the principle of physics that underlies its operation. From this viewpoint there are

1. Electrodynamic tachometers: tachogenerators, drag cup, inductive pulse
2. Mechanical tachometers: the fly-ball
3. Magnetic tachometers: the Hall effect
4. Optical tachometers: the optical pulse and the stroboscope

Tachometers may be permanently installed and fixed on the machine group for constant monitoring of speed. For this purpose, any kind of tachometer but a stroboscope will do. Or they may be intermittently used to check the speed, without the need for permanent mechanical setting. For this purpose, the stroboscope and the optical reflected pulse tachometers are suitable, as they do not need moving pieces and only need an appropriate light source to be pointed at the machine to get a speed reading. Although not so well adapted, other types of tachometers with rotating internal parts may be found in portable tachometers. In these portables, the axle of the rotating parts protrudes from the apparatus box and has a rubber wheel on the end of it. To use the meter, the operator must put this wheel in contact with the rotating shaft, which then transmits the motion to the inner mechanisms of the device. The output then appears on the tachometer display. Care must be taken to avoid getting grease and dirt on the rubber wheel, which would cause it to slip on the shaft.

A vehicle's rotational speed is generally much slower than a machine shaft rotation. This vehicle motion speed can be measured with two kinds of sensors: the rate gyroscope, based on the mechanical conservation of angular momentum; and optical gyros, based on the interference of two optical beams whose wavelengths are modified differently by the motion. Although

these are not considered to be tachometers, they do measure a rotation speed and therefore will be described later in this article.

TACHOGENERATORS

Tachogenerators are small auxiliary machines acting as generators, driven by the shaft of the group whose rotation speed is to be measured. They provide an output voltage that has certain characteristics (constant value, rms value, frequency, etc.) proportional to that speed. There are several types of tachogenerators.

Dc Tachogenerators

Direct current (dc) tachometers are essentially dc generators (dynamos). Driven by a rotating axle, they provide an output dc voltage proportional to the rotation speed. A dynamo (Fig. 1) is composed of an iron rotor (armature), where a number of copper coils are wound, rotating inside a stator (inductor), which creates a magnetic induction flux ϕ that is stationary in space and constant in time. This induction flux can be generated by extra coils in the stator, which act as electromagnets being fed by direct currents; or it can be created by permanent magnets in the stator poles.

According to Faraday's law of induction, in each rotating coil an alternative emf (electromotive force) $e(t)$ is induced whose maximum is $E = kN\phi$, where k is a constant dependent on the way the device is constructed and N is the rotating speed in rpm.

The coils of the rotor have their terminals connected to different insulated copper segments of a cylinder, coaxial with

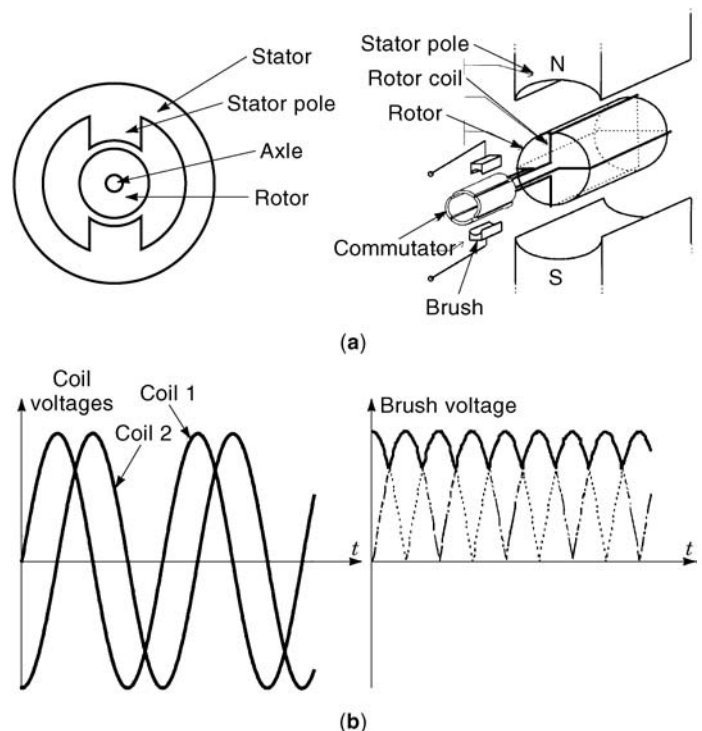


Figure 1. Dc tachogenerator (dynamo). (a) Internal constitution. (b) Voltages.

the rotor, called the commutator [Fig. 1(a)]. A pair of stationary carbon brushes slips over the segments, with little friction. The brushes are in contact with only one pair of segments at a time and, therefore, they establish electric contact between an outside stationary circuit and the two terminals of only one coil each time.

The brushes are positioned so that the two segments come in contact with the two terminals of a coil exactly when that coil's emf goes through its maximum value. As the rotor's rotation tends to decrease that emf, it also moves those segments away from the brushes, making way for new segments, which establish contact with the next coil where the emf is now at a maximum. In this way, the commutator acts like a mechanical rectifier, picking up only the maximum value of emf [Fig. 1(b)]. So the output voltage of the dynamo, measured between the stationary brushes at no load condition, is $V = E = kN\phi$.

Keeping the flux ϕ constant, the voltage is proportional to the rotation speed N . As can be seen by Faraday's law, inversion of the rotation direction (correspondent to a switch of algebraic sign of N) causes an inversion of the output voltage polarity. So the voltage is a measure of both direction and magnitude of the rotational speed.

As the number of armature coils is finite, the brushes are in contact with them not only at the moments when their emf is at a maximum, but also a little before and a little after that. Therefore, the voltage at the brushes is not exactly constant but has a ripple, which decreases as the number of coils built into the rotor increases. In practice, other sources of ripple exist, such as eccentricities, asymmetries, and irregularities of the brushes contacts. The ratio of the amplitude of that ripple (difference between maximum and minimum) to the mean voltage value is called undulation and is an index of performance of the tachometer. This undulation value can reach as high as 2%.

Typical sensitivities of the dc tachometer generators are about 5 V to 10 V per 1000 rpm, with ranges of typically 0–5000 rpm up to 0–10,000 rpm.

To behave as a tachometer, the dynamo must be kept at constant flux condition. So the flux is usually generated by permanent magnets made of appropriate alloys or sintered materials; and no load current must be allowed, so the tachometer must only be connected to high-impedance voltmeter devices.

To minimize the mechanical load of the tachometer on the group, the rotor is built with as little a moment of inertia as possible. One typical construction solution is to wind the armature coils over a hollow plastic rotor while the magnetic circuit is completed by an iron stationary cylinder inside the rotor.

Dc tachometers have the advantage of furnishing a dc signal proportional to the speed, with nonlinearity usually less than 1% of full scale and with a polarity that shows the direction of movement. Therefore, they are appropriate for control purposes. However, they have the disadvantage of having a voltage ripple, which may demand electric filtering, causing a time delay that may be inconvenient. Another disadvantage is that the construction of the windings and of the commutator is difficult, making the devices expensive. For these reasons, these devices tend to be replaced by digital, pulsed types.

Ac Tachogenerators

Alternating current (ac) tachometers are ac generators coupled with the group, producing an ac voltage that is a measure of the speed. Ac machines of two types, synchronous and induction,

can be used to build two corresponding types of ac tachogenerators.

Ac Synchronous Tachogenerators. The ac synchronous generator, also called an alternator, generates an ac voltage that is synchronized with the rotation speed of the rotor. Generally it is built with the inductor in its rotor and the induced coils in the stator. By means of dc rotor currents or permanent magnets at the poles, the rotor creates an induction field B in the airgap, which is constant in time and is fixed with respect to the rotor. The revolution movement of the rotor inside the stator causes a similar rotation of the magnetic field. This field is expected to have sinusoidal spatial distribution.

The induction flux ψ linked to a coil in the stator with n turns is $\psi = n\phi \cos \alpha$, where ϕ is the simple flux produced by a rotor pair of poles and α is the electric angle between the coil and the pole magnetic axes. With the rotor rotating at an angular speed ω , the angle is $\alpha = \omega t$. However, if the rotor has p pairs of poles, the magnetic north-south sequence passing through the coil in the stator occurs p times in each mechanical revolution. Then its electric frequency is $p\omega$, and then $\alpha = p\omega t$. According to Faraday's law of induction, the emf induced in the coil is then

$$e(t) = (p\omega)n\phi \cos(p\omega t\pi/2)$$

If no current is drawn from that coil, its no load voltage equals this emf. So in the static coil, an alternative voltage is produced, with both its magnitude and its frequency proportional to the rotating speed, provided the flux ϕ is kept constant. To keep a constant flux ϕ , the ac synchronous tachogenerators have permanent magnets on the rotor, built with the same technologies referred to in the section on dc tachometers; and their stator coils must be connected to high-impedance voltmeter devices. To increase the induced voltages, they generally have more than one pair of poles, typically four. Their sensitivities and ranges are similar to those of the dc tachometers.

Ac synchronous tachogenerators are less expensive than the dc tachometers because they do not need the costly commutator. The speed information is obtained by measuring the tachogenerator voltage amplitude with an ac voltmeter (sensitive to rms values, or a dc voltmeter provided the voltage is electronically rectified). For control purposes, the alternative voltage must be conditioned by means of an ac/dc converter and filtering circuit, which may cause undesirable delays.

Alternatively, the rotating speed may be obtained by measuring the voltage frequency, which is necessarily p times the speed in rps. This is a more accurate means of measurement, if a digital frequency meter is used, and is appropriate for use in digital measuring and control systems.

Neither the voltage nor the frequency measures give any information about the direction of the rotation. This is a disadvantage of the ac synchronous tachogenerators when compared with the dc ones. To overcome this limitation, synchronous tachogenerators with two stator coils are used. Voltages induced in both coils are equal in amplitude and frequency but have different phases (typically they differ by 90°). A circuit sensing which one is leading and which one is lagging provides the information about the direction of rotation.

Ac Induction Tachogenerators. Ac induction tachogenerator tachometers have the structure of a two-phase induction motor: The stator has two different coils whose magnetic axes are

perpendicular in space; the rotor has no salient poles and also has a two-phase winding, which is kept short circuited. Indeed, as all the rotor conductors are short circuited at their terminals, there is no need to individualize the coils. These can be replaced by a uniform distribution of rotor conductors, forming the so-called squirrel cage rotor. The currents induced in the cage, regardless of the number of its conductors, have the same magnetic effect as a two-phase winding. The same effect is obtained with an even smoother current distribution, like the one obtained in a continuous conducting sleeve over the rotor, the so-called drag-cup rotor.

One of the stator coils is fed with a sinusoidal voltage $u_1 = U \cos(\omega t)$ while the other is kept in open circuit. Because the two coils are orthogonal, the voltage induced in the second coil will be null if the rotor is stopped. However, if the rotor is moving with angular velocity Ω , the currents induced in it will cause a magnetic link with the second coil. If resistive voltages can be neglected compared with inductive ones, then a voltage will be induced in the second coil that is $u_2 = \Omega k U \cos(\omega t + \pi/2)$, where k is a constant that depends on the construction of the device. That is, a voltage is obtained that has the same frequency as the feeding voltage and whose amplitude is proportional to the speed to be measured. The direction of the rotation is obtained by comparing the phases of u_1 and u_2 . Indeed, reversing the direction of the movement is equivalent to changing the algebraic sign of Ω , which is the same as switching the phase of u_2 by 180° .

The preceding expressions assume that the second coil is kept without current, so a high-impedance voltmeter device must be used. A constant rotating speed Ω is also assumed; this means that if Ω changes, the frequency of the changes of Ω must not approach the frequency ω of the input voltage. This is generally the case. For the most common cases, power line frequency is used to feed the tachometer, as the main machine mechanical speed cannot have rapid oscillations on the order of tens of hertz.

This type of tachometer may be built with a squirrel cage rotor, thus providing an inexpensive instrument. However, the drag-cup rotor is more common because it provides better results with the advantage of having a lesser inertia than the former. The drag-cup rotor consists of a hollow sleeve of aluminum providing the rotor electric circuit. The cup rotates around a stationary iron core that completes the magnetic circuit.

With drag-cup rotors, ac induction tachometers can have an accuracy up to 0.1% of full-scale deflection, with ranges typically between 0 rpm and 5000 rpm and sensitivities of up to 10 V per 1000 rpm.

THE DRAG-CUP TACHOMETERS

The drag-cup tachometer (Fig. 2), also called eddy-current tachometer, is not a tachogenerator. It does not furnish an electric signal proportional to the speed. Instead, its response is a visual indication of speed by means of the position of a needle over a scale. It can be used in addition to any position transducer if an electric signal is needed, but this is not a common solution. The drag-cup tachometer is a very common device and is generally used as a speed and engine rpm indicator in automobiles, airplanes, and other vehicles.

The instrument consists of a rotating cylindrical permanent magnet driven by the rotating shaft. The magnet rotates inside

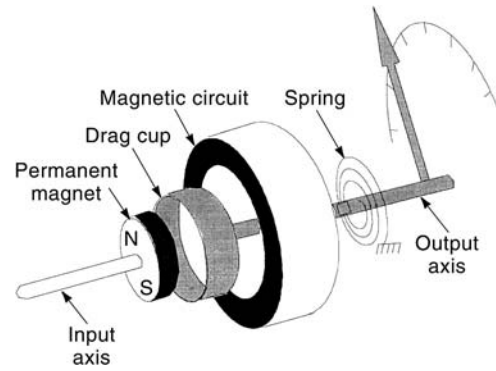


Figure 2. The drag-cup tachometer (exploded view).

a hollow conductive sleeve (a cup, usually made of aluminum), coaxial with the shaft. This cup can also rotate, but its movement is restrained by a spring. Surrounding the conductive cup, a fixed hollow iron cylinder closes the magnetic circuit of the magnet's induction field lines.

When the magnet rotates, its revolving magnetic field induces eddy currents in the conductive cup. The amplitude of these currents is proportional to the time derivative of the induction flux (that is, proportional to the rotation speed). These currents interact with the magnetic field and generate an electrodynamic torque proportional to the field amplitude (which is constant) and to the currents' amplitude. Therefore, the torque generated is proportional to the rotating speed. This electrodynamic torque causes a displacement of the cup to a position where the spring resisting torque balances the first one. The cup is attached to a pointer, whose position on a scale is an image of the equilibrium position of the cup, thus furnishing a reading of the speed.

This mechanism is similar to an induction motor: The rotating field of the permanent magnet is similar to the rotating field generated by a current-fed polyphase stator winding, with currents of a frequency similar to the rotating speed of the magnet. The cup behaves as a short-circuited rotor at standstill, experiencing the starting torque. Under these conditions, induction motor equations show that the torque is approximately proportional to the stator frequency (in this case, to the speed of the magnet), provided the currents induced in the cup do not have magnetic effects comparable with the field of the permanent magnet.

The permanent magnet used in drag-cup tachometers generally is made of Al-Ni, or other appropriate alloys, and has up to five pairs of poles. The dependence of its magnetic properties on the temperature is corrected by means of a thermoperm compensating disk. This is a disk made of a material whose magnetic conductivity depends on the temperature, placed close to the cylindrical magnet and partially shunting it to a degree that also depends on the temperature. The dependencies practically cancel each other.

Drag-cup tachometers are used in speed ranges up to 10,000 rpm and have a measuring accuracy on the order of 1%.

FLY-BALL TACHOMETERS

The fly-ball tachometer is a mechanical device that has been in use for a long time as a speed-governing device for thermic

engines (watt regulator). Used as a tachometer, it gives a visual information of the speed, positioning a pointer on a scale. Position sensors can be added to convert the pointer position into an electrical signal if this is needed.

The fly-ball tachometer consists of two massive spheres, connected through a system of levers to the shaft. The shaft rotation causes the spheres to rotate around it. The centrifugal force drives the spheres away from the shaft, but their radial displacement is restrained by the levers acting against a restitution spring. At equilibrium, the spheres rotate at a distance where the spring force and the centrifugal force are equal. This distance is shown by the position of a pointer driven by the lever-spring system.

Normal springs exercise a restitution force proportional to their deformation x , $F_s = K_s x$. Centrifugal force acting on constant masses is proportional to the rotation radius, R , and to the square of the rotation speed, Ω : $F_c = K_c R \Omega^2$. Note that the spring deformation and the rotation radii of the spheres R are related in a way depending on the mechanical design of the gearing, $R = R(x)$. At equilibrium $F_s = F_c$, which leads to a nonlinear equation $x = (K_c/K_s)R(x)\Omega^2$. Therefore, these tachometers have a nonlinear velocity scale. In some models, the nonlinearity is compensated for, to a certain extent, by using a nonlinear spring.

INDUCTIVE PULSES TACHOMETERS

The inductive pulses tachometer is a type of tachometer in which there is no mechanical contact between the moving shaft and the stationary sensor (Fig. 3). The output of the tachometer consists of a train of pulses, which can be processed either by analog or digital means. It consists of an iron-toothed wheel mounted in the shaft and rotating with it. Placed near it is a permanent magnet, around which a sensing coil is wound [Fig. 3(a)]. The magnet may be placed, as suggested by the figure, at a distance of about a millimeter from the outer face of the teeth. Alternatively, the magnet may be U shaped, with the wheel's teeth passing between the two poles.

As the wheel rotates, the teeth and gaps pass successively in front of the magnet, causing a periodic variation of the magnetic reluctance of the flux lines' path. Therefore, the induction flux through the coil wound around the magnet varies, thus causing periodic pulses of voltage to be induced at its terminals [Fig. 3(b)]. Both the amplitude and the frequency of the voltage impulses are proportional to the rotation speed and to the number of teeth in the wheel.

The rotation speed can then be obtained by means of any ordinary method of measuring the voltage peak value, time-average value, or rms value. However, this is not the most

common way, as the pulse amplitudes depend on circumstantial factors, like the temperature or variations of the sensor position relative to the toothed wheel. Another drawback of this method is the difficulty of measuring very low speeds when the pulse amplitudes are not high enough to be measured accurately.

Alternatively, the speed can be obtained, and usually with better accuracy, by measuring the pulse frequency, f , which is equal to the shaft's number of rotations per second times the number of teeth n of the toothed wheel. Expressing the rotation speed N in rpm, the following is obtained:

$$f = nN/60$$

When using the frequency measurement method, generally the pulses are first processed through some sort of electronic comparing circuit, like the Schmitt trigger or equivalent, in order to produce a train of sharper impulses, whose amplitude is now independent of the speed. The frequency of the pulses is then measured, either analogically or digitally. The ways of implementing these measures will be described later. The difficulty of measuring very low speeds persists, however, because the amplitude of the pulses from the pick-up coil may be too low to switch the electronic trigger. (This can be overcome if the field variations are sensed, not by means of the inductive process just described, but with a Hall effect sensor. The pulses produced by the Hall effect sensor have a constant amplitude, independent of speed, so only their frequency varies with the rotation velocity.)

To increase the sensitivity of the sensor at low speeds, wheels with more teeth are used. However, this procedure has time and space limitations. If there are too many teeth, the distance between them may become too small compared with the size of the polar piece. The polar piece would then face more than one tooth at a time, causing a space integration and averaging effect. The device would not be able to distinguish the teeth from the gaps and thus would become insensitive to the rotation. For this reason, the polar piece generally has a sharp tip, of a conical or other similar shape. A similar averaging effect occurs because of a time limitation. The pulses have nonzero duration. A wheel with a number of teeth, few enough to avoid the space integration but numerous enough to measure low speeds, will cause pulses of increasing frequency when the shaft speeds up. At high velocity, the pulses can become too rapid and start to overlap, and then the sensing circuit is no longer able to distinguish the too frequent pulses. For these reasons, a wheel with an appropriate number of teeth must be chosen for the range of rotation speeds to be measured.

The shape of the wheel's teeth is also designed according to the needs of the electronic circuitry to be used: Smoother teeth may cause an almost sinusoidal voltage to be induced, and teeth with sudden geometrical changes induce sharper pulses.

Some inductive pulse tachometers are provided with a second pick-up coil, generally placed in such a way as to induce pulses in quadrature with those of the first coil. The phases of both signals may be compared in some electronic circuitry to find if the second leads or lags behind the first one, thereby obtaining information about the direction of the rotation.

As there is no mechanical contact between moving and fixed parts, this kind of tachometer is suitable for use in low-torque mechanisms. If low inertia is also required, a plastic wheel is used, in which ferromagnetic pieces are regularly inserted in radial positions near the periphery. These pieces have the same

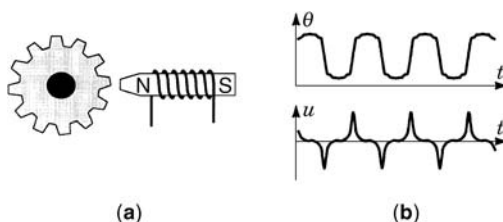


Figure 3. Inductive pulses tachometer. (a) Schematic constitution. (b) Flux and voltage at pick-up coil.

effect of regularly modifying the magnetic reluctance of the flux path, thereby causing a periodic variation of the flux through the pick-up coil with the consequent induction of a voltage train of pulses. This kind of wheel is more expensive than an iron one, but, in addition to the lower inertia it offers, it has the advantage of producing sharper voltage pulses.

Inductive pulse tachometers are widely used because they need almost no maintenance, they need no power feeding circuits, and the sensor device is not expensive. Having no electric contact between moving and stationary parts, they produce no sparks and therefore can be used in potentially explosive environments. The magnet and pick-up coil are usually assembled in a sealed case, which is mounted near the wheel, without touching it and with no magnetic barrier between the wheel and the magnet.

HALL EFFECT PULSE TACHOMETERS

The Hall effect pulse tachometer is similar to the inductive pulse tachometer, except that the magnetic induction field variations are sensed by means of a Hall effect sensor (Fig. 4) instead of a pick-up coil.

The Hall effect consists of the appearance of a voltage v_H through a material when an electric current I flows through it in the presence of an external magnetic induction field B . The electric carriers that constitute the current, because they are moving with a velocity v inside the field B , sense a Lorentz force. This is equivalent to an electric field $E = v \times B$, perpendicular to both the induction field and the current direction and that generates the external voltage that is felt.

The Hall effect can be better sensed in semiconductor than in conductor materials. The charge carriers will feel a force $F = qE$, which deflects their movement to one side of the material. If the material is p type [Fig. 4(a)], the carriers will be positive and the direction of velocity is the same as that of the current. Suppose that the B field is from the upper side; the positive traveling charges will be deflected to their left. The left side will then become positive with respect to the right side.

If the material is n type [Fig. 4(b)], the carriers will be negative; the direction of velocity is opposite to the current. The Lorentz electric field will be opposite to that of the previous case. As a consequence, the external voltage produced will also be reversed.

The voltage is directly proportional to the current I and to the field B . If a sensor based on the Hall effect device is fed with a constant current I , then the output voltage will be a measure of the induction B . Note that with inductive sensors,

it is the time derivative dB/dt that is sensed, while with Hall sensors it is the field itself that is sensed, regardless of its rate of variation. Typical sensors are supplied with currents on the order of 10 mA and have an output sensitivity on the order of 10 V/T.

The construction of Hall effect pulse tachometers is similar to inductive pulse tachometers. Like them, a toothed ferromagnetic wheel rotates in front of a permanent magnet. Instead of the pick-up coil used in inductive pulse tachometers, a Hall sensor is placed between the permanent magnet and the rotating toothed wheel. The airgap variation of the magnetic flux path, caused by the succession of teeth and gaps of the wheel, causes a variation in the intensity of the B field over the magnet's pole. The output voltage of the sensor, being proportional to the B intensity, follows its variations. Another option is to use a wheel with magnetic poles in its periphery and avoid the stationary magnet. The Hall sensor will sense the magnetic field from the wheel's poles.

The output of the Hall effect tachometers is a train of voltage pulses with a frequency f that is equal to the shaft's speed (in rps), times the number of pole pairs n of the magnetized wheel, or times the number n of teeth of the toothed wheel. Or, using the speed N in rpm,

$$f = nN/60$$

This frequency then is measured by one of the methods described later in this article. The pulses all have the same intensity regardless of the shaft's speed and, therefore, the device is sensitive to low as well as high speeds.

The sensor output is a measure of the rotation speed but gives no information about rotation direction. This is obtained with a second sensor placed in a position so that its signals are in quadrature with the first sensor's output. Electronic processing may distinguish which one is leading and which one is lagging, thus determining the direction of movement.

Hall effect tachometers have the same advantages of contactless sensing as inductive tachometers, but they have the additional advantage of sensing all speeds equally. However, they have the disadvantage of needing a power supply. When compared to optical-based tachometers, they have the advantage of being less sensitive to environmental conditions, like humidity, dust, or vibrations, and of having characteristics that vary less over time than the optical tachometers.

OPTICAL PULSE TACHOMETERS

Optical pulse tachometers (OPT) (Fig. 5), like the Hall effect tachometers, generate a train of constant amplitude pulses

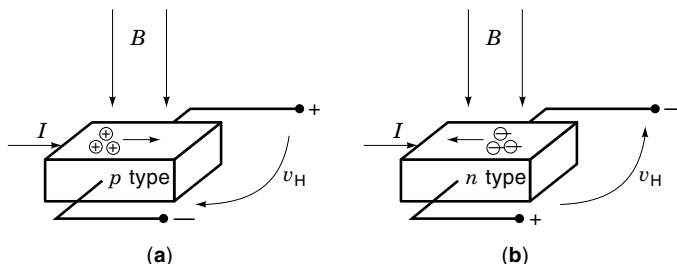


Figure 4. Hall effect sensor. (a) With p -type semiconductor material. (b) With n -type semiconductor material.

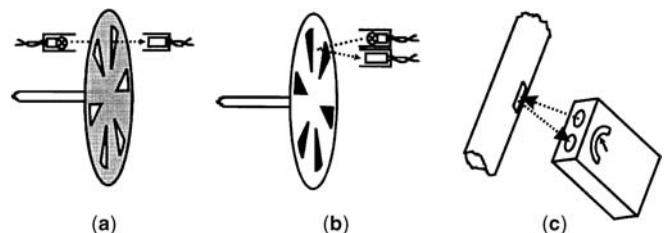


Figure 5. Optical pulse tachometers. (a) Incident light method. (b) Reflected light method. (c) Portable reflected light OPT.

whose frequency is proportional to the rotating speed to be measured. The frequency of the pulses is measured either by analog or digital means, to be described later.

The pulses may be generated by two alternative ways: the incident light method or the reflected light method (also known as the transmittance and the reflectance methods).

The incident light type of optical tachometers uses the repetitive interception of a light beam to generate the pulses. A disk [Fig. 5(a)], which has at its periphery a sequence of transparent radial windows and opaque sectors, is mounted to the shaft and rotates with it. The entire disk may be made of transparent acrylic plastic, with black radial masks painted on it. Or the disk may be made of an opaque material, plastic or metallic, with slots cut in it; sometimes a toothed wheel is used. At one side of the disk, a light source emits a light beam parallel to the shaft. The light beam falls on a light sensing device. As the disk rotates, the transparent slots and the opaque zones alternatively cross the beam, alternatively transmitting and interrupting it, which causes a pulsing response of the light detector.

The reflected light solution uses the light source and detector on the same side of the disk [Fig. 5(b)]. The disk now has no slots, but its face has alternate light-absorbent zones and light-reflecting ones. The rotation of the disk causes the light sensor to detect alternate pulses of reflected light. Marks alternatively painted black and white usually create the desired effect. If increased reflection is needed, strips of reflective tape may be glued over the disk. The surface of this tape is covered with small spherical lenses, and these lenses reflect the incident light back in the same direction, even if that direction is not perpendicular to the tape surface. The reflected beam solution may be implemented without the disk. The shaft itself may be painted with light-absorbent and -reflecting strips, or have reflecting tape strips fixed to it.

As in inductive and Hall effect pulse tachometers, the frequency f of the optical tachometer's output is equal to the shaft's rotations per second times the number n of transparent slots in the disk (for the incident light method) or of reflecting marks (for the reflected light method). Using the rotation speed N in rpm,

$$f = nN/60$$

For use in mechanical settings where access is difficult or where inflammable atmospheres proscribe the presence of electric devices, the two light beams, one from the source and the other for the sensor, may be carried by optical fibers between the shaft proximity and the remote electronic setting.

The incident light method demands the disk to be mounted on the shaft and the light source and sensor to be installed carefully, in a fixed and correct position. Therefore, it requires more mechanical care than the reflected beam method, which may be used with an improvised mechanical setting, specially if the reflection is made on the shaft itself, without the use of the disk. However, the incident method device, once correctly installed, is more reliable because it provides steadier pulses than the reflected beam setting, which is more prone to errors due to vibration, misalignment, dust on the reflecting surfaces, and so on.

The ease of use of the reflected light method allows for the construction of stand-alone, battery-operated portable tachometers [Fig. 5(c)]. They include a light source, the reflected

light sensor, the signal processor, the frequency meter, and a digital or analog display where the speed can be directly read, normally in rpm. Usually these tachometers are furnished with a set of reflecting tape strips. The operation consists only of pointing the instrument at the shaft, where the tape has been previously attached, pressing a switch to emit the light, and reading the desired rotating speed.

A number of construction solutions are available for these tachometers. Any light source may be used. Generally an electric bulb has been used, but an increasing number of models use light emitting diodes (LEDs). The sensor used to be a photoelectric cell, but photosensitive semiconductor devices are now most common.

Photoelectronic devices having both a photodiode and phototransistor are available on the market and are also used in the construction of tachometers. Some of these devices are made in a U shape, with the diode at one extreme and the transistor at the other, and are suitable for the construction of incident beam tachometers, the slotted wheel passing between the two extremities of the U. Other devices are made with the diode and the transistor side by side and are appropriate for reflected beam tachometers. There are also reflected beam tachometers that use a laser beam, generally from a laser diode, when increased distance range is desired.

Like other pulse tachometers, optical tachometers measure the speed of rotation but not its direction. Another light source and sensor placed in quadrature is needed for phase detection and direction determination.

Optical tachometers have an advantage over the magnetic ones of easier installation and operation, but they have the disadvantages of being susceptible to environmental pollution, dust, and grease and of experiencing accelerated aging as the light source properties generally decay in time.

OTHER PULSE TACHOMETERS

Strain Gage Tachometers

A strain gage is a dc-fed Wheatstone bridge whose resistors vary their resistance when stretched or compressed. The resistors are fixed to a mechanical piece, which is to be deformed, and are mechanically installed in such a way that the deformation of the piece does not equally deform the four bridge arms. In so doing, a strain in the piece unbalances the bridge and causes a voltage output proportional to the deformation.

If the rotation of a shaft can be transformed into a periodic deformation of a stationary piece, then this deformation can be detected with a strain gage, causing an output signal that is synchronized with the rotation. The frequency of the signal is, therefore, equal to the rotation speed, but nothing can be said about its amplitude (this is preferably constant, but mechanical resonance may cause it to vary with the speed). This can be an appropriate way of sensing the rotation speed when the mechanical installation, by its own design characteristics, naturally causes a vibration or deformation of one of its stationary components. Then no further mechanical settings are necessary, except for the setting of a strain gage. If that mechanical deformation does not naturally occur, the method can still be implemented by introducing an extra piece and forcing it to suffer a deformation. The strain gage then is fixed to a cantilever-mounted beam, which is in contact with the periphery of an eccentric wheel or cam, driven by the shaft. The

rotation of the eccentric wheel or cam causes a sinusoidal vibration of the beam, synchronized with the rotation.

The output characteristics of these tachometers depend on the sensitivity of the strain gage, and on the mechanical properties of the mechanical setting. The amplitude of the signal is not a measure of the speed, but its frequency is, and it should be read by one of the methods described later. However, transient mechanical vibrations or resonance with harmonics may introduce spurious frequencies and thus cause a false speed reading.

Variable Capacitance Tachometers

The usual design of variable capacitors consists of a set of fixed conductive parallel plates, among which another set of parallel plates can rotate. These are parallel to the fixed plates, and their rotation axle is perpendicular to both sets of plates. The capacitance of the capacitor is proportional to the total juxtaposing area of fixed and movable plates, and this area depends on the angular position of the rotating plates.

If the movable plates are driven by a rotating shaft, then the capacitor will have a periodically changing capacitance, varying synchronously with the shaft rotation. This capacity variation may be sensed and translated into a measure of speed, in two different ways. One way is to feed the capacitor with a constant voltage U . Then its current i_C will be

$$i_C = \frac{dq}{dt} = \frac{d(CU)}{dt} = U \frac{dC}{dt} = U \frac{dC(\alpha)}{dt} \cdot \frac{d\alpha}{dt} = U \cdot \omega \cdot \frac{dC(\alpha)}{dt}$$

that is, proportional to the rotation speed ω , and will depend on the way the capacitance varies with the position.

If a smooth varying current is desired, the geometric design of the plates is such that a sinusoidal dependence $C(\alpha)$ is obtained, which causes a sinusoidal current to flow in the tachometer, with both amplitude and frequency proportional to the speed. The measure of any of these values indicates the desired speed.

For low speeds, these sinusoidal change derivatives may be too low to cause an easily measurable current. For this reason, the plates' design may include a smooth capacitance variation (for instance, a linear one, followed by a sudden change). The resulting current will be a smooth, low current during most of the period, followed by a sudden and sharp spike. The frequency of these pulses is proportional to the rotation speed and is fed to any frequency meter device to obtain the speed value. In general, the current must be first converted into a voltage, for measurement purposes.

An optional method of implementing a variable capacitance tachometer is to connect the capacitor to a Wien, Maxwell, or other ac capacitance measuring bridge. The bridge is fed with a ac voltage of a frequency higher than the rotation rhythm (more than 10 times the rotation pace) and is balanced for an intermediate capacitance value of the variable capacitor. The rotation of the shaft and the consequent variation of the sensor capacity periodically unbalance the bridge and cause it to output an ac voltage with the frequency of the power source and with an amplitude that periodically changes at the pace of the rotation. This is an amplitude modulation, with the feeding frequency acting as a carrier and the rotation speed acting as a modulator. This output signal is then demodulated to extract the modulator. The frequency of the modulator is then read by

any of the usual processes, and that frequency is the desired speed value.

Switch Tachometers

The switch tachometer is a device that, when it is driven by the rotating shaft, alternatively makes and breaks electrical contacts in a circuit, causing current pulses, whose frequency is measured by any usual method. Generally the switch is connected to an RC circuit, and the current provoked is the periodic charge and discharge of the capacitor.

One type of switch tachometer has a pair of rotating contacts, driven by the shaft and connected to a capacitor pair of terminals. A dc voltage is fed through a resistor into a stationary pair of terminals. The rotating capacitor terminals contact the fixed ones, and the rotation causes their periodical inversion. This causes the capacitor to be alternately charged and discharged through the resistor, with the resulting pulsed current.

Another type of switch tachometer uses a fixed circuit, with commuting switches driven by the shaft. A common way is to use reed switches. The reed can oscillate between two positions, each one making contact with a different circuit topology. The reed can be mechanically driven by a cam or yoke. A popular design is to make the reed magnetically actuated by a magnet fixed to the shaft. Being very flexible, the reed is attracted by the magnet and then released, which makes it alternatively touch two contacts.

The switch tachometer is referred to later in this article, in the section on frequency measurement.

STROBOSCOPES

The stroboscopic method of speed measurement is based on the inertia of the human eye. Therefore, it demands the presence of a human operator in the measurement loop. If a rotating or vibrating object is illuminated by a pulsing light whose flashing rhythm coincides with the rotation or vibration speed, the object will be illuminated whenever it is in the same position. Hence, it will always be seen in that position. The human eye, being incapable of distinguishing successive flashes if they are too rapid, will give the sensation of seeing an object standing still.

Stroboscope tachometers are made of a flashing lamp, a device to control its rhythm, and a frequency meter. Most stroboscopes have all the components integrated in the same portable box. The operation consists of pointing the light at the rotating shaft, regulating the flash frequency until the shaft appears stationary, and reading the frequency. For easier detection of the apparent immobility, the shaft must have some distinguishable feature on its outer surface. Normally, a white mark is painted or a reflective tape is fixed to the shaft. Old stroboscopes used a slotted rotating wheel to mask periodically the vision of the shaft or to interrupt periodically an illuminating light beam. Modern stroboscopes use a gas-filled flash lamp, which has a grid between its cathode and anode.

A high voltage, from a capacitor that is periodically charged, is applied between the cathode and the anode. Voltage pulses from a pulse generator are applied to the grid, each pulse permitting an electrical discharge from the capacitor through the bulb gas, with the resulting emission of an intense and short (about 10 μ s) strobe of light. Before the next voltage pulse is applied to the grid, the capacitor is again charged.

The frequency of the triggering pulses may be read directly on a dial mounted on the adjustment knob of the flashing rhythm, or the stroboscope may include an extra frequency meter, either analog or digital. Generally this frequency may be directly read in impulses per minute, to coincide with the usual way of expression rotation speeds.

Sometimes it is difficult to determine what the speed is, because the shaft may look like it is standing still for different flash frequencies. This may happen when different harmonics of the rotation rhythm are used as flash frequencies (in this case, the image of the mark in the shaft may be repeated in more than one position). Then a frequency f_1 for apparent immobilization must be found, and then the next and closest stationing frequency f_2 must be located. These two frequencies, because they are as close as possible, are two consecutive harmonics of the rotation rhythm N ; that is, $f_1 = KN$ and $f_2 = (k + 1)N$; therefore, the speed is $N = f_2 - f_1$ (assuming all the quantities are in the same units, that is, minute^{-1} or second^{-1}).

Inversely, if the flashing rate is a subharmonic of the rotation speed, $N = Kf_1$, then the same frozen image aspect will occur, with the mark visible in only one position. Then the frequency of flashes must be decreased until the next frozen image appears, which happens at frequency f_2 , such as $N = (K - 1)f_2$. Then $1/N = 1/f_2 - 1/f_1$ or $N = f_1 f_2 / (f_1 - f_2)$ is obtained.

Stroboscopes are an easy and straightforward method of measuring speed that can be used without any mechanical setting and applied to any periodically moving mechanism. However, they have the disadvantage of not furnishing an automated output. The range, precision, and light intensity of these devices vary from maker to maker. Ranges from 5 Hz (300 rpm) to 300 Hz (18,000 rpm) or even up to 417 Hz (25,000 rpm) are available, with accuracy on the order of 1%. The light intensity of each flash may be on the order of 1500 lx.

FREQUENCY MEASUREMENT

Analog Frequency Measurement

There are several methods for analog frequency measurement, but we shall limit the discussion to those most often used in tachometers. These are the conversion of the frequency into a proportional dc voltage, which is measured with a common dc voltmeter, or into a dc current, which is measured with a dc ammeter.

Frequency Voltage Conversion. The frequency voltage converter (Fig. 6) is based on the fact that the average value V of a train of pulses of constant amplitude U , constant duration T_0 , and period T [Fig. 6(a)] is $V = UT_0/T$. With these assumptions, $UT_0 = A$ is a constant (the constant area of each pulse), and $1/T = f$ is the frequency to be measured. Hence, $V = Af$ results. Each constant amplitude, constant duration pulse is produced by a monostable circuit, which must be accurate and stable. Its pulse duration must, of course, be shorter than the minimum period corresponding to the maximum frequency to be measured.

The input signal is fed to a shaping circuit [Fig. 6(b)], generally a Schmitt trigger and a differentiator, in order to get new short and sharp pulses synchronized with the incoming ones. The shaped pulses attack the monostable circuit, each pulse causing the output of a constant-area pulse. The train of these constant-area pulses is fed to an averaging circuit

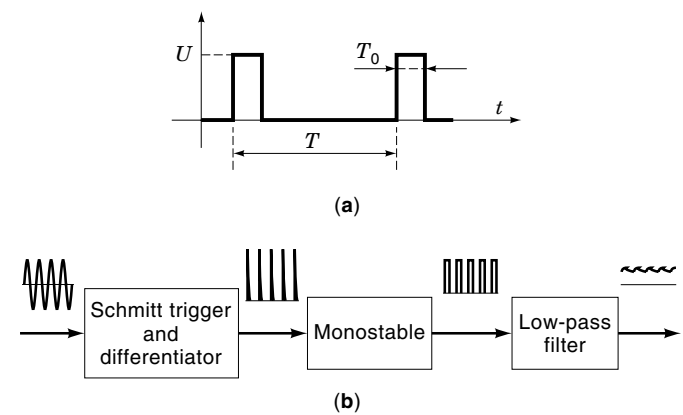


Figure 6. Frequency measurement by frequency-voltage conversion. (a) Time diagram of a train of constant-area voltage pulses. (b) Block diagram of a frequency-voltage converter.

(an RC circuit, or an integrating amplifier with discharging resistor), functioning as a low-pass filter of time constant τ . Its output will be the average value V , with a ripple $V_R = V/(f\tau)$.

For the propose of visual reading of the frequency with a voltmeter, there is no need for a high-quality filter, as the mechanical inertia of the meter itself has an averaging effect. For use in an automated measurement chain, a tradeoff must be made between the presence of output ripple and response time of the filter.

Frequency-Current Conversion. Another analog method of frequency measurement, similar to the first, is to convert the frequency into an average value of a capacitor charging current (Fig. 7). The alternative input signal whose frequency is to be measured, by any way, controls a switch that can be commuted between two positions [Fig. 7(a)]. The positive part of the signal causes the switch to apply a dc voltage U to an RC series circuit. The negative part of the signal commutes the switch to a position where it short circuits the same RC circuit. The time constant $\tau = RC$ remains the same during the charge and the discharge, and the circuit is dimensioned to be $\tau < T/2$ where T is the period of the signal.

For fast signals, the switch must be electronic. For signals up to a few hundred hertz, reed relays can be used. This is generally the frequency range of the signals involved in rotation speed measurements; therefore, the speed can be measured by using some kind of generating device driven by the shaft to produce an alternating voltage and using this voltage to drive an electromechanical reed relay. However, it is much simpler to actuate mechanically the reed of the switch with a cam geared

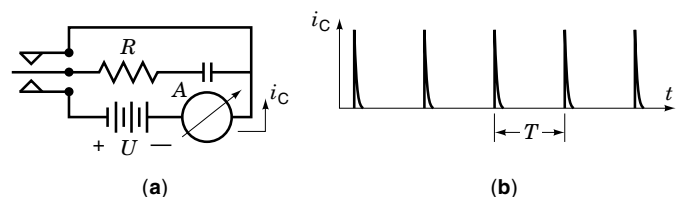


Figure 7. Frequency measurement by frequency-current conversion. (a) Circuit. The reed actuator is not shown. (b) Time diagram of the charging current of the capacitor.

with the shaft, or magnetically actuate the reed by a magnet that is fixed to the shaft and rotates with it. Thus, the frequency of the rotation can be measured without the need of generating an electric signal just for the purpose of actuating a relay.

If $\tau < T/2$ one can assume that the capacitor is totally discharged when the switch causes the voltage U to be applied and that the capacitor totally charges before it is shorted again. In this assumption, a relative error $e = e^{-T/2\tau}$ is neglected compared with 1, for which it must be $T > 10\tau$, if the error e is to be $e < 0.5\%$, or $T > 14\tau$, if it is to be $e < 0.1\%$. Therefore, the charging current i_C is a quick pulse that ends before the next commutation [Fig. 7(b)], and it carries the total final charge $Q = CU$.

The charge being the time integral of the current, and there being no more charging current for the rest of the interval T , we obtain

$$\int_T i_C dt = Q = CU$$

that is to say, the area of the pulse is constant and is equal to CU . Therefore, the average value I_{av} of the current i_C during a period is proportional to the frequency, as it is

$$I_{av} = \frac{1}{T} \int_T i_C dt = \frac{CU}{T} = CUf$$

For control purposes, this current must be converted into a voltage and averaged, to obtain a dc voltage signal. For reading purposes, however, it is enough to read the charging current with an averaging ammeter, such as the moving coil type.

Digital Frequency Measurement

Digital frequency measurement (Fig. 8) consists of counting the number of the signal periods that fit in a certain time interval

(for instance, 1 s). For this, the input signal of frequency f_s is fed into a conditioning and shaping circuit [Fig. 8(a)], which generates a train of pulses synchronized with that signal, and with appropriate amplitude for the next electronic blocks. This first circuit generally includes an attenuator, an amplifier, a trigger, and a differentiator.

The synchronized train of pulses is fed into a digital counter, through an electronic gate that is generally an AND gate. The gate is kept opened for a unit period time T_c , because, in its control input, a signal from an internal clock is applied. An edge-triggered flip-flop maintains the gate-opening control signal on for an entire clock period T_c [Fig. 8(d)]. In this way, the digital counter counts the number of input signal periods during a unit time (that is, measures the frequency).

Generally, the clock consists of a high-accuracy crystal oscillator of 1 MHz (or of 100 kHz) and a series of six (or five) frequency dividers [Fig. 8(c)]. Each one receives a signal from the previous divider (or from the oscillator, if it is the first divider), divides the signal frequency by 10, and passes the resulting signal to the next divider. In this way, clock signals of 1 MHz (or 100 kHz) down to 1 Hz are available. The digital counting error is of one unit, and the counting is $f_s T_c$. Therefore, the relative error is $\epsilon = 1/(f_s T_c)$. For increased accuracy, the measurement should be made by counting as many input signal periods as possible, which means counting during the longest internal clock period possible.

If the input signals come from a pulse tachometer that generates only a few pulses per rotation, then the input frequency may be low and the machine may be turning very slowly. Therefore, the speed measurement may demand too much time for use in control purposes. As it is the number of turns per time interval that is counted, it is the average speed in that time interval that is obtained. In the case of slow motion and long counting time intervals, this average speed may differ significantly from the usually desired instantaneous speed. In that

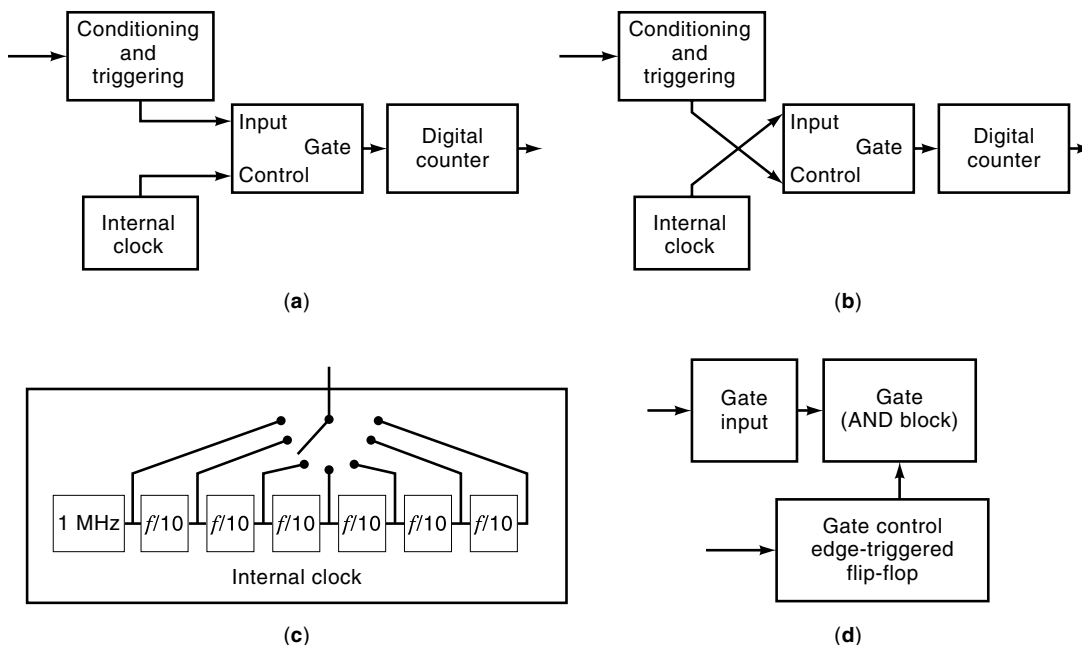


Figure 8. Digital frequency measurement. (a) Block diagram of the circuit for frequency measurement. (b) Block diagram of the circuit for period measurement. (c) Block diagram of the clock. (d) Block diagram of the gate.

case, instead of measuring the number of turns per unit time, the inverse measurement is more appropriate. The duration of a single rotation is measured. This is done by reversing the roles of the input-synchronized and the clock pulses [Fig. 8(b)]: The input-synchronized pulse is applied to the edge-triggered flip-flop, keeping the gate opened during an input period $T_s = f_s$, and the counter counts the number of clock ticks that fit in an input period. The result is the time between input signal pulses, or signal period, and the frequency is its inverse.

Using a clock of frequency $f_c = T_c$, the count will now be $f_c T_s$, and the error of one unit represents a relative error $\epsilon = 1/(f_c T_s) = f_s T_c$ (the inverse of the former case). Good accuracy requires that a quick clock signal be used, and so the measurement may be done in a short time. This is appropriate for the measurement of slow signals, such as those from a tachometer installed in a slow machine.

Inverting the time measured, an (almost) instantaneous speed of rotation is obtainable. This is sometimes called the division by time technique for speed measurement. Because the result of the operation is a digital number, the output circuitry may include a processor to execute the operations needed to convert the count into the appropriate speed units, usually rpm, if the number is to be displayed on the tachometer.

GYROMETERS

Gyrometers are devices for measuring the velocity of rotation of a vehicle with respect to an inertial frame. There are several types of gyrometers, such as mechanical spinning gyroscopes (rate gyros), mechanical vibrating gyroscopes, gas-jet nongyro rate transducers, and optical interferometers gyros (optical gyros).

Spinning Rate Gyros

Mechanical rate gyros are single degree of freedom gyroscopes. A gyroscope is essentially a spinning wheel (Fig. 9) whose spin axis coincides with its principal geometric axis of rotation. The wheel's axle is supported by bearings located in a frame that surrounds the wheel. The supporting frame itself has an axle, orthogonal to the wheel's axle. The frame can also be rotated about its axle, which is supported by a second pair of bearings. This second pair of bearings can be fixed to the device case. Therefore, the frame is free to rotate about its axle, permitting the wheel's spinning axis to have a rotation movement on a plane perpendicular to the frame's axle. The device just described is called a single degree of freedom gyroscope [Fig. 9(a)].

Alternatively, the bearings supporting the frame may be fixed to a second outer frame. This second frame has a (third) axle of its own, perpendicular to the axes of both the wheel and first frame, and now this third axle is supported by a third pair of bearings fixed in the mechanism case. With this doubly articulated suspension, the wheel's spinning axis can rotate in two perpendicular planes and can thus assume any direction in space. This is the double degree of freedom gyroscope [Fig. 9(b)]. The supporting frames are called gimbals. They are also called Cardan suspensions, after the sixteenth-century mathematician and physician Gerolamo Cardano (1501–1576), discoverer of the solution of the third-degree equation and inventor of this suspension.

If the wheel's angular velocity is Ω , then its angular momentum equals $\mathbf{H} = \Theta \cdot \Omega$, where Θ is the momentum of inertia of

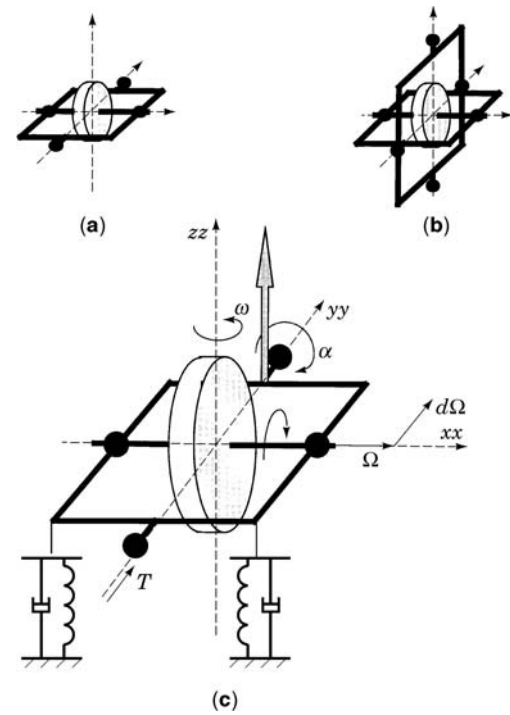


Figure 9. Spinning wheel gyroscope. (a) Single degree of freedom. (b) Double degree of freedom gyroscope. (c) Rate gyro.

the wheel. Both $\mathbf{H}$ and Ω are axial vectors, and Θ is a tensor. However, if the component of the rotational velocity of the wheel around its axis is much greater than any other components, then $\mathbf{H}$ may be considered parallel to Ω , and Θ may be considered a scalar constant. This requires that the wheel be spun at high speed.

Newton's second law applied to rotating bodies states that $\mathbf{T} = d\mathbf{H}/dt$, where $\mathbf{T}$ is the torque acting on the body and $d\mathbf{H}/dt$ is the time derivative of the angular momentum computed in an inertial reference frame. It follows from this law is that if no torque is applied to the wheel (i.e., $\mathbf{T} = 0$), then its angular momentum will be kept constant. The wheel's spin will maintain its speed and its axis will maintain its direction with reference to the distant galaxies.

This principle is the basis for the use of a double degree of freedom gyroscope as an inertial navigation device. The case of this apparatus is fixed to a moving vehicle. The vehicle's rotational motion causes the axis of the gimbals to rotate with it, with respect to an outer, inertial reference frame. Assuming that the bearings are frictionless, the movement of the gimbals transmits no torque to the spinning wheel. Therefore, in the interior of the revolving gimbals, the spinning wheel maintains the direction of its axis of rotation. The angles between the axis of the wheel and the outer case indicate the attitude (angular orientation) of the vehicle. In practice, the friction in the bearings and other causes impart a small torque to the wheel, and, therefore its axle will have slow motions, which are the major source of error in this type of mechanism. Special care is taken in their construction to reduce and compensate for these torques. However, the spatial attitude normally is measured with rate-integrating gyros (to be described later) because less error is obtained using three rate-integrating gyros (one for each spatial axis).

For measuring the turning speed (turning rate) of a vehicle, single degree of freedom gyros can be used in a configuration that is called rate gyro [Fig. 9(c)]. This device can sense a vehicle's rotation speed around one axis. To sense the rotation around the three space directions, three orthogonal rate gyros are needed.

A rate gyro is a spring-restrained single degree of freedom gyro. The gyro's single frame can execute small rotations around its own axle, but this movement is restrained by restitution springs, which impart a restitution torque proportional to the inclination of the frame about its axle. Generally, this movement is also slowed down by mechanical or electric dampers connected to the frame. The entire apparatus is fixed to its case and to the vehicle. If the vehicle turns in a plane parallel to the frame plane, the whole set is forced to turn with it.

Consider a moving referential xyz fixed to the gyroscope, with the xx axis aligned with the wheel's spinning axle, and the xOy plane coincident with the gimbal. Then the frame can oscillate about the yy axis, and the device will be able to detect the vehicle's motion about the zz axis.

In this moving referential, the vector $\mathbf{H}$ can be expressed as $\mathbf{H} = H \cdot \hat{\mathbf{u}}_x$ where H is its magnitude and $\hat{\mathbf{u}}_x$ is the unitary vector along the xx axis. Then Newton's law can be written as

$$\mathbf{T} = \frac{d\mathbf{H}}{dt} = \frac{dH}{dt} \cdot \hat{\mathbf{u}}_x + H \cdot \frac{d\hat{\mathbf{u}}_x}{dt} = \Theta \left(\frac{d\Omega}{dt} \cdot \hat{\mathbf{u}}_x + \Omega \cdot \frac{d\hat{\mathbf{u}}_x}{dt} \right)$$

The first parcel in the right-hand side of the equation represents a variation of the magnitude of the wheel's spinning speed Ω caused by any possible accelerating or braking component of the torque along the xx axis. If the torque is perpendicular to the spinning axle, however, then the speed will be kept constant in magnitude, $d\Omega/dt = 0$, but the spinning axis will deflect in the direction of the torque. It is equivalent to saying that the direction ϕ of its axis $\hat{\mathbf{u}}_x$ will be rotated toward the yy axis with angular speed $\omega = d\phi/dt$, resulting in

$$\frac{d\hat{\mathbf{u}}_x}{dt} = \frac{d\phi}{dt} \hat{\mathbf{u}}_y = \omega (\hat{\mathbf{u}}_z \times \hat{\mathbf{u}}_x)$$

Substituting this relation in the former expressions and using the axial vector $\omega = \omega \cdot \hat{\mathbf{u}}_z$ to represent the angular turning speed of the entire frame (and of the vehicle) around the zz axis, it follows from Newton's law that $\mathbf{T} = \omega \times \mathbf{H}$.

The final conclusion is that, by keeping the wheel's spinning speed constant, the angular motion of the vehicle about the zz axis (called the rate input axis) causes the frame to move about the yy axis (called the output axis), under the influence of a torque proportional to the turning rate. The frame inclination about the output axis reaches an angular value α at which the restitution torque from the restraining springs, $T_r = K\alpha$, balances the gyroscopic torque, which is $T_g = H\omega$ (assuming that the turning rate and the angular momentum are kept perpendicular). A pointer fixed to the frame shows, on a scale fixed to the case, this angle α , which measures the turning rate ω about the input axis. At equilibrium $\alpha = S\omega$, where $S = H/K$, the device static sensitivity has the dimensions of time and is measured in seconds.

Note that these results are true only if the angle α is kept small enough to allow to consider approximately $\sin(\pi/2 - \alpha) = 1$. As the gyroscopic torque results from a vectorial product, its magnitude is indeed $T_g = H\omega \sin(\pi/2 - \alpha)$, which only can be

taken as $T_g = H\omega$ for a small inclination α . Using the approximation $\sin(\pi/2 - \alpha) = 1$ introduces a relative error ϵ , which is $\epsilon < 1\%$ if $\alpha < 8^\circ$, and $\epsilon < 1.5\%$ if $\alpha < 10^\circ$ in the expressions deduced.

Note also that if the gyroscope's frame is allowed to rotate unrestrained until the angle α reaches 90° , a singular situation is reached: The input axis and the spinning axis coincide at that point, the angle between them becomes null, and, therefore, the torque becomes $\mathbf{T} = \omega \times \mathbf{H} = 0$. Without torque, this situation will not be further modified, and all the orientation information will be lost. This situation is referred to as "the gimbals lock."

Instead of a pointer, a potentiometric position sensor may give an electric signal that measures the frame's deflection. Alternatively, the restitution springs may be replaced by restitution electric motors, usually called torquers, controlled by a feedback system that maintains the frame in its original position. These motors must generate a torque with the same magnitude as the gyroscopic torque, for which they require a proportional feeding current. Therefore, this current is the measure of the turning rate. The use of these motors has the advantage of measuring the torque without requiring a real displacement α , thus allowing the orthogonality between the input and the spinning axis to be maintained. Otherwise, the restraining springs require the displacement to generate the torque.

The operation of the device intrinsically involves a rotation movement of the frame about the output axis. Therefore, its dynamic motion is determined by the equilibrium of the acting gyroscope torque $T_g = H\omega$, the resisting spring torque $T_r = K\alpha$, the friction torque opposing the movement $T_f = D(d\alpha/dt)$, and the inertia torque $T_i = \Theta_y(d^2\alpha/dt^2)$, where D is the dumping coefficient and Θ_y is the moment of inertia of the whole set about the yy (output) axis. This leads to a second-order differential equation, the solution of which shows the possible existence of oscillations with a natural frequency $\omega_{a0} = (K/\Theta_y)^{1/2}$ and a dumping degree $\beta = D/(2(K\Theta_y)^{1/2})$. As is customary in moving measurement apparatus, the dumping degree is usually chosen to be $\beta = 0.707$, and D is computed to achieve this goal.

The preceding expressions show that a high sensitivity $S = H/K$ is obtained using a spinning wheel of high angular momentum, which must be achieved by spinning it at as high speed as possible. Reducing the spring constant K will also increase the sensitivity, but this will also reduce the resonance frequency $\omega_{a0} = (K/\Theta_y)^{1/2}$.

To obtain high angular momentum, the wheel must have a high momentum of inertia and a high spinning speed, but these goals must be traded off against excessive device mass and size, as well as the spinning motor power required. The momentum of inertia may be increased by the use of wheels provided with a thick rim around them. The wheel may be driven by a variety of motors. It can be connected to a turbine moved by a jet of compressed air or by combustion exhaust gases. In short-life gyros used in rockets, the hot gas jet may be produced by a pyrotechnic charge. The most common solution is an electric drive, ac as well dc motors being used. In some devices, the spinning wheel is the rotor of an induction motor.

The typical spinning speed for the inertia wheel is on the order of tens of thousands rpm. Rate gyros have typical rate resolution of 0.01° per second and range up to 2000° per second. Their natural frequency is on the order of some units to a few tens of hertz.

Rate-Integrating Gyros. The measurement of a vehicle's attitude can be made with a pair of double degree of freedom gyroscopes. However, spurious torques cause unwanted movements of precession and mutation of these gyroscopes, diminishing their accuracy. Therefore, it is usual to compute the attitude as a time integral of the turning rates about the three space axes.

This integration can be made electronically and is generally so done when the sensors are part of a wider measurement or control system. Sometimes, especially if only a visual indication is needed, a rate-integrating gyro is used. This rate-integrating gyro has the same structure as the rate gyro just described, except for the fact that it has no restraining springs. Instead, the dumping coefficient D of the frame rotation is increased. This increased dumping can be achieved electrically or mechanically (for instance, by making the frame rotation drive a paddle wheel rotating inside a viscous liquid).

Therefore, the acting torque developed has the same expression as in the former gyro, $T_g = H\omega$. But in the equation of its dynamic motion, $T_g = T_r + T_f + T_i$, the restraining-spring torque T_r is null, leaving $H\omega = D(d\alpha/dt) + \Theta_y(d^2\alpha/dt^2)$. In equilibrium, the acceleration $d^2\alpha/dt^2$ will be null, leaving the frame rotating at the speed $d\alpha/dt = (H/D)\omega$, proportional to the vehicle's turning rate. Except for initial values that can be zeroed at the beginning of the operation, the integrals of both rotating speeds will also be proportional. Therefore, the vehicle's attitude angle ϕ and the frame position angle α will keep the relation $\alpha = (H/D)\phi$. The constant (H/D) is the sensitivity of the rate-integrating gyro and increases with the inertia wheel's speed. The differential equation is of the first order, which means the device has no oscillations but has a time constant $\tau = \Theta_y/D$.

Vibration Gyros

These devices make use of pendulums, instead of rotating wheels, to conserve momentum. A pendulum without a transversal torque applied will conserve its plane of oscillation in a inertial reference frame—in the same way that a spinning wheel conserves the direction of its axis. If the plane of oscillation is forced to rotate about an axis coincident with the pendulum's equilibrium position, then a torque will result, which tends to bend the pendulum. Instead of free pendulums, vibrating cantilever beams are used. The rotation rate of the vibration plane causes bending of the beams. This generates a strain that can be measured and that is proportional to the turning rate and to the frequency of vibration.

This principle allows for the construction of very small and accurate vibrating gyroscopes. The cantilever beams used are bars a few centimeters or millimeters long. They have a piezoelectric material incorporated that, being fed by an electric power source, drives the vibration at a frequency on the order of tens of kilohertz. Other piezoelectric pieces are fixed to the other faces of the vibrating bar and act as strain gages to sense the transversal strain. Generally, these sensors are fixed to opposite faces of the beam and are connected in a bridge configuration for higher sensitivity to changes of strain and noise cancellation.

This technology allows for the production of very small and lightweight devices, with size on the order of $2\text{ cm} \times 3\text{ cm} \times 6\text{ cm}$ and weight about 40 g, including the necessary circuitry. Their maximum input is about 90 degrees per second, their output is on the order of $20\text{ mV/degree}\cdot\text{s}^{-1}$, and their linearity is on the order of 0.1% of the range.

Gas-Jet Rate Transducers

Gas-jet nongyro rate transducers are inertial devices, but instead of conserving the angular momentum as the gyros and vibrating beams do, they are based on the conservation of linear momentum: A particle thrown in free space will travel along a straight line. The particles they use are the molecules of a stream of gas, usually helium, which is blown through the top of a cylindrical metallic box by a nozzle that forms a laminar flow beam. The gas beam travels in a straight line to the opposite side of the cylinder, where it reaches an exhaust hole to which it is pumped and recirculated. The nozzle format and the pumping are crucial, as the travel time of the gas along the cylinder must be kept constant. Generally, the pump is a vibrating circular crystal driven by an external oscillator.

If the box is made to rotate about an axis perpendicular to the gas beam, the beam will no longer reach the same point of the opposite face. The gas will travel in a straight line with reference to an inertial frame but, as it travels, the target point will move in the sense of the angular rotation. Seen in the box referential, the gas beam seems to have been deflected in the lagging direction of the rotation. As the traveling velocity of the gas is kept constant, the lag of the beam is proportional to the speed of the rotation.

To sense the lagging of the beam, two equal and parallel wires are positioned, one at each side of the beam and equidistant to the equilibrium position of the beam. Both wires have equal currents, so they reach the same temperature. They are connected in a bridge configuration, so their voltages balance each other.

When the set is rotating, the gas blows preferably over the lagging wire than over the leading wire. Therefore, the lagging wire cools and its resistance decreases, while the leading wire heats up and its resistance increases. This unbalances the bridge and causes a voltage output that is a linear measure of the magnitude and direction of the turning rate.

These devices have no moving parts, so they are very rugged and have long operating lives. They are closed systems, having only electric contact with the outside world, and thus are adequate for operating in unfriendly environments, such as under water, in outer space, or in aggressive atmospheres. A variety of models and sizes are available, and small devices can be found. The cylinder can be a few centimeters in diameter, though its length is generally a little longer, and it can be as light as 60 g. A typical range is 10° to 5000° per second, with an output of about $0.01\text{ mV/degrees}\cdot\text{s}^{-1}$ to $0.5\text{ mV/degree}\cdot\text{s}^{-1}$, and linearity of 1% full scale.

Optical Gyros

The principle of operation of optical gyros is the intrinsic modification of some characteristics of light due to the motion of the medium in which the light travels (Sagnac effect). This modification is sensed through the observation of light interference patterns.

When a surface is illuminated by monochromatic light coming from a small light source, the distance the light must travel from the source to each point on the surface is point dependent. Therefore, the light reaches different points with different delays. Because the light is a sinusoidal oscillation of the electromagnetic field, the different delays cause the sinusoids to have different phases at the different surface points. If there is a second light source of the same wavelength illuminating the

same surface, the second light beam will also reach different surface points with different phases. At the points on the surface where the two impinging beams are in phase, the beams interfere constructively, their fields are added, and the illumination of those points is increased. At the points where the two beams are in opposition, the beams interfere destructively, their fields are subtracted, and the illumination of those points decreases. If both the frequency and phase difference of the two beam sources are constant, the zones of constructive and destructive interference stand still, and a stationary pattern of dark and bright fringes becomes visible.

If the phase of one of the sources is shifted relative to the other, the correspondent wave will strike each surface point with the same phase shift. Therefore (assuming that the shift was not an integer of 2π), at the points where the two light beams were in phase, they will cease so being and no longer interfere constructively. However, the two beams will meet together in phase at new points on the surface. As a consequence, the interference fringes move about the illuminated surface, their spatial shift being proportional to the phase shift that occurred at the light source. The measurement of these interference fringes' shift is at the heart of optical gyros, because the shift is due to the motion of the optical devices relative to an inertial reference frame, as shown next.

A beam of light will take a time $t = L/c$ to travel along a path of length L , where c is the speed of light in the propagation medium. If the medium itself is moving with velocity v in the same direction as the light is, then the end of the path will have traveled the displacement $dL = vt = Lv/c$ by the time the light reaches it. This is an extra length that the light must travel, and therefore the total length traveled by the light is $L_1 = L(1 + v/c)$. In an analogous way, if the light travels in a direction opposite to the movement of the medium, the total length is now $L_2 = L(1 - v/c)$. Two beams traveling in opposite direction will go through different distances, the difference being $\Delta L = L_1 - L_2 = 2Lv/c$. If the path is circular with radius r , then $L = 2\pi r$. And if the path is rotating about its center, in a fixed plane, with angular speed Ω , then $v = \Omega r$, and therefore $\Delta L = 4\pi r^2 \Omega / c = 4A\Omega / c$, where A is the area enclosed by the path. Although the preceding equation was derived for a circular path, it can be shown that the final result holds for any planar closed path. The difference ΔL can now be used in different ways to measure the motion of a body by exploring interference phenomena, originating the laser ring gyros (LRGs) and the fiber-optical gyros (FOGs).

Laser Ring Gyros. Laser ring gyros are usually used as attitude sensors. In the case of motion on a single plane, they measure the angle between an initial direction that is taken as a reference and the current direction of the main body axis of the moving vehicle or platform. At a particular instant, this angle is proportional to the number of wandering interference fringes that will have already passed in front of an optical detector, as described next.

In laser gyros, the light source is a gas laser tube, the most usual choice being an He-Ne gas laser. The tube emits in both directions and is inserted in a closed optical path. The path is a closed polygon, usually a triangle, made of ceramic glass. Mirrors at the corners of the triangle reflect the light around the corners. One of the mirrors is a partially transmitting mirror. The light coming from one direction of the path goes partially through this mirror directly into an optical interference

sensor. Light coming from the other direction crosses partially the mirror, is reflected back by a prism, and impinges on the same optical sensor.

The whole optical path works as a resonant cavity, tuned in such a way that two laser beams are established, both having a wavelength λ that is an integer submultiple of the length L of the path. The two beams travel in opposite directions.

When the whole path rotates, the beams travel different distances and therefore each beam is tuned to a different wavelength. The difference of the two wavelengths is proportional to the difference in the distance traveled: $\Delta\lambda/\lambda = \Delta L/L = 4A\Omega/(cL)$. Because $\lambda = cf$, it follows from the preceding expressions that $\Delta f/f = -\Delta\lambda/\lambda$, where f is the frequency of the light. Therefore, the rotation of the laser ring makes the two beams acquire a frequency shift $\Delta f = -4A\Omega/(\lambda L)$.

This frequency shift is very small, as the preceding equations show: In devices about 10 cm across, movements of one rotation per second result in $\Delta f/f \sim 10^{-9}$. With such a small frequency difference, for practical purposes the two beams may be viewed as having the same frequency but a nonconstant phase difference. Thus, they behave as if one of them is lagging the other by an uniformly growing phase difference $\Delta\phi = 2\pi\Delta f t$. The speed at which the phase lag increases is $2\pi\Delta f$.

The interference of the two beams at the receiving surface will show a fringe pattern that will be shifted by a distance proportional to the phase lag. Therefore, as the phase lag uniformly increases, the pattern will wander over the receiving surface. The wandering velocity of the fringes over the surface is proportional to the rate of the phase difference. Over a particular point, a new fringe passes each time the phase difference $\Delta\phi$ increases by 2π . Therefore, the fringe rate, expressed in the number of fringes over a point per unit time n_f/s , equals Δf . The final result is that the wandering rate of the fringes is proportional to the rotation speed of the optical laser gyro with respect to an inertial frame; that is,

$$\frac{n_f}{s} = \frac{4A}{\lambda L} \Omega$$

The interference sensing device consists of an optical sensor, typically a photodiode, that senses the bright fringes and generates an electric pulse for each fringe that passes over it. The wandering rate of the fringes can be measured by counting the number of pulses per unit time, and this provides a measure of the angular speed of rotation. However, the most usual utilization of laser ring gyros is for measuring angular position, not angular speed. Since the angular position α is the time integral of the rotation speed, α can be determined from the number N of fringes that have passed over the diode since the beginning of operation. Thus, by simply maintaining a permanent account of the fringes that have passed, the angular position is measured. A double photodiode is needed, with both detectors mounted in quadrature, to sense the wandering direction of the fringes and therefore properly increment or decrement the counter.

From the preceding equations, it can be seen that $\alpha = N(L\lambda/4A)$. The factor $S = L\lambda/4A$ is called the nominal scale factor of the ring laser gyro. For an He-Ne laser ($\lambda = 633$ nm) and a triangular ring of 10 cm side length, $S = 1.1 \times 10^{-5}$ rad/pulse, or $S = 2.3(\text{arcsec})/\text{pulse}$.

Laser ring gyros are very sensitive devices; they are small and light. They have a fast warm-up and rapid response, and they are insensitive to linear motion. They also are free from

multiaxis cross-coupling errors. Their price is decreasing and is competitive with the price of other types of gyros. Therefore, laser ring gyros tend to replace other inertial attitude sensors in air- and spacecrafts.

Rate Fiber-Optic Gyros. Fiber-optic gyros are commonly used as rate gyros; the phase difference between two light beams provides a measurement of the turning angular velocity of a platform (this is the principle of the interference fiber-optic gyros, or IFOGs). However, the phase measurement cannot be simply executed by the procedure of counting interference fringes, as will be explained next. The new measurement procedures can lead to resonant and Brillouin fiber-optic gyros.

Fiber gyros are a variant of the Sagnac ring interferometer. A light beam from a laser source is divided in two beams by a beam splitter. The two new beams enter in opposite extremes of an almost closed path (in former Sagnac rings, a polygon with mirrors on the corners was used), travel in opposite directions around the path, and come back to the beam splitter. The splitter combines the two arriving beams and projects them on the screen of an interferometer, where the beams interfere, creating an interference fringe pattern.

The two beams travel equal distances, so they are in phase when they interfere. If the closed path rotates on its plane, one beam travels a longer distance than the other, and therefore its phase on arrival will lag the other beam's phase.

Notice that the two beams result from the splitting of an already existing beam generated by a single laser source. Therefore, contrary to what happens with the laser ring gyro, traveling different distances causes no frequency shift between the two beams, but only a phase shift. The shift between the beams' phases causes a stationary shift of the position of the interference fringes on the plane of the screen (and not the permanent wander of the fringes, as is the case for the ring laser's frequency shift).

As seen before, the difference of the paths' length is $\Delta L = 4A\Omega/c$, where A is the area enclosed by the closed path. With the same wavelength for the two beams, this difference in distance corresponds to a difference of phase $\Delta\phi = 2\pi\Delta L/\lambda = (8\pi A/c\lambda)\Omega$. As a result, the phase shift is proportional to the rotating rate of the gyro, Ω . However, for small devices with about 10 cm across, the proportionality constant is very small, on the order of 10^{-4} rad/rad·s $^{-1}$, causing the phase shift $\Delta\phi$ to be very small when the rotation rate Ω is kept moderate.

One way of increasing the sensitivity is to increase dramatically the length L of the path, or the area A it encloses, but without increasing the size of the device. To achieve this purpose, the path is made with an optical fiber wound as a coil of several turns. Modern silicon fibers present an attenuation on the order of 0.2 dB/km, at a wavelength of 1550 nm. Therefore, several hundreds or even thousands of meters of optical fiber can be used, almost without attenuation. The fiber is wound a number N_t of turns, and the total area enclosed to be considered in the equations is now $A' = N_t A$, where A is the area of the surface physically occupied by the device, which remains on the order of 10 cm or 20 cm in diameter. However, noise and other sources of signal perturbation increase with the length of the fiber. Even when special techniques are used to minimize these effects, they end up putting a limit on the maximum length of the fiber.

The rotating rate Ω cannot be measured by counting the number of fringes that shift in front of a photodiode, as was

done in the case of the laser ring. Indeed, the distance of two consecutive fringes on the screen corresponds to a phase shift of $\Delta\phi = 2\pi$. For example, considering a one-coil path with a diameter of 10 cm and a wavelength $\lambda = 1550$ nm, that phase shift corresponds to the astounding turning rate of 15000 rad/s per fringe. Therefore, it is necessary to measure phase shifts much smaller than 2π by any method other than a simple digital count of the fringes.

In the fringe pattern, the light intensity does not vary abruptly between the bright and dark fringes. Instead, the light intensity has a sinusoidal variation between the fringes. Sensing both the illumination intensity and its gradient around a point is a way of detecting the position of that point with respect to the fringe pattern and therefore computing the phase shift. Many processes of detecting and measuring very small phase shifts, on the order of 10^{-7} rad, have been developed and implemented, some of them using sophisticated modulation techniques. Some of these modulation techniques rely on using some external physical action to modify the fiber's light propagation characteristics in a manner that these characteristics become direction dependent. Being nonreciprocal (direction sensitive), this additional action increases the difference in the behavior of the two beams, hence facilitating the measurement of the length variation ΔL .

One of the techniques consists of having two rings optically coupled so as to form a multipass optical resonator. One of the rings is a small rotating ring of fiber (only about 10 m long) that acts as rotation sensing coil. The Sagnac effect changes the frequency of resonance by an amount Δf proportional to the rotation rate Ω , and the frequency shift is measured (the equation for Δf is the same that was determined for the laser ring gyro). These devices are called resonance fiber-optic gyros (R-FOGs).

Another technique is based on the fact that mechanical vibrations (acoustic waves) can interact with light waves, causing a frequency shift of the light (Brillouin effect). This effect is induced in such a way that the two beams are affected differently. The result is that the two beams acquire different frequencies, behaving in a way similar to the laser ring gyros, although the physical mechanism is different. Devices based on this technique are called Brillouin fiber-optic gyros (B-FOGs).

Using fibers up to a thousand meters, and accurate interferometer methods of reading phase shifts, a sensitivity of up to the order of 10^{-2} degrees per hour is reached for I-FOGs, which is adequate for inertial navigation systems for aircraft. This kind of gyro is currently in use in commercial airplanes. R-FOGs and B-FOGs can have a sensitivity on two or three orders of magnitude higher.

BIBLIOGRAPHY

1. G. Asch and P. Desgoutte, *Les Capteurs en Instrumentation Industrielle*, Cap. 9, "Capteurs Tachymétriques," Dunod, France; Georges Asch, 1987.
2. A. Boros, *Electrical Measurements in Engineering*, New York: Elsevier, 1985.
3. S. Ezekiel and H. J. Arditty, *Fiber-Optic Rotation Sensors and Related Technologies*, New York: Springer-Verlag, 1982.
4. P. Hariharan, *Optical Interferometry*, New York: Academic Press, 1985.

5. P. Hauptmann, *Sensor: Principles & Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1991.
6. K. Hotate, J. Dakin, and B. Culshaw, in *Optical Fiber Sensors Applications, Analysis, and Future Trends*, New York: Artech House, 1997.
7. R. T. H. Kersten, J. Dakin, and B. Culshaw, in *Optical Fiber Sensors: Principles and Components*, New York: Artech House, 1988.
8. A. S. Morris, *Principles of Measurement and Instrumentation*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
9. H. N. Norton, *Sensor and Analyzer Handbook*, Englewood Cliffs, NJ: Prentice-Hall, 1982.
10. R. Pallás-Areny and J. G. Webster, *Sensors and Signal Conditioning*, New York: Wiley, 1991.
11. J. Powers, *An Introduction to Fiber Optic Systems*, 2nd ed., Chicago: Irwin, 1997.
12. T. Senbon and F. Hanabuchi, *Instrumentation Systems: Fundamentals and Applications*, New York: Springer-Verlag, 1991.
13. B. Stieler and H. Winter, *Gyroscopic Instruments and Their Application to Flight Testing*, AGARD Flight Test Instrumentation Series, Vol. 15, North Atlantic Treaty Organization.
14. D. Wobschall, *Circuit Design for Electronic Instrumentation, Analog and Digital Devices from Sensor to Display*, 2nd ed., New York: McGraw-Hill, 1987.

MÁRIO VENTIM NEVES
Technical University of Lisbon

FIBER-OPTIC SENSORS

Numerous parameters of light guided by optical fibers or transmitted by optical devices may depend strongly on the external environment. In some applications, this dependence is not welcome, as with fiber-optic telecommunications systems, where, for instance, attention is typically directed to avoiding bend- or stress-induced attenuation in optical cables. In other applications, however, this process may purposely be amplified, usually to make a specific light parameter selectively sensitive to a chosen environmental measurand. When optical fibers are used either to transmit the modulated light signals or to modulate some parameters of these signals directly via a specific physical phenomenon, the process is known as *fiber-optic sensing* (FOS) and the corresponding devices are known as *fiber-optic sensors*.

The need for fiber-optic sensors in the broad area of engineering is now widely accepted, especially in those applications where the benefits offered by this still novel technology are already more significant to the users than the higher costs of fiber-optic instrumentation in comparison with other sensor systems. One such application area is monitoring of civil engineering structures, including highway and railway bridges, dams, tunnels, large buildings, and underground mines. Such structures must resist environmental and in-service loads due to winds, earthquakes, traffic, thermal effects, construction, or environmental damage. For safety and maintenance their response must be carefully monitored. This information is needed to assess their overall "health" and to increase engineering knowledge for future projects. Pressure and strain gauges, stress cells, extensometers, accelerometers, and tiltmeters are among the variety of electrical and nonelectrical transducers that can be directly embedded in concrete, earth, rock, bore-hole or tunnel lining to sense and monitor structural loads and responses. None are well suited, however, to more demand-

ing environments. Nonelectrical hydraulic devices are costly in labor-intensive inspection, maintenance, and repairs; they are difficult to multiplex and impossible to use for dynamic measurements. Electrical measuring devices are out of the question in open-air structures (risk of lightning) and in mining environments (risk of explosion). For these and similar applications, fiber-optic sensor technology offers strong potential for significant metrological improvement: electrical passivity, high bandwidth, safety in corrosive or explosive environments, immunity to electromagnetic interference, high sensitivity, miniature dimensions, possibility of remote operation, and direct compatibility with increasingly common fiber-optic data transmission and communication networks.

BASIC PRINCIPLES

A fiber-optic sensor is a device capable of converting a given, variable physical quantity into a modulated optical signal and then decoding it as a normalized electrical signal. For measurement purposes, this output signal is usually calibrated against a primary standard of the physical quantity of interest. The sensor then becomes a true secondary measuring instrument of this quantity. Figure 1 shows a schematic view of a basic fiber-optic sensor, where the sensing element S can be an optical fiber or another external optical element.

Definition

The sensing element S of the optical fiber sensor shown in Fig. 1 can best be formally described by the propagation matrix $M(\mathbf{V}, \lambda)$ which normally depends on the physical environment represented by the vector $\mathbf{V}$ (depending on every imaginable external parameter) and on the spectral properties of the transmission medium (λ is the wavelength of propagating light beam). We may then describe the action of the sensor using the following equation:

$$\mathbf{E}^M(\lambda) = M(\mathbf{V}, \lambda)\mathbf{E}(\lambda) \quad (1)$$

where $\mathbf{E}(\lambda)$ is the wavelength-dependent input electrical field entering the sensor, and $\mathbf{E}^M(\lambda)$ is the measurand-modulated output electrical field exiting the sensor. By applying appropriate optical and electronic signal processing we can then recover M , $\mathbf{V}$, and finally those components of $\mathbf{V}$ that depend on the specific quantity we intend to measure. It is worth noting that $M(\mathbf{V}, \lambda)$ can conveniently be represented by a product of three separate factors as

$$M(\mathbf{V}, \lambda) = T \exp(i\phi_m) B \quad (2)$$

Each of the three factors on the right side of the above equation is both environmentally sensitive and dispersive (dependent on the wavelength of transmitted light). The first factor, T , is the scalar transmittance, the parameter ϕ_m is the mean phase

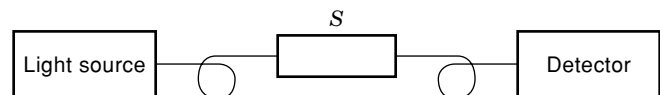


Figure 1. Basic configuration of a fiber-optic sensor; S denotes a sensing element.

retardance, and finally B is the birefringence matrix characterizing the sensing element.

Classification

One of the most straightforward classifications of fiber-optic sensors can be derived directly from Eq. (2): *intensity sensors* are those where the measurand-modulated parameter is the scalar transmittance T , and consequently the intensity of transmitted light; *interferometric sensors* are those based on phase modulation; and, finally, *polarimetric sensors* are devices based on birefringence modulation. Both phase and birefringence effects may be recovered using interferometry or polarimetry, depending on specific application. This of course is a somewhat simplified approach; in reality all those parameters may be simultaneously affected, but one type of modulation usually predominates.

Fiber-optic sensors are often classified according to the nature of the sensing element: in *intrinsic* sensors, this element is an optical fiber, and the light is never permitted to exit the fiber into free space; in *extrinsic* sensors, it is a completely different optical part, and consequently in these hybrid structures the fibers are used exclusively for light-guiding purposes. In addition to standard sensors designed to measure a given parameter in one particular location only, we often mention *multipoint* or *multiplexed sensors*, which make it possible to take measurements in several different locations. Finally, it is worth noting that no other classical sensor systems have the same capability as fiber-optic systems to serve as truly *distributed sensors* and to enable the simultaneous detection not only of the magnitude of the modulating factor but also of the precise location of its occurrence and even of its spatial distribution.

Depending on the type of propagation occurring in the optical fibers, it is customary to differentiate between *single-mode* and *multimode sensors*. Depending on the degree of temporal coherence of the light sources involved, we can categorize sensors as *high-coherence*, *low-coherence*, or *incoherent*. In the rest of the paper, we will rely primarily on the notions introduced above, although many other less important approaches to classifying fiber-optic sensors, such as those based on the particular technology or on the application area, can be found in Ref. 1.

FIBER-OPTIC SENSOR COMPONENTS

Many FOS systems are based on standard commercially available components developed primarily for applications in optical fiber telecommunications. However, specifically designed and/or modified optical and fiber elements are also in use, since they can significantly improve the sensitivity and selectivity of

the sensor and the stability of the system. Some examples of the components most often used in optical fiber sensing systems are described below.

Optical Fibers

Different types of FOS require different optical fibers for the transmitting, sensing, and receiving waveguides of the measuring system. For *extrinsic intensity* FOS, standard cylindrical fibers are normally used. The light signal can be intensity-enhanced when large-core multimode fibers or fiber bundles are applied for light-guiding purposes. A similar situation occurs with the fibers used to lead in and lead out the light in *intrinsic intensity* sensors. Specially prepared optical fibers such as doped fibers; hollow-core fibers filled with metal, liquid, gas or liquid crystals; and fibers with partially removed cladding or D-shaped cladding are often used as sensing elements. Since the lasing properties of erbium-doped fibers depend on various external parameters (such as temperature, stress, pressure, etc.), these fibers also can be used as sensing parts of the FOS despite the fact that they were developed originally for applications in optical fiber lasers and amplifiers.

For *polarimetric* and *interferometric* FOS, the methods of signal detection usually require a well-defined state and degree of polarization of the guided light. Therefore, optical fibers designed for these applications should have well-defined birefringence. In ideal cylindrical fibers [using the description of linearly polarized (LP) modes], the fundamental LP_{01} optical mode contains two degenerated orthogonally polarized modes propagating at the same phase velocity. In real fibers, both these modes have randomly different phase velocities, causing fluctuations of the polarization state of the light guided in the fiber. In highly birefringent (HB) fibers (2), the phase velocities for the two orthogonal light polarizations are high enough to avoid coupling between these two modes. Hence, if only one mode is excited, then the polarization of light guided in the fiber is preserved. On the other hand, if both modes in highly birefringent fibers are excited, then the polarization state of the light varies along the propagation distance. The state of the polarization depends on the phase difference $\Delta\phi$ between the two orthogonal linearly polarized modes. Since both these modes "see" the different refractive indices n_x and n_y , the phase difference between them changes at the distance z according to the formula

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta n \cdot z \quad (3)$$

where $\Delta n = n_y - n_x$. Figure 2 shows the changes in the polarization states corresponding to the phase difference and to the propagation distance. These changes are periodical, with the

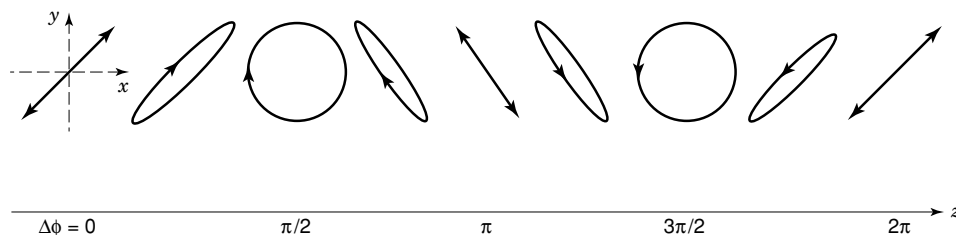


Figure 2. Evolution of the polarization states along the propagation axis in a HB fiber; input polarization is linear at 45° to the principal birefringence axes.

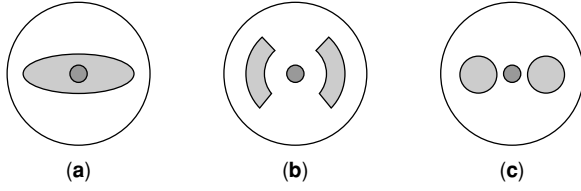


Figure 3. Cross sections of three HB fibers: (a) elliptical, (b) bowtie (stress-induced), (c) side-hole (stress-induced).

spatial period defined as the beat length L_B :

$$L_B = \frac{\lambda}{\Delta n} \quad (4)$$

The birefringence of the fiber expressed by Δn is strongly influenced by external forces, and consequently the state of the polarization at the output fiber also depends on the local birefringence changes. This transduction mechanism is utilized in *polarimetric* sensors and will be explained later. High birefringence in optical fibers can be obtained by designing a core and/or cladding with noncircular geometry, or by using optically anisotropic core or cladding material (with intrinsic or induced anisotropy). Several cross sections of typical highly birefringent optical fibers are presented in Fig. 3. While in standard optical fibers the average beat length varies from 0.1 m to 1 m, in highly birefringent fibers it is only a few millimeters. Table 1 shows typical beat lengths at wavelength $\lambda = 830$ nm for commercially available highly birefringent fibers.

Other types of specialty fiber are low-birefringence (LB) fibers with birefringence lower than in standard optical fibers, and circularly birefringent fibers. Both of these fiber types are applied in sensors for magnetic field or electrical current measurements, where the magnetic field modifies the circular birefringence (Faraday effect). Another specialty fiber, recently introduced to the market by 3M, is the so-called polarizing fiber, which at a specific wavelength has a significantly different attenuation for two orthogonally polarized modes. After nonpolarized light is launched into 2 m to 3 m of such fiber, only the light in the low-loss polarized mode is transmitted and the mode with orthogonal polarization is completely attenuated (an extinction ratio of 1:1000 can easily be obtained). This type of fiber is designed with an absorbing metal layer placed close to the fiber core, or as a bent birefringent fiber. The polarizing effect is observed in a specific wavelength region only. Typical spectral characteristics of the attenuation in polarizing fibers are presented in Fig. 4.

Optical Sources and Detectors

In almost every FOS application, semiconductor devices are used as sources and detectors. Other sources and detectors are

Table 1.

Fiber Type	Beat Length L_B (mm)	Birefringence Δn (10^{-3})
Andrew E-type	5.08	0.16
Andrew D-type	4.71	0.18
Corning	2.99	0.28
Fibercore bowtie	2.10	0.39

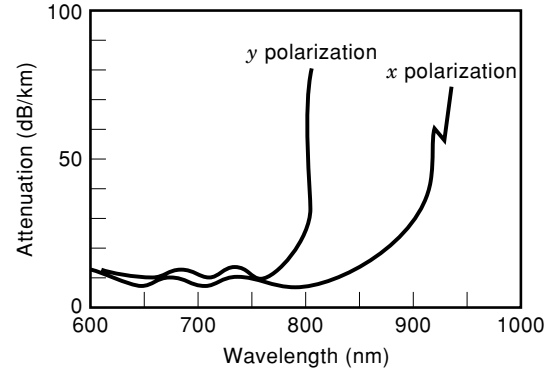


Figure 4. Spectral characteristics of the attenuation of two perpendicularly polarized eigenmodes propagated by a polarizing fiber. This fiber should be operated at about 830 nm for best performance.

occasionally used in specialized FOS for which the required parameters of light are not yet available from semiconductor sources (as in the case of sensors measuring luminescence with short-wavelength light sources).

The chosen light source for a specific sensor configuration should have the required stable output power, as well as the required light wavelength and coherence, which is especially important for polarimetric and interferometric systems. The coherence of the light source is described by the spectral linewidth $\Delta\lambda$ or equivalently by the coherence length $L_c = \lambda^2/\Delta\lambda$. The coherence length L_c is the maximal difference of optical path length at which the light can interfere. When the light is divided into two different beams propagating in different fibers then these two beams can interfere according to the phase difference

$$\Delta\phi = \frac{2\pi}{\lambda}(n_1L_1 - n_2L_2) \quad (5)$$

where $L_{1(2)}$ is the length and $n_{1(2)}$ is the effective refractive index of the fibers. The interference can be observed only if the phase difference $\Delta\phi$ is smaller than the phase change at the coherence length:

$$\Delta\phi \frac{2\pi}{\lambda} L_c \quad (6)$$

Similarly, the state of polarization of the light propagating in a birefringent fiber can be defined only if the phase difference between the two orthogonal linearly polarized modes [Eq. (3)] meets the conditions of Eq. (6). Otherwise, the light exiting at the fiber end is depolarized and polarization-based signal processing is clearly not possible.

The coherence length for semiconductor sources varies from micrometers for light-emitting diodes (LEDs) to a few meters for single-mode diode lasers (LDs). The typical coherence of multimode LDs is a few millimeters, while superluminescent diodes (SLDs) have a coherence length of the order of a few tens of micrometers. High coherence is not always required for interferometric and polarimetric measuring systems. Low-coherence light sources are often intentionally used in certain interferometric FOS to reduce signal errors due to the presence of unwanted interference fringes (as in optical fiber gyroscopes). Also, in coherence multiplexing techniques the coherence

should be low enough to avoid interference between the different measuring points. Therefore SLDs or edge-emitting light-emitting diodes are typically used in these systems. Recently even erbium-doped fiber lasers (with high-power 980 nm or 1480 nm LD pumps) began finding application in this area due to the facility of their coupling with other fibers. They are used as superluminescent fiber sources in interferometric systems (3) or fiber-optic amplifiers in FOS networks (4).

The important requirements of optical detectors used in FOS are bandwidth, noise, and dynamic range. The bandwidth determines the detector response time, which can be very long for low-cost discrete sensors but for certain types of distributed sensors and for time-domain multiplexed sensor systems should be very short. Typically, single detectors such as *pin* photodiodes, avalanche photodiodes, and photoconductors are still often used, but fast charge-coupled detector (CCD) lines and matrices are seeing greater use due to their increasing availability. The spatial light distribution or changes in it registered by the CCDs can be directly decoded as a measured quantity or may be further processed in demultiplexing techniques for distributed FOS.

Optical Fiber Components

Such elements as beamsplitters, wavelength filters, wavelength analyzers, polarization controllers, polarizers, phase and intensity modulators, and amplifiers are frequently used in optical sensor systems for optical processing or delivering the light beam to a specific location. All these elements are increasingly manufactured in optical fiber technology (*all-fiber* components) rather than in bulk optics (5) in order to obtain more mechanical and thermal stability as well as to satisfy the typical requirement for low optical losses in sensing systems.

One of the most frequently used all-fiber components is the *directional coupler* (6), which plays the role of a wavelength-selective beamsplitter. A schematic drawing of the bidirectional coupler is presented in Fig. 5. The directional coupler consists of two optical fibers placed close enough to each other to allow tunneling of guided light between the two fiber cores. The input light intensity I_i is partially transferred to both outputs I_{o1} and I_{o2} . For two identical lossless fibers, the energy transfer between them is described by the formulas

$$\begin{aligned} I_{o1} &= I_i \cos^2 \kappa L \\ I_{o2} &= I_i \sin^2 \kappa L \end{aligned} \quad (7)$$

where L is the effective coupler length. For parallel fibers, L

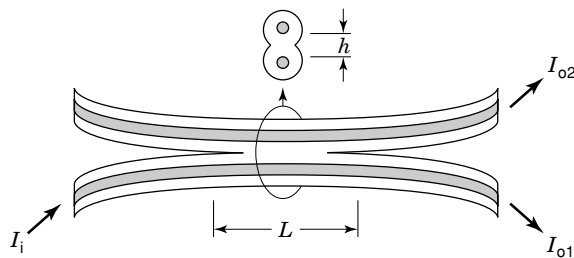


Figure 5. Schematic view of a bidirectional fiber coupler and its cross-section. I_i is input light intensity, I_{o1} and I_{o2} are output light intensities.

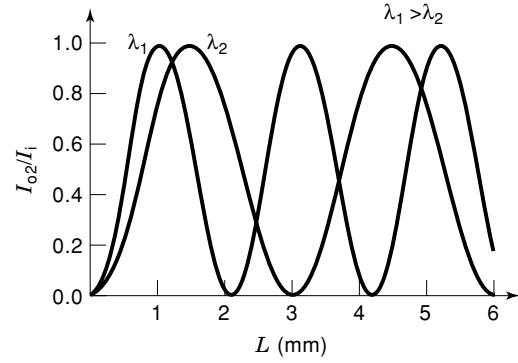


Figure 6. Power transmission ratio for a bidirectional coupler as a function of its effective length at two different wavelengths of light.

is simply the length of the coupling region; for curved fibers with a bend radius R , it is given by $L = (\pi R/\gamma)^{1/2}$. The coupling coefficient κ is estimated in terms of coupled mode theory as an integral over the coupler cross section:

$$\kappa = \frac{\pi}{\lambda n} \int (n_c^2 - n_{cl}^2) e_1 e_2 dS \quad (8)$$

where n is the effective refractive index, $e_{1,2}$ are the normalized field envelopes of the guided modes in the fibers, and n_c and n_{cl} are the refractive indices of the fiber core and the surrounding cladding. In the simplest approximated form, coupling coefficient κ depends on the fiber's core separation distance h as follows:

$$\kappa = \frac{\lambda}{2\pi n} \frac{v^2 - \gamma^2}{a^2 - v^2} \frac{K_0(\gamma h)}{K_1^2(\gamma a)} \quad (9)$$

where $v = 2\pi(n_c^2 - n_{cl}^2)^{1/2}/\lambda$, $\gamma = 2\pi(n^2 - n_{cl}^2)^{1/2}/\lambda$, a is the fiber core radius, and K_0 and K_1 are modified Bessel functions. For large core separations $\gamma h \gg 1$, the coupling coefficient is proportional to $\pi/2\gamma h e^{-\gamma h}$.

Note that the energy transfer between the two fibers depends on the length of the directional coupler, on the distance between the fibers, and on the wavelength of the light. This fact allows for convenient design of a directional coupler with the required transmission at the given wavelength. Figure 6 shows that by choosing the effective length of the directional coupler it is possible to obtain almost any power transmission ratio for two different wavelengths. Therefore the directional coupler can play the role of a bidirectional beamsplitter, a wavelength filter, or a wavelength-division multiplexer/demultiplexer. Directional couplers are most often made by polishing the fibers close to the core, by thermal fusion of two fibers, or by chemical etching. Polarization-maintaining couplers using highly birefringent fibers are also manufactured for specific applications. Figure 7 shows three examples of coupler use in fiber-optic sensor systems.

Fiber polarization controllers are often needed in sensing systems dealing with polarized light. The polarization controller transforms any given state of polarization into another state that is required in the particular application. This can be done either by modifying the fiber birefringence or by changing the birefringence axis. The first method is applied in controllers based on electromagnetic squeezing of the fiber. The

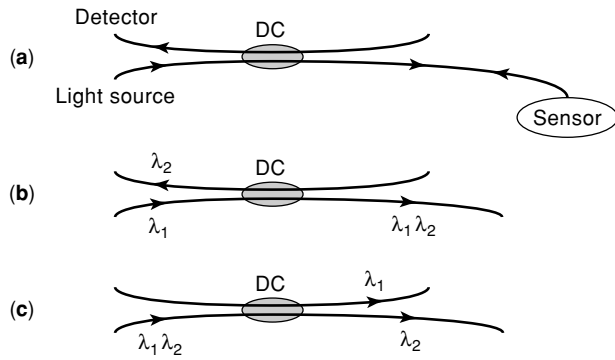


Figure 7. Three examples of possible applications of fiber directional couplers (DC) in fiber-optic sensor systems: (a) as a beamsplitter, (b) as a wavelength multiplexer, (c) as a wavelength demultiplexer.

external stress induces proportional birefringence change in a fiber. Therefore the controlled squeezing produces a required value of birefringence in the fiber and consequently the required phase shift between two perpendicular polarizations. The state of polarization has two degrees of freedom, and therefore two such squeezers rotated in relation to one another at a 45° angle are required for polarization controller.

Another type of polarization controller is based on changes of the birefringence axis of two elements playing the roles of a $\lambda/4$ plate and a $\lambda/2$ plate (three $\lambda/4$ plates can also be used for this purpose). The $\lambda/4$ plate introduces a $\pi/2$ phase retardation between two orthogonally polarized waves, while the $\lambda/2$ plate introduces a π phase shift. In the simplest and the most commonly used configuration, both plates are designed as coiled fibers. In bent fibers, the stresses induce birefringence and selection of the appropriate radius of curvature can produce the required phase retardation. Both fiber coils can be rotated relative to the z axis. The desired state of polarization at the output is obtained by adjusting the angle between the coils. Optical polarizers rather than polarization controllers are required in systems with elements depending on the polarization state. In all-fiber systems, often the polarizing fibers play the role of such polarizers, the extinction ratio increasing with the fiber length. Fiber polarizers can also be manufactured as polished fibers with the metal layer attenuating one polarization.

Intensity modulators, phase modulators, and frequency shifters are fiber devices frequently used in sensor systems to improve the performance of their detecting systems. Fiber-optic phase modulators utilize the refractive index changes caused by stretching or squeezing the fiber, and polarization controllers based on electromagnetic squeezing can often be used for these purposes. Typically, the phase modulators are designed as devices using an optical fiber wrapped around a piezoelectric ring, which changes its dimensions, and consequently the fiber length, in response to externally applied voltage. The frequency shifters, on the other hand, are manufactured primarily in integrated optics technology. The integrated optical modulators most widely used are based on the electro-optical Pockels effect occurring in lithium niobate crystals in channel waveguides surrounded by electrical electrodes. The coupling between the integrated channel waveguide in lithium niobate and the external leading optical fibers is inevitably associated with additional losses in the optical system and constitutes an obvious drawback; an important advantage, however, of the integrated

optical chip is that numerous other elements such as integrated polarizers, directional couplers, or polarization controllers can be simultaneously designed and manufactured on the chip.

FIBER-OPTIC SENSORS: OVERVIEW

In this section we consider a representative sample of the most recent and successful embodiments of fiber-optic sensing devices. This overview, however, is by no means exhaustive; researchers worldwide have studied and explored a large number of sensing concepts during the last 15 years. Books such as Refs. 7 and 8 are suggested for anyone wishing to study this subject in more detail.

One focus of early fiber-optic sensor research was periodic measurands, but this work often ignored the issues of stability, repeatability, and temperature drifts and has not led to many practical installations. For this review, we have focused on practicability and on the potential for real implementations, which lead to an emphasis on absolute and quasistatic techniques. A significant majority of the conventional electrical and hydraulic measurement techniques that fiber-optic sensing technology is trying to outperform are in fact absolute and quasistatic, with countless applications including industrial process control as well as stress analysis in civil engineering and geotechnics.

Intensity Sensors

Fiber-optic sensors based on intensity modulation are the simplest sensors, but they are often very successful and cost-effective devices. They can be developed using any kind of optical fiber, and noncoherent sources and simple detector units are usually adequate as other components of these systems. Most research effort in the area of intensity sensors has focused on overcoming one major drawback: finding a reliable referencing technique to compensate for a variety of analog signal instabilities. These instabilities include intensity and wavelength fluctuations of the light source and uncontrolled changes in detector responsivity as well as environmentally sensitive losses in optical fibers, couplers, and connectors.

Loss-Based Fiber-Optic Sensors. In bent optical fibers, higher-order modes tend to leak to the cladding when the bending radius decreases. This effect can be purposely amplified by applying a periodic microbending structure on an optical fiber, with the degree of induced attenuation depending upon the action of the parameter to be measured. If the bending is periodic with the bend pitch p , that is

$$p = C\pi r n (NA)^{-1} \quad (10)$$

where NA is the numerical aperture of the fiber, n is the refraction index of the fiber core, r is the core radius, and C is an index profile factor, then (9) the resonance condition can be achieved when the attenuation is maximum. Figure 8 shows a typical structure of a microbending force sensor using multi-mode fiber: in such a device, an external perturbation increases the bending of the fiber and causes progressive coupling of light energy into the radiation modes leaking out of it.

Based on this principle, many practical embodiments of the microbending sensor have been studied and developed. A good and timely review by Berthold (10) gives a complete analysis of this technology. One successful microbending sensor developed

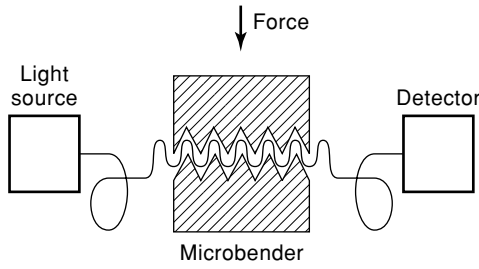


Figure 8. Basic microbending fiber-optic force sensor.

just recently (11) uses a multimode fiber as a sensing element and single-mode fiber as a lead-in. Such a configuration makes it possible to reduce noise significantly and to use more coherent sources; more importantly, it allows for a sixfold increase in sensitivity over microbending sensors built entirely of multimode fiber.

Like other intensity-based sensors, microbenders are sensitive to the optical power level in the system, and a large amount of research has been devoted to finding adequate referencing techniques to alleviate the problem. This has not been fully successful, since the possible origins of power fluctuations in the optical system are numerous and varied. Changes in output of the optical source, changes in transmission and bending loss along the fiber links, changes in coupling between optical connectors and other optical elements external to the sensor, and finally changes in the modal filtering properties of system components can all be responsible for these perturbations.

One successful self-referencing method, based on time-division normalization, was introduced by Spillman and Lord (12). The method is based on a 2×2 coupler and on introducing a specific time delay between the signals reflected back from the sensing and referencing arms and arriving back at the coupler. The ratio of the two signals is relatively independent of external perturbation sources, provided the time delay is short. Another similar method based on wavelength normalization utilizes a broadband source with two narrow spectral channels, dedicated to the reference and to the sensor signal, respectively (10).

The reported applications of microbending sensors include measurement of pressure (13,14) and temperature (15), as well as acceleration (16), strain (17), and speed (18). The main advantage of microbenders is their relatively high dynamic range of about 40 dB; the reported sensitivities go up to $10^{-10} \text{ m/Hz}^{1/2}$ (10).

Evanescent Sensors. An interesting, quite simple, and surprisingly effective group of sensors is based on the phenomenon of frustrated total internal reflection, where light energy is extracted from the propagating beam in the form of an evanescent wave. This extraction occurs in a predesigned location (a fiber element stripped of cladding or a separate optical element put in contact with a surrounding medium with a higher, or measurand-dependent, index of refraction).

Figure 9 shows a liquid-level sensor developed by Raatikainen et al. (19), where light delivered via a multimode input fiber undergoes two consecutive total internal reflections while passing through a sensing tip and then returns to the

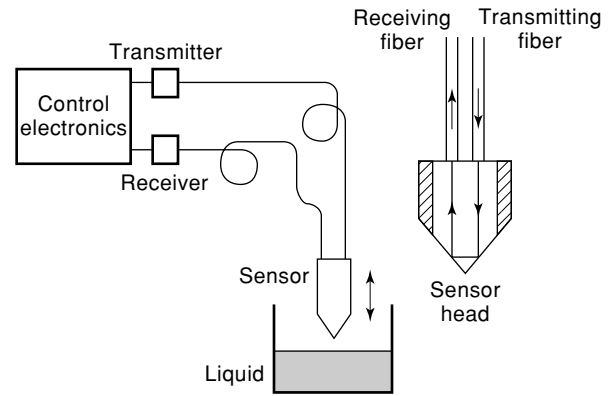


Figure 9. Schematic view of a fiber-optic liquid-level sensor. The magnification shows the light path within the sensor tip.

detection electronics via a multimode output fiber. When the sensing tip hits a liquid, the conditions for total internal reflection no longer exist and the output signal is attenuated at the level of 10 dB due to the evanescence of higher-order modes.

A more advanced configuration of a digital level transducer has been recently developed by Betta et al. (20). It is based on an optical fiber extended over the whole depth of a tank, with the optical cladding removed in 40 zones 0.5 mm long and with 25 mm of spacing between them. Each time the liquid level reaches or leaves one of those zones, the change in output power and the direction of this change are registered, giving information sufficient to determine the absolute value of the liquid level. A prototype with 1 m range and 25 mm resolution was manufactured and tested. It displayed no hysteresis, good repeatability, and accuracy of 25 mm.

Absorption Sensors. It is well known that the energy band structure in some semiconductor materials, especially in gallium arsenide (GaAs), depends strongly on the parameters of the external environment. Several successful approaches to the design of fiber-optic temperature and pressure sensors (21) rely on this effect, and some of them have already been commercialized (22).

The working principle behind this type of sensor is the displacement of the wavelength transmission limit in a GaAs crystal induced by pressure and/or temperature change. Pure GaAs crystal is transparent to infrared radiation and opaque to visible light; the optical absorption edge at 300 K and atmospheric pressure is 867 nm. Assuming isotropic compressibility of the crystal, a pressure change at constant temperature causes a change in the bandgap described by the following coefficient:

$$\gamma_p = \frac{dE_g}{dp} = 11 \times 10^{-5} \text{ eV/MPa} \quad (11)$$

The energy bandgap in GaAs will broaden with pressure, contrary to the conventional wisdom that all materials should become "more metallic" under pressure. Broadening of the bandgap will result in a displacement of the optical absorption edge illustrated in Fig. 10 and following the equation

$$\frac{d\lambda}{dp} = -0.0667 \text{ nm/MPa} \quad (12)$$

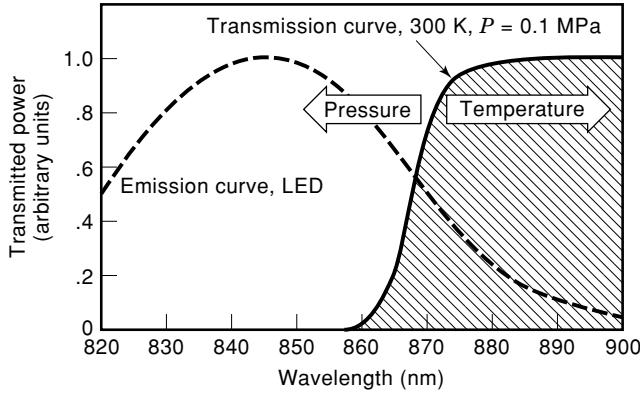


Figure 10. Transmission curve of GaAs and its dependence on temperature and pressure. The emission spectrum of the light source (dashed line) has to be carefully chosen to assure proper sensor operation.

The negative value of $d\lambda/dp$ indicates that the band edge will move toward shorter wavelengths with increasing pressure. Similarly, the GaAs energy bandgap will narrow with increasing temperature with the coefficient

$$\gamma_T = \frac{dE_G}{dT} = -4 \times 10^{-4} \text{ eV/K} \quad (13)$$

which corresponds to a positive shift of the optical absorption edge at a rate of 0.28 nm/K.

The emission spectrum of the LED light source used with the sensor should overlap on both sides of the optical absorption edge. The displacement of this edge induced by temperature or by pressure changes modifies the fraction of light that falls into the absorption band, meaning that the transmitted intensity will change. A successful temperature sensor system based on this principle is presently manufactured by Nortech Fibronic Inc. (Canada) and allows a 32-channel measurement of temperature in range from -200° to 250°C . The maximum length of the optical cable between a sensing probe terminated with a semiconductor tip and the processing unit is 1 km (22).

Another configuration has been used to develop a pressure sensor based on the same principle (21), with special attention paid to compensation of the temperature artifact and to amplification of the pressure effects. If we assume that pressure and temperature effects on a sensor are factorizable, the output signal at a detector will be proportional to the intensity of the source multiplied by the factors related to temperature and pressure:

$$I(p, T) = I_0 f(T) g(p) \quad (14)$$

Figure 11 shows a two-source compensation setup of a GaAs pressure sensor developed using this assumption and following the suggestion by Culshaw et al. (23). Two separate light sources supply light signals modulated at different frequencies f_1 and f_2 . By using two reference arms in addition to two sensor arms, the influence of different individual detector responsivities is eliminated. It can be shown that the ratio of two signals registered at two detectors a and b is described by

$$\frac{I_a}{I_b} = g(p) \quad (15)$$

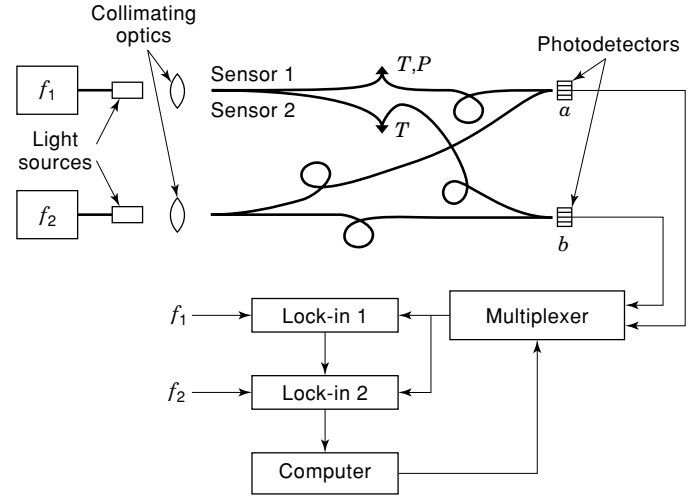


Figure 11. Compensation setup of a GaAs pressure sensor in two-source configuration.

and depends directly on the pressure factor, assuming the light entering the two fibers is of equal intensity, with other effects canceled out. This assumption of equal intensity is problematic due to the fluctuations in ambient temperature and due to laser mode hopping resulting in uncontrollable deviations of compensation; however, equal intensity can be achieved by separately adjusting the coupling of light into each of the fibers. The output signal of the thermally compensated pressure sensor (21) clearly illustrates the benefits of the presented configuration. Good compensation has been achieved, especially at the lower range of temperatures from 5° to 25°C : in the worst case, the temperature error was reduced by more than an order of magnitude to less than 0.15 MPa/K. Further improvement can be achieved by controlled doping of the semiconductor material and by digital processing of the temperature-calibrated sensor signal.

Polarimetric Sensors

Polarimetric sensors may function in a single-mode or in a few-mode regime of propagation. Depending on the selected regime of operation, the different propagation constants shown in Fig. 12 have to be considered in order to understand the

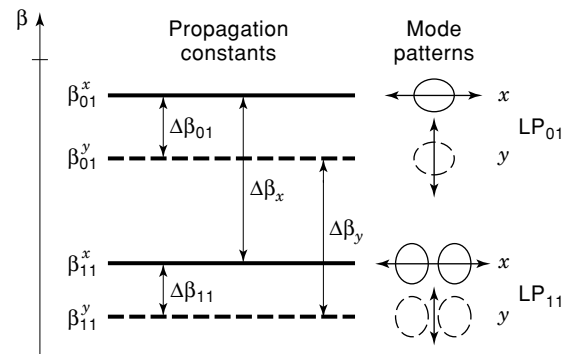


Figure 12. Propagation constants and mode patterns in HB bimodal fibers. Arrows indicate different possibilities of interference and show the corresponding phase shifts.

behavior of a specific sensor (24). Single-mode operation occurs when only one spatial mode LP_{01} or LP_{11} is excited at the fiber's input. No intermodal interference is observed in this case, as the second mode is absent.

Single-Mode Operation. If quasimonochromatic light linearly polarized at an angle ϕ with respect to the fiber's x axis is launched into the fiber and an analyzer turned to an angle α is placed at the output of the fiber, then the optical intensity detected will be

$$I = \frac{1}{2}(1 + \cos 2\alpha \cos 2\phi + |\gamma| \sin 2\alpha \sin 2\phi \cos \Phi_0) \quad (16)$$

where $\Phi_0 = \Delta\beta_{01} L$ is the phase. When external perturbations are introduced, they cause changes in the phase $\Phi_0 = \Delta\beta_{01} L$ of the fundamental LP_{01} mode (or correspondingly $\Phi_1 = \Delta\beta_{11} L$ for the LP_{11}). These will lead to a cosine variation of the observed intensity I measured after the analyzer, a variation that is in fact a polarization interference. The setup is then a *polarimetric sensor*. The interfering waves in this case are the LP_{01}^x and the LP_{01}^y polarization modes. With $|\gamma|$ we represent the correlation function between the polarization modes. This is a function of the product of the length L of the fiber, its polarization dispersion $\delta\tau$, and the spectral half-width $\delta\lambda$ of the source. The visibility V of the observed polarimetric response is

$$V = \frac{I_+ - I_-}{I_+ + I_-} = |\gamma| \frac{\sin 2\alpha \sin 2\phi}{1 + \cos 2\alpha \cos 2\phi} \quad (17)$$

An input polarizer (if the light is not linearly polarized) acts as a splitter, and the analyzer acts as a recombiner. If we define $k_1 = \sin^2 \phi$ and $k_2 = \sin^2 \alpha$ as the power coupling coefficients of the splitter and the recombiner, then the expression for visibility becomes

$$V = |\gamma| \frac{4\sqrt{k_1 k_2 (1 - k_1)(1 - k_2)}}{1 + (1 - 2k_1^2)(1 - 2k_2^2)} \quad (18)$$

This expression is analogous to the well-known formula for visibility in classical two-beam interferometry (8). So maximum visibility is obtained if $k_1 = k_2 = 0.5$, that is, for $\phi = \alpha = 45^\circ$. In other words, maximum visibility occurs when the LP_{01}^x and LP_{01}^y polarization modes are equally excited by launching $\pm 45^\circ$ linearly polarized light. In this case, the observed intensity and the visibility become

$$I = \frac{1}{2}(1 + |\gamma| \cos \Phi_0) \quad \text{with} \quad V = |\gamma| \quad (19)$$

Evidently, if a monochromatic source is used, $\delta\lambda = 0$ and then the visibility will be $|\gamma| = 1$.

Two-Mode Operation. For a two-mode regime of operation, a bimodal sensing fiber must be used and the interference between either x - or y -polarized LP_{01} and LP_{11} spatial modes is observed. By letting $\Phi_x = \Delta\beta_x L$ and $\Phi_y = \Delta\beta_y L$, the intensity observed at the output of the fiber excited with x - or y -polarized quasimonochromatic light can be obtained (25) as

$$\begin{aligned} I_x(x, y, z) &= \eta_0 f_0^2(x, y) + \eta_1 f_1^2(x, y) + 2\eta_{01} f_0(x, y) f_1(x, y) \cos \Phi_x \\ I_y(x, y, z) &= \eta_0 f_0^2(x, y) + \eta_1 f_1^2(x, y) + 2\eta_{01} f_0(x, y) f_1(x, y) \cos \Phi_y \end{aligned} \quad (20)$$

In the equations above, η_0 and η_1 are the relative optical powers carried by the spatial modes. Depending on the detection setup, different expressions for the visibility can be obtained. Assuming the Gaussian approximation (26), the distributions of the two modes are expressed as

$$\begin{aligned} f_0(x, y) &= \frac{1}{\sqrt{\pi \omega_x \omega_y}} F(x, y), \quad f_1(x, y) = \sqrt{\frac{2}{\pi \omega_x \omega_y}} \frac{x}{\omega_x} F(x, y) \\ F(x, y) &= \exp \left[-\frac{1}{2} \left(\frac{x^2}{\omega_x^2} + \frac{y^2}{\omega_y^2} \right) \right] \end{aligned} \quad (21)$$

The following three important subcases can readily be outlined: visibility at a point (pixel), visibility when half the pattern is detected, and visibility when an offset single-mode fiber is used. The visibility is then presented in the form $V(z, \eta_1) = |\Gamma_{x,y}| V_0$, where V_0 is a factor depending on the relative excitation coefficient of the second mode as well as on the detection scheme while $|\Gamma_{x,y}|$ are the correlation functions between the two interfering modes. In all of the cases, maximum visibility is obtained if the source is monochromatic and both modes are equally excited, that is, $\eta_0 = \eta_1 = 0.5$.

It is important to note that a fiber may be used simultaneously in the single- and the two-mode regime by operating it at two wavelengths at the same time. Thus e -core, D-shape, and bowtie fibers designed for single-mode operation at 850 nm (laser diode) will be bimodal when operated at a shorter wavelength of 633 nm or 670 nm, depending on the laser source. Alternatively, a polarimetric setup can be designed for 850 nm, and a polarization-sensitive intermodal interferometer for 633 nm. The setup illustrated in Fig. 13 could be utilized for this purpose. Note that the input and output polarization conditions

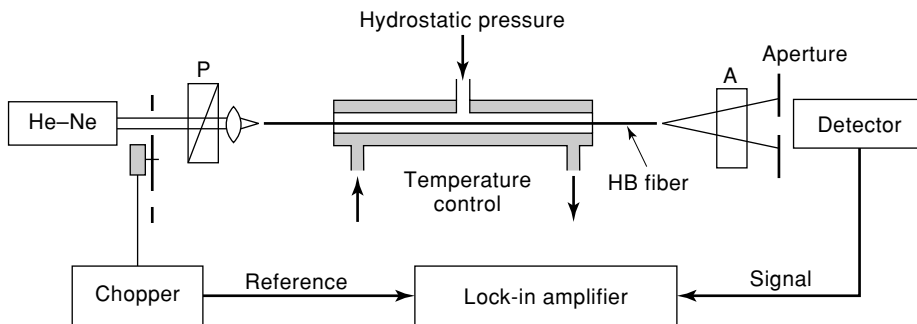


Figure 13. Basic configuration of a single-mode or bimodal polarimetric pressure sensor (A—analyzer, B—polarizer).

in the single- and two-mode regimes are different. If the fiber is operated at both wavelengths and all possible polarizations, then three independent cosine responses can be obtained.

Differential Phase Shifts. When an external perturbation such as pressure, strain, or temperature is applied to a fiber of length L , it will change the phases $\Phi_i = \Delta\beta_i L$ ($i = 0, x, y$) by the amount

$$\delta\Phi_i = \delta(\Delta\beta_i)L + \Delta\beta_i\delta L \quad (22)$$

If the external perturbation is hydrostatic pressure p , then an increase by δp will cause both $\Delta\beta_i$ and L to change by $\partial(\Delta\beta_i)/\partial p$ and $\partial L/\partial p$. So from Eq. (22) we can obtain

$$\frac{\delta\Phi_i}{\delta p} = \frac{\partial(\Delta\beta_i)}{\partial p}L + \Delta\beta_i\frac{\partial L}{\partial p} = \frac{2\pi}{T_{i,p}} = \Lambda_{i,p} \quad (23)$$

The above equation can also be rewritten as

$$\delta\Phi_i = \left(\frac{\partial(\Delta\beta_i)}{\partial p}L + \Delta\beta_i\frac{\partial L}{\partial p} \right) \delta p = \frac{2\pi}{T_{i,p}}\delta p = \Lambda_{i,p}\delta p \quad (24)$$

Note that $T_{i,p}$ ($i = 0, x, y$) has the dimensions of pressure, while $\Lambda_{i,p}$ has the inverse dimensions. These are experimentally measurable parameters and determine the sensitivity of the sensor to a given external perturbation [equations similar to Eq. (24) describe the phase changes induced by temperature or strain].

Polarimetric Pressure Sensor. The design of a practical high-hydrostatic-pressure polarimetric sensor with a predesigned sensitivity as expressed by the parameter $T_{i,p}$, is based on an important relation expressed simply as (24):

$$T_{i,p}L = C_i \cdot \lambda \quad (25)$$

where C_i is a constant. Thus, a longer fiber will exhibit greater sensitivity and vice versa. Note that if the source is not monochromatic, a longer fiber will produce a lower value of the correlation function $|\gamma|$, which in turn will null the cosine response of the sensor. Therefore, proper choice of the laser source is essential.

For pressure measurement (considering temperature to be a disturbing parameter), temperature- and pressure-induced

phase shifts $\Delta\phi_S(p, t)$ will then be transformed into intensity changes $I_S(p, t) - I_0$ in accordance with the equation

$$I_S(p, t) = I_0[1 - \cos \Delta\phi_S(p, t)] \quad (26)$$

Figure 14 shows a topology of a polarimetric pressure sensor (PPS) in both reflection and transmission configurations, employing the principle of temperature self-compensation (27). The sensing (L_2) and compensating (L_1) parts of the sensor are assumed to be equal. The advantage of the reflection configuration is that only one fiber leadthrough is required to connect the sensor to the laser source and to the detection electronics. The advantage of the transmission version lies in the much higher level of optical signal, allowing for longer transmission distances and/or multiplexing of several sensing devices. The polarization axes of the sensing and the compensating fiber elements are rotated by 90° , while the input and output fibers are rotated at 45° relative to the sensing and compensating parts, respectively. Ideally, if equal sensing and compensating elements remain at the same temperature, their temperature-induced phase retardations will cancel out, and if they are placed under different pressures, the pressure-modulated output signal will be immune to temperature changes. In reality, however, if such an arrangement is to satisfy the requirements for an accurate pressure-measuring instrument, several important first- and second-order coefficients have to be considered to allow for simultaneous pressure and temperature effects on fiber birefringence. In a polarimetric cross-spliced sensor, the resulting unwanted sensitivity to temperature will still be present and can be described by differentiation of the phase retardation with respect to temperature (28):

$$\begin{aligned} \rho_t &= \frac{\partial \Delta\phi_S}{\partial t} = \Delta L K_t + L K_{tp} \Delta p \\ K_t &= \frac{1}{L} \frac{\partial \phi}{\partial t}, \quad K_{tp} = \frac{1}{L} \frac{\partial^2 \phi}{\partial t \partial p} \end{aligned} \quad (27)$$

where $\Delta L = L_1 - L_2$, K_t is the first-order phase sensitivity to temperature, and K_{tp} is the temperature–pressure cross-sensitivity coefficient. This simple but informative equation provides important insight into designing a sensor with minimum temperature error. The first term in it can be minimized by choosing a small ΔL . The second term depends on fiber properties but will always be nonzero, and can only be minimized

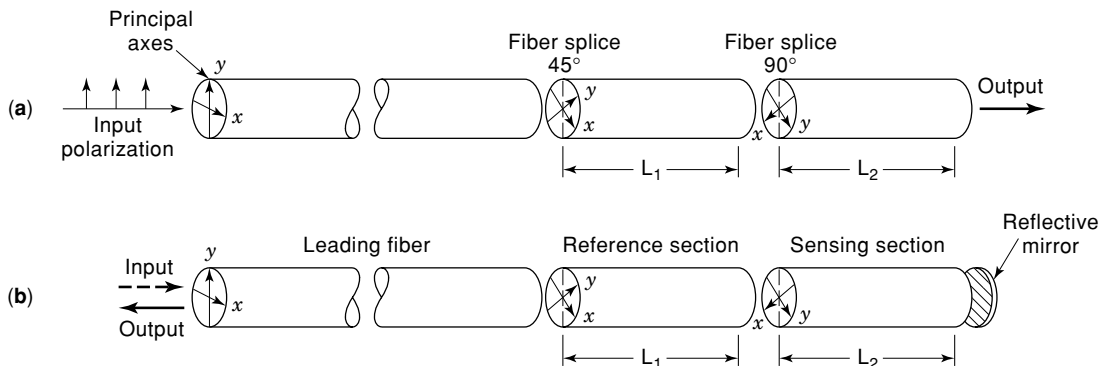


Figure 14. Temperature-compensated polarimetric fiber-optic sensor in (a) transmission and in (b) reflection configuration.

by carefully adjusting the fiber's technological and construction parameters.

Polarimetric Current Sensor. The need for fiber-optic sensing technology for high magnetic field and large current monitoring is now well documented. Conventional magnetic field and current sensor systems suffer from high susceptibility to electromagnetic interference, may lack the necessary bandwidth, are difficult to miniaturize, and cannot accommodate large numbers of measuring points at remote locations. Such monitoring can be especially valuable for protection, control, and fault detection in power plants, high-power transmission lines, substations, and distribution lines, where the high intrinsic electrical insulation of optical fibers is a significant advantage. Several techniques for fiber-optic magnetic field sensing have already been proposed (29), but only two approaches appear to be viable. The first is based on detection of a magnetic field by magnetostrictive effects, involving measuring the longitudinal strain produced in the optical fiber to which a magnetostrictive material has been bonded. The performance of such sensors is limited by the coupling efficiency of the magnetostrictive material and the optical fiber. Although various bonding and coating techniques have been explored, all usually lead to substantial hysteresis, temperature drift, and changes of fiber birefringence. The second approach is based on the well-known Faraday effect (30), consisting of a nonreciprocal circular birefringence induced in the propagation medium by a magnetic field and characterized by the Verdet constant V . The most convenient detection approach in this case is polarimetric sensing.

The Faraday effect may occur directly in standard or specifically doped optical fibers, but as V in silica fiber is very small, this type of sensor needs to be very long and as such will be prone to a variety of instabilities. For N turns of fiber around a conductor with a current I the Faraday rotation is given by

$$\Phi_F = \int_{L_F} V(\lambda, T) \mathbf{H} \cdot d\mathbf{L}_F \quad (28)$$

where L_F is the propagation path and $\mathbf{H}$ is magnetic field. The material-dependent Verdet constant $V(\lambda, T)$ is dispersive and often varies strongly as a function of temperature. To assure successful operation of a sensor based on the fiber-sensitive element, it is extremely important to avoid intrinsic birefringence induced by core ellipticity or stress in the core-cladding area, and extrinsic birefringence induced by packaging and mounting. This parasitic effect can be alleviated to some extent by annealing the fibers at an elevated temperature (31). Figure 15 shows a simple polarimetric sensor based on this principle. Coils with 5 mm to 10 cm diameter and up to 300 turns have been produced, and a temperature sensitivity of $1.7 \times 10^{-4} \text{ K}^{-1}$ has been demonstrated over a -20° to 120°C range. This technology has been transferred from NIST to the 3M Corporation for commercialization.

An approach based on extrinsic polarimetric sensors seems, however, to be more promising. The Faraday effect in ferrimagnetic garnets such as single-crystal yttrium iron garnets (YIG), Bi-doped YIG (BIG), and gallium-substituted YIG (Ga:YIG) (32,33) has been explored, since their Verdet constant can be greater than that of silica fibers by a factor of several thousand. Although the large V gives such sensors an impressive advantage over the all-fiber configurations, inherent problems asso-

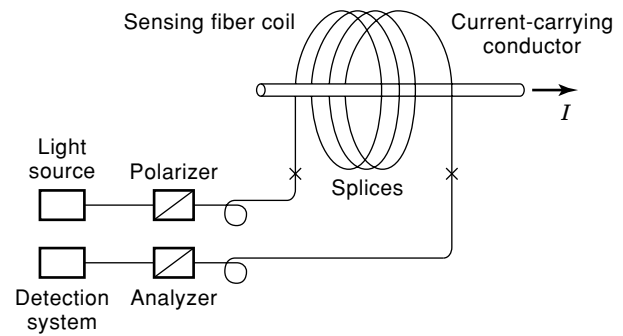


Figure 15. Polarimetric fiber-optic current sensor.

ciated with the ferrimagnetic domain structure of these materials are large hysteresis, temperature drift of about $0.2\%/K$, nonlinearities, saturation at relatively low fields of about 0.1 T , and vanishing of response at frequencies of about 500 MHz . Recently, however, Inoue and Yamasawa (34) reported obtaining a Bi-doped garnet by liquid-phase epitaxy that shows a flat temperature characteristic from -10° to 80°C . Also, since the recent discovery of the large Faraday effect in $\text{Cd}_{1-x}\text{Mn}_x\text{Te}$ (CMT) (35), interest in this group of materials has been steadily growing with a view to their application not only in optoelectronic devices such as modulators and magneto-optic insulators, but also for polarimetric current sensors. These materials have improved thermal stability and can work at much higher fields (up to 25 T) and at higher frequencies (up to 5 GHz).

Interferometric Sensors

Fiber-optic interferometric sensors are usually designed following classical configurations of optical interferometers (8). In these devices, a range of physical measurands can induce modulation of phase in a sensing signal light beam, while the reference light beam remains unchanged. This phase change then has to be electronically processed, often by complex and sophisticated systems, to produce a useful intensity-type output signal from the interferometer proportional to the measurand. Although this technique offers very high sensitivity, it is extremely difficult to use outside the laboratory due to the unavoidable interference caused by environmental perturbations. One notable exception is low-coherence (or white-light) interferometry, which is presented later in more detail.

Classical Fiber Interferometers. Figure 16 shows the three best-known configurations of two-beam fiber-optic interferometers: (a) the reflective all-fiber Michelson interferometer, (b) the all-fiber Mach-Zehnder interferometer, and (c) a remote Fabry-Perot interferometer. Another configuration—the reciprocal Sagnac interferometer—is discussed separately. The demodulation systems for interferometric sensors are usually based on *homodyne* detection, where the sensing and reference beams have the same frequency, or on *heterodyne* detection, where the reference beam has a fixed frequency that is different from that of the signal beam.

The all-fiber Michelson interferometer is based on a bidirectional single-mode fiber coupler that divides the input light beam from the laser source into two components—one propagated by the sensing arm, the other by the reference arm of the interferometer—and then combines the two reflected beams so

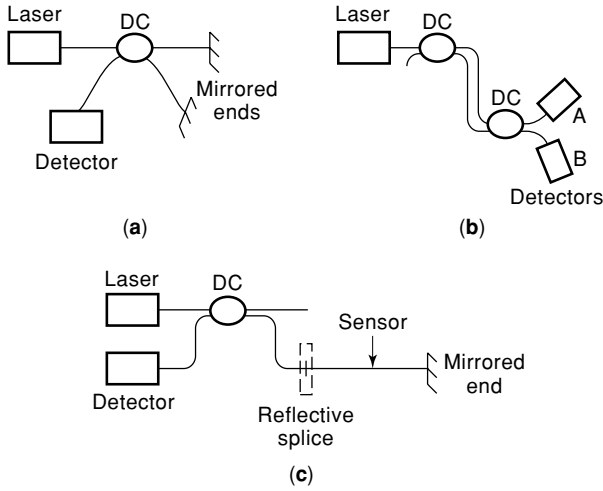


Figure 16. Three configurations of two-beam fiber interferometers: (a) Michelson, (b) Mach-Zehnder, (c) Fabry-Perot.

that their interference can be registered by the photodetector. Assuming for simplicity that the polarization effects can be ignored, the electric fields propagating in the signal arm and in the reference arm can be treated as scalars and described as

$$\begin{aligned} E_S &= E_{S0} \cos \left[\omega t + 2 \left(\frac{2\pi L_S}{\lambda} \right) \right] \quad \text{and} \\ E_R &= E_{R0} \cos \left[\omega t + 2 \left(\frac{2\pi L_R}{\lambda} \right) \right] \end{aligned} \quad (29)$$

In this equation, E_{S0} and E_{R0} denote the amplitudes, λ is the wavelength of the light, and ω is its angular frequency. The phase difference will then be proportional to the path difference ΔL and can be expressed as follows:

$$\Delta\phi = \frac{4\pi}{\lambda} (L_S - L_R) \quad (30)$$

It may finally be shown that the irradiance at the detector is proportional to the absolute value of the squared sum of the superimposed electrical fields. The resulting formula is given by

$$I_d = I_0(1 - V \cos \Delta\phi) \quad (31)$$

(and corresponds to that of the polarimetric and dual mode sensors (Eq. (19)), where V is the visibility defined as in Eq. (17) and depending on both the reflectivity of the fiber ends and the coherence properties of the light source, but independent of the splitting ratio of the coupler. The maximum visibility occurs at the so-called quadrature condition when $\Delta\phi = \pi/2$ or when the path difference is equal to $\lambda/4$.

The Mach-Zehnder interferometer is based on two bidirectional fiber couplers, the first to divide the light beam into two components and the second to recombine the two beams exiting from the sensing arm and from the reference arm of the system. The sensitivity of this interferometer is only half that of the Michelson interferometer, as light propagates in each arm only once, and the phase difference is consequently

described by

$$\Delta\phi = \frac{2\pi}{\lambda} (L_S - L_R) \quad (32)$$

The Mach-Zehnder configuration has, however, two significant advantages that more than compensate for the lower sensitivity. Two antiphase output signals from two photodetectors,

$$\begin{aligned} I_A &= I_0(1 - V \cos \Delta\phi) \\ I_B &= I_0(1 + V \cos \Delta\phi) \end{aligned} \quad (33)$$

can conveniently be used to provide a feedback loop for assuring operation at maximum sensitivity (quadrature condition when $\Delta\phi = \pi/2$). This configuration is also characterized by much lower backreflection into the laser diode, which assures the higher wavelength and power stability of the system.

A useful and more practical example of the multiple-beam technique is the fiber Fabry-Perot interferometer (FFPI). A resonant cavity of this device may be contained within the fiber, with semireflective splices (36), cleaved or mirrored (37) end faces, or Bragg gratings (38) serving as reflective surfaces. This cavity may also be external to the fiber, taking the form of an air gap between two cleaved fiber end faces, or between a fiber end face and a thin moving or deformable diaphragm (39). Following classical textbooks (40), the transfer function of an FFPI for the transmitted signal can be expressed by

$$I = \frac{I_0}{1 + F \sin^2(\phi/2)} \quad (34)$$

where F is a parameter describing the phase resolution and known as the *fineness* of the interferometer, and ϕ is the phase retardance after the light has passed through the cavity twice. When attenuation is disregarded, F may be described in terms of the mirror reflectance R :

$$F = \frac{4R}{(1 - R)^2} \quad (35)$$

In the case of the frequently used low-finesse interferometer, the reflective surfaces may simply be the normally cleaved uncoated fiber ends, for which $R = 0.04$. The FFPI is then operated in a reflective configuration with visibility approaching 1 as the reflectivity is decreased. For $R \ll 1$ the FFPI signal may be approximated by the signal of the two-beam interferometer (Eq. (31)): $I_r = I_0 2R(1 - \cos \phi)$ for the reflective case and $I_t = I_0[1 - 2R(1 - \cos \phi)]$ for the transmitted signal.

Recently several quite interesting low-finesse Fabry-Perot-based fiber-optic sensors have been reported for a wide range of different measurands. Beard and Mills (41) developed an extrinsic optical fiber ultrasound sensor based on a thin transparent polymer film serving as the resonant cavity and mounted at an end face of a multimode fiber. The device can then be used for the detection and measurement of ultrasound with a sensitivity of about 60 mV/MPa. A phase-shifted extrinsic FFPI for dynamic strain measurements has been developed (42) using two laterally displaced single-mode fibers inserted into a hollow silica tube. Through an air gap these fibers face a multimode fiber acting as a reflector and forming a Fabry-Perot cavity. A high-sensitivity low-pressure sensor up to 13 kPa (100 Torr) has been developed (37) using a diaphragm bonded

to the strained fiber containing the interferometer with two in-fiber mirrors, with the motion of the diaphragm producing a change in the length of a Fabry-Perot cavity. Even a fiber-optic sensor for electric field measurement has been developed around the concept of a low-coherence FFPI using a miniature cantilever beam, the deflection of which depends on the electrical charge induced by the external field to be measured (39).

For most of these sensors, however, active phase control and complex polarization control systems are usually required to keep them at their optimum operating points, at the expense of significant cost, high system complexity, and low potential for wide practical implementations. Recently Fürstenau and Schmidt (43) made a significant breakthrough in this area by proposing a two-wavelength passive quadrature demodulation system for FFPI vibration sensors, based on a low-coherence source in combination with adjustable interference filters and an electronic switching unit. This novel and cost-effective sensing system is now undergoing tests in an application for airport ground traffic monitoring.

White-Light Interferometry. Although low-coherence or white-light interferometry (WLI) has been known in fiber-optic sensing since the late 1980s (44), a significant research effort in this area became noticeable only around 1990. By then, the deficiencies of high-coherence classical interferometry, especially its inability to deal with absolute measurements and the enormous costs of the decoding equipment involved, were becoming more apparent. Since that time, significant advances in WLI sensor design, signal processing, and sensor multiplexing have been reported, and all prove that this technology has indeed matured to the point of becoming one of the most promising in fiber-optic sensing, thanks to its practicality and cost-effectiveness. Due to its youth, this technology has not yet received much book coverage, although several excellent reviews such as Ref. 45 have been devoted to this topic.

The principle of a WLI sensor system is illustrated in Fig. 17 in a reflective configuration. A broadband source, typically an SLD or a LD operated below threshold, launches light into a bidirectional coupler and into the sensing interferometer located in the measurand field. The modulated output signal of this interferometer is then coupled back through the coupler into the local receiving interferometer. When the optical path differences of the two interferometers are matched one to another within the coherence length of the white-light source, interference fringes with their amplitude modulated by the contrast function will be observed at the output of the receiving

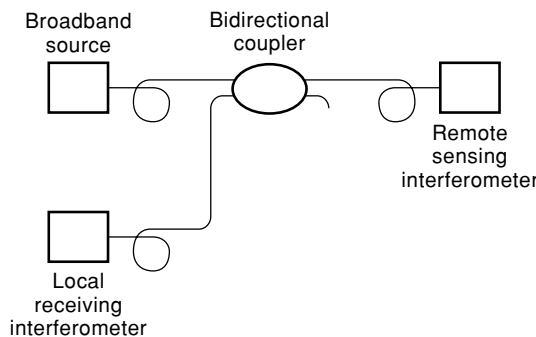


Figure 17. Basic configuration of a white-light interferometric fiber-optic sensor.

interferometer assuming a Gaussian spectral distribution of the source. The interferogram obtained at the photodetector can then be described (45) by

$$I = I_0 \left\{ 1 + \frac{1}{2} \exp \left[- \left(\frac{2\Delta X}{L_c} \right)^2 \right] \cos(k\Delta X) \right\} \quad (36)$$

where L_c is the coherence length of the source, $\Delta X = X_1 - X_2$ is the mismatch between the optical path differences of the sensing and the receiving interferometers, k is the wave number, and I_0 is the total optical intensity arriving at the photodetector. Note that the necessary condition for the system to work is $\Delta X < L_c$. It is clearly evident that any change of the external measurand acting upon the sensing interferometer will result in the interference pattern being shifted. This shift can then be easily registered or compensated for at the receiving interferometer, giving an absolute measurement of the external parameter in question. Many application-specific sensor systems have been developed to measure temperature (46), strain (47), pressure (48), and force (49) using an often complex variety of signal-processing techniques (45). In recent years, processing based on electronic scanning for low-coherence sensing systems has attracted much attention, since it does not require any piezoelectric or mechanical scanning devices and makes the system more compact, more stable, and less expensive.

An example of an electronically scanned WLI strain sensor based on HB fibers and with a Wollaston prism serving as a receiving interferometer (50) is shown in Fig. 18. The light source is an SLD with a central wavelength $\lambda_0 = 820$ nm and a coherence length of about $15 \mu\text{m}$. The input light is linearly polarized by the polarizer (or by a polarizing fiber) aligned at 45° to the polarization axes of the sensor in order to excite both polarization modes equally. The sensor itself is composed of two exactly equal lengths of the HB fiber spliced with their corresponding polarization axes rotated at 90° and giving a total phase shift at the output of the system close to zero. Light from the output of the sensor is collimated by lens L and passes through the Wollaston prism WP, which is made of crystalline quartz. The polarization axes at the output of the sensor are aligned to be parallel to the polarization axes of the WP. Therefore, the x - and y -polarized modes of the sensor are spatially separated by the Wollaston prism with the bisection angle α equal to 2° . The two modes interfere after passing through the analyzer A, which has its transmission azimuth aligned at 45° to polarization axes of WP. The cylindrical lens CL focuses the output beam on the CCD camera (or a CCD array) with a horizontal resolution of 1024 pixels, enabling registration of the resulting interference pattern with an adequate signal-to-noise ratio.

At an arbitrary point along the horizontal direction of the CCD camera, the two polarization modes intersect at an angle α , and the resulting intensity of the interference pattern may be written as

$$I(y) = I_0(y) [1 + \gamma(\phi_s + \phi_{WP}) \cos(\phi_s + \phi_{WP})] \quad (37)$$

where $I_0(y)$ is the bias intensity profile of the output beam; ϕ_s is the phase imbalance between the polarization modes at the output of the sensor; $\phi_{WP}(y)$ is the phase difference introduced

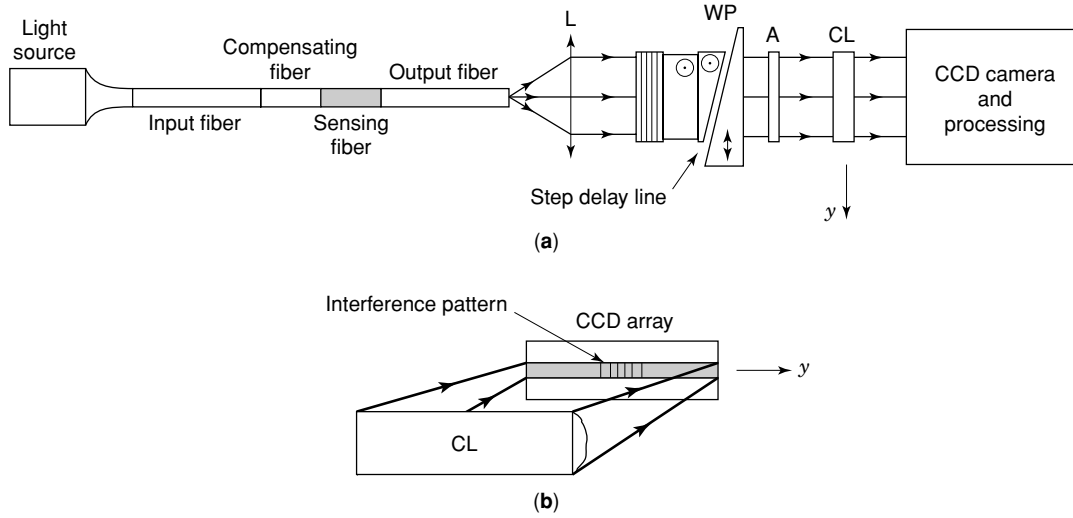


Figure 18. Electronically scanned white-light interferometric strain sensor in temperature-compensated configuration: (a) general setup, (b) detail of the output pattern exiting the receiving interferometer.

by the Wollaston prism, which can be expressed as

$$\phi_{WP} = \frac{2\pi}{\lambda_0} \alpha (y - y_0) \quad (38)$$

and γ is the coherence function of the light source. The longer wedge of the Wollaston prism may be shifted transversely across the beam, allowing for adjustment of the zero position of the interference pattern with respect to the scanning area of the CCD camera. The parameter y_0 in Eq. (38) indicates the position of the center of the longer wedge with respect to the edge of the scanning area. Assuming that the sensor is initially perfectly balanced, the phase shift occurring at the output due to the elongation of the sensing part of the fiber will be

$$\phi_s = L_0 \left(\Delta\beta + \frac{\partial \Delta\beta}{\partial \epsilon} \right) \epsilon \quad (39)$$

where L_0 is the initial length of the sensing part of the fiber and ϵ is the strain applied. An increase in the output phase shift ϕ_s results in a transverse shift Δy of the white-light interference pattern, proportional to the absolute value of ϵ :

$$\Delta y = \frac{\lambda_0 L_0}{2\pi \alpha} \left(\Delta\beta + \frac{\partial \Delta\beta}{\partial \epsilon} \right) \epsilon \quad (40)$$

Consequently, it is clear that processing of the signal digitized by the CCD camera needs to establish the location of the center of the white-light interference pattern determined by the symmetry axis of the coherence function $\gamma(\cdot)$. This can be achieved in a three-step procedure as discussed in detail in Ref. 45. First, the bias intensity function is determined by simple numerical filtering-out of higher harmonics from the output intensity. Then, in the second step, the coherence function is determined, and finally, in the third step, the upper part of the coherence function is approximated by a parabolic function, the symmetry axis of which determines the central position of

the white-light interference pattern. This position may thus depend on the action of the selected measurand. For instance, in the case of strain measurement illustrated by Fig. 19, an operating range of 4000×10^{-6} was achieved with an absolute accuracy of 0.5% of full scale (FS) (including temperature drift of 40°C) and a resolution of about 10^{-3} with no hysteresis effect. Similar performance has been achieved in pressure measurement up to 100 MPa with an almost identical configuration, the only difference being that the sensing element was subjected not to elongation but to hydrostatic pressure (51).

Bragg Grating Sensors

Periodic changes of the refractive index profile in optical waveguides are equivalent to gratings and constitute a source of scattering for the guided light. When the spatial period of such a grating is matched with the light wavelength, the scattered light can be guided as a backreflected wave. This phenomenon is applied for instance in distributed feedback resonators in semiconductor lasers, in distributed fiber Bragg reflector structures, and in sensing structures (52).

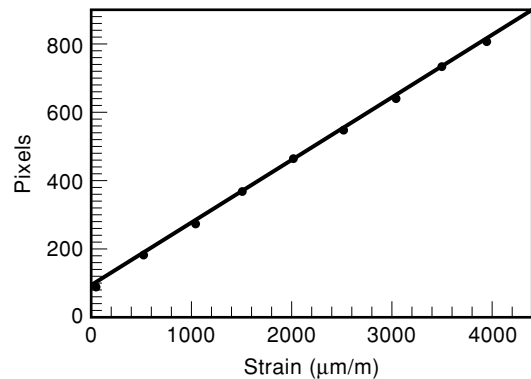


Figure 19. Example of output signal of the WLI strain sensor: position of the center of interference pattern versus strain.

Assuming that the refractive index profile n_p in an optical fiber changes with the spatial period Λ as

$$n_p = n_0 + \Delta n \cos\left(\frac{2\pi}{\Lambda} z\right) \quad (41)$$

the complex amplitudes of the forward mode A and the backreflected mode B fulfill the coupled equations (53)

$$\begin{aligned} \frac{dA}{dz} &= -i\kappa B \exp(2i\delta z) \\ \frac{dB}{dz} &= i\kappa A \exp(-2i\delta z) \end{aligned} \quad (42)$$

A normalized frequency δ represents the mismatch between the wavelength λ and the spatial period of the grating:

$$\delta = \frac{2\pi n}{\lambda} - \frac{\pi}{\Lambda} \quad (43)$$

where n is the effective refractive index of the guided wave. The coupling coefficient κ depends on the effective depth of the refractive index modulation:

$$\kappa = \frac{\pi}{\lambda n} \int 2n_0 \Delta n e_1 e_2 dS \quad (44)$$

where $e_{1,2}$ are the normalized electric field envelopes of the guided modes, and the integration is over the fiber cross section. Solution of Eq. (42) for a grating with length L gives the following expression for the reflectivity:

$$R = \frac{|B(z=0)|^2}{|A(z=0)|^2} = \frac{\kappa^2}{\gamma^2 \coth^2 \gamma L + \delta^2} \quad (45)$$

where $\gamma^2 = \kappa^2 - \delta^2$. The reflectivity depends strongly on the length of the grating, and consequently the backreflection can be high even when the index of refraction has a small modulation amplitude Δn . The maximum reflection, as illustrated in Fig. 20, is obtained at a specific wavelength λ_B that is twice the period of the grating within the fiber:

$$\lambda_B = 2n\Lambda \quad (46)$$

Bragg gratings also make it possible to couple two modes copropagating inside a multimode fiber core, or to couple core

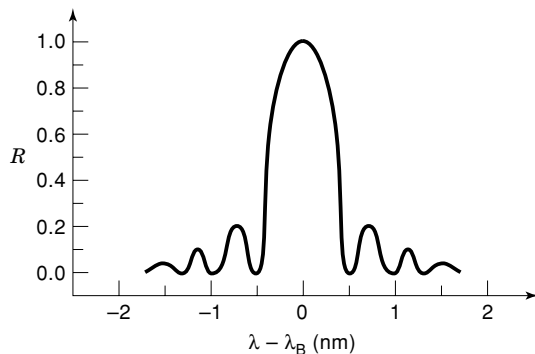


Figure 20. Theoretical dependence of the reflectivity in Bragg grating structures on wavelength ($\kappa = 1 \text{ mm}^{-1}$, $L = 2 \text{ mm}$, $\lambda_B = 1330 \text{ nm}$).

and cladding modes. This phenomenon occurs when Eqs. (43) and (46) are fulfilled with $2n$ replaced by the difference between the effective refractive indices of the two coupled modes, $n_1 - n_2$. In this case, the resonant Bragg wavelength λ_B corresponds to longer grating periods Λ . Typically the spatial period Λ in such long periodic gratings is hundreds of micrometers long, while in reflecting fiber Bragg gratings a period of only hundreds of nanometers is typical.

Fiber Bragg gratings are manufactured by writing a permanent periodic modulation of a fiber's index of refraction within the fiber using a photoinduced effect. This effect is particularly useful in germanium-doped silica fibers. The resulting pattern of the refractive index is induced by the interference fringes generated by the interfering ultraviolet or near-ultraviolet radiation. The effect may be obtained using either a standing wave created by the light guided by the fiber, or the diffraction and interference of light externally illuminating the fiber.

The wavelength λ_B of the backreflected light depends strongly on temperature T , a fact that constitutes a major drawback of this method:

$$\frac{\partial \lambda_B}{\partial T} = 2 \left(\Lambda \frac{\partial n}{\partial T} + n \frac{\partial \Lambda}{\partial T} \right) \quad (47)$$

Obviously this wavelength also depends (in a manner similar to that expressed in the above equation) on a variety of potential external measurands such as strain or pressure. Several complex techniques have been proposed (54), with very limited success, to solve this fundamental problem of discrimination between temperature effects and those induced by the measurand in question. Proposed solutions involve (1) using two gratings, one isolated from strain and playing the role of a reference grating; (2) using two gratings for different wavelengths; and (3) using two gratings for the same wavelengths, but with different strain sensitivities (e.g., formed in different fibers). Measured sensitivities of 10 pm/K, 1 pm/(μm/m), and 5 pm/MPa have been reported (55) to temperature, strain, and hydrostatic pressure, respectively. The difficulty of the measurement is increased by the requirement for high-precision wavelength measurement: λ_B may change by as little as 1 nm within the full scale of the intended measurement. The simplest FOS systems with fiber Bragg gratings use a broadband source to encompass the Bragg wavelength and a wavelength-selective detector to measure the reflected wave (Fig. 21). Using several different gratings with different λ_B , we can construct a quasidistributed sensor system with wavelength-division multiplexing techniques.

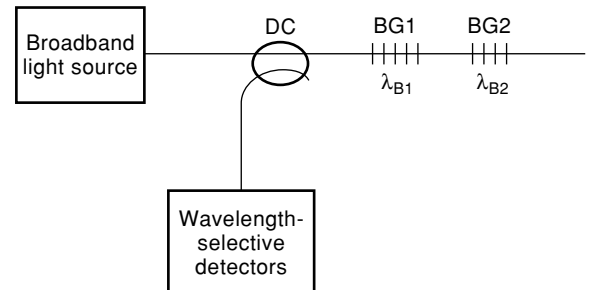


Figure 21. Basic configuration of a fiber Bragg grating sensor, with two Bragg gratings BG1 and BG2.

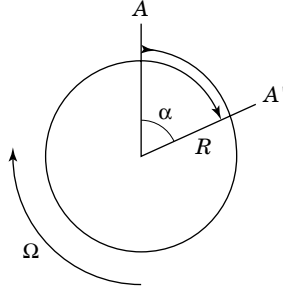


Figure 22. Sagnac effect in optical fiber ring.

Fiber-Optic Gyro

Fiber-optic gyros (FOGs) are most mature of all fiber-optic sensors and are now widely considered in altitude and heading reference systems, and for car navigation systems with reduced requirements. A solid-state configuration gives them a significant advantage over conventional mechanical gyros based on the principle of inertia of spinning masses. FOGs have found practical application in aircraft, tactical missiles, trains, autonomously guided vehicles, and marine gyrocompasses (56).

The principle of operation of FOGs is based on the rotationally induced phase shift of the light propagating in the fiber coil (Sagnac effect) (8). When the fiber ring in Fig. 22 rotates in the inertial system, then light propagated in the same direction will traverse a longer path to return back to the initial point A (moved to A') at the ring. The additional path is equal to αR , where R is the ring radius and the angle $\alpha = \Omega\tau$ depends on the rotation rate Ω and the time τ it takes the light to traverse the ring. For a ring length L , the time τ is equal to Ln/c , where c/n is the light speed in a medium with the refractive index n . Consequently, the phase shift ϕ for a light wave with the angular frequency ω that traverses the rotating ring can be expressed as

$$\phi = \left(L + \Omega RL \frac{n}{c}\right) \omega \frac{n}{c} \quad (48)$$

Similarly, for a counterpropagating wave we obtain

$$\phi_c = \left(L - \Omega RL \frac{n}{c}\right) \omega \frac{n}{c} \quad (49)$$

A FOG operates by measuring the phase difference $\Delta\phi = \phi - \phi_c$ between two counterpropagating waves, the difference being proportional to the rotation rate Ω of the fiber coil:

$$\Delta\phi = \phi - \phi_c = 2\Omega RL\omega \left(\frac{n}{c}\right)^2 \quad (50)$$

Three principal passive configurations for a FOG are an interferometer (I-FOG) with an open loop, an I-FOG with a closed loop, and a resonator (R-FOG). In the I-FOG with the open loop (Fig. 23), light from the laser source is divided by a directional coupler into two beams with equal intensity counterpropagating in a fiber coil composed of many turns. Then the outgoing light is mixed again in the directional coupler and the interference of the two waves is measured by the detector. The interference, similarly to the Eq. (31) and (32) for two-beam

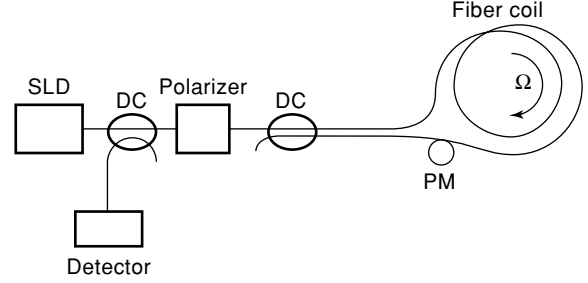


Figure 23. Basic configuration of an I-FOG with an open polarization-maintaining fiber loop. PM is a phase modulator.

interferometers, can be expressed as follows:

$$I = I_0(1 + V \cos \Delta\phi) \quad (51)$$

where V is the interferometric coefficient and I_0 is the input intensity modified by any losses in the optical system. To avoid parasitic dependence of interference on the polarization state, the configuration requires use of a polarizer, and the fiber in the coil has to be polarization-maintaining. The detection system can be improved by incorporating a phase modulator (PM) with a fixed frequency f_m at the end of the fiber loop. This induces an additional phase shift

$$\Delta\phi_m = \phi_m \cos(2\pi f_m(t + \tau)) - \phi_m \cos(2\pi f_m t) \approx F \cos(2\pi f_m t) \quad (52)$$

where ϕ_m is the modulation depth and $F = 2\phi_m \sin(\pi f_m \tau)$. In this case, the output signal at the detector can be expressed as follows:

$$I(t) = I_0\{1 + V \cos[\Delta\phi - F \sin(2\pi f_m t)]\} \quad (53)$$

Then the detected signal is modulated with an amplitude that depends on the magnitude of the rotation rate Ω , with the phase dependent on the direction of rotation.

The I-FOG with the open loop is the simplest configuration and the most popular to date. For applications that require higher performance and dynamic range, the I-FOG with the closed loop is more promising. In this configuration, a tunable frequency shifter is added at the fiber coil end. The frequency shifter changes the angular frequency of the first wave when it enters the fiber ring, and of the second wave when it exits the fiber ring. Therefore the phase shift of the first wave is modified in the optical fiber coil as follows:

$$\phi = \left(L + \Omega RL \frac{n}{c}\right) (\omega - \omega_s) \frac{n}{c} \quad (54)$$

where ω_s is the frequency shift induced by the frequency shifter, and the second counterpropagating wave has a phase shift given by Eq. (49). Then the phase difference $\Delta\phi$ is equal to

$$\Delta\phi = 2\Omega RL\omega \left(\frac{n}{c}\right)^2 - L\omega_s \frac{n}{c} \quad (55)$$

Changing the frequency ω_s , one can null the phase-shift difference, and the obtained frequency ω_0 for $\Delta\phi = 0$ gives the

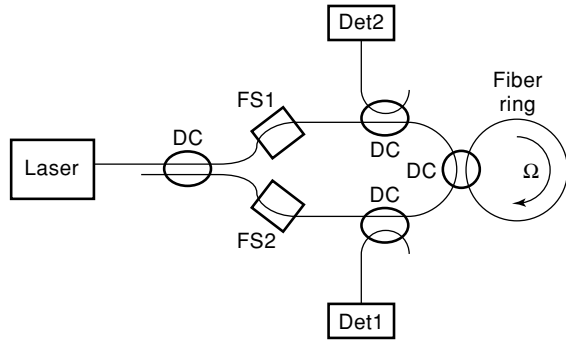


Figure 24. Basic configuration of the resonator-based gyro (R-FOG).

rotation rate:

$$\omega_0 = 2\Omega R\omega \left(\frac{n}{c} \right) \quad (56)$$

The frequency shifter is typically made as an electro-optical modulator in integrated optics technology, mounted together with a Y junction, which plays the role of a beamsplitter (similar to a directional coupler), and with a polarizer. The closed-loop configuration gives a very good bias and stable scale factor (i.e., proportionality between the rotation rate and the output). In both types of I-FOG, low-coherence light sources (such as SLDs or erbium-doped fiber lasers) are used to minimize noise due to reflected and backscattered light. However, the use of low-coherence sources causes fluctuations due to the potential wavelength drift. The sensitivity of I-FOGs depends on the length of the polarization-maintaining fiber in the coil. Increasing the length of the fiber increases costs (polarization-maintaining fibers are expensive in comparison with other elements in FOGs) and also increases noise due to the variable temperature distribution. Some of the above problems can be solved by resonator FOGs.

In R-FOGs (Fig. 24), the phase shifts of the two counter-propagating waves are detected independently. The detector Det1 measures the interference between the light traversing the fiber ring and that incoming from the light source in the same direction. The frequency shifter FS1 changes the frequency to maximize the intensity measured by Det1. Similar processing occurs with the counterpropagating wave, and the difference between the frequencies shifted by FS1 and FS2 is directly decoded as a rotation rate. The R-FOG configuration requires highly coherent and stable sources of light, but the polarization-maintaining fiber loop is much shorter than in I-FOGs. R-FOG systems are still, however, in the research stage. Many other concepts, such as ring laser systems, Brillouin ring lasers, or the use of depolarized light in a fiber coil, could be applied in advanced FOG development. It should be noted that the fiber-optic ring resonators and coils could also be used as sensors for measuring other measurands than the rotation.

Distributed Sensors

Due to the unidimensional structure of an optical fiber, fiber-optic sensing technology offers the unique possibility to simultaneously measure not only a given parameter but also its real-time spatial distribution. Distributed fiber sensors are intrinsic fiber-optic sensors that allow sensing of the variation of

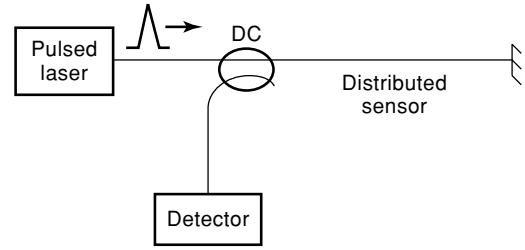


Figure 25. Principle of an OTDR-based distributed sensor.

the measured quantity along the fiber as a continuous function of distance (57). They rely in principle on optical time-domain reflectometry, which is commonly used for diagnostics in fiber telecommunication systems.

Figure 25 shows the principle of operation of the optical time-domain reflectometer (OTDR). A high-power light pulse is launched into the sensing fiber and propagates with the group velocity v_g . The light is reflected at discontinuities and is scattered mainly in the elastic process of Rayleigh scattering, which is caused by the microscopic fluctuations of the refractive index in the fiber. The detector measures the time dependence of the backscattered light signal, and the time of the measurement t determines the distance $z = tv_g/2$ at which the light pulse was backscattered.

Figure 26 presents a sample of the return signal measured by the OTDR. Since the light is attenuated in accordance with an exponential law, the backscattered power P measured by the detector is calculated as (8)

$$P = P_0(1 - \eta)\eta RD \exp(-2 \int \alpha(z) dz) \quad (57)$$

where α is an attenuation coefficient, P_0 the initial pulse power, D the length of the pulse, R the backscattering reflection coefficient per unit length, and η the coupling ratio in the directional coupler. The slope of the logarithm of the detected power at constant reflection R is proportional to the attenuation:

$$\frac{d}{dt} \left(\log \frac{P}{P_0} \right) = -\alpha(z)v_g \quad (58)$$

Thus those parts of the fiber with higher losses are recognized as regions where the detected characteristic has a larger slope,

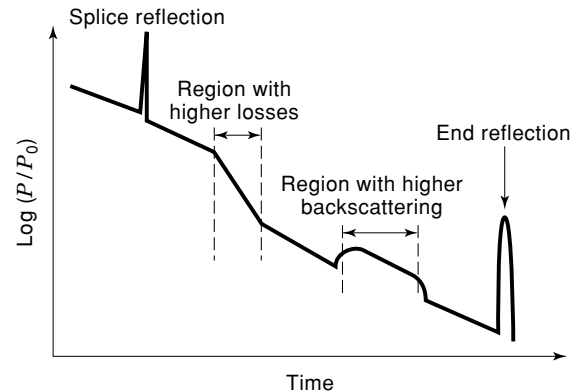


Figure 26. Sample of the detected signal from a distributed fiber-optic sensor.

and those parts with higher backscattering are recognized as regions where there is a higher value of detected power. Any discontinuities (e.g., splices or a fiber end) produce high reflections and jumps in the characteristic of the detected signal.

The OTDR sensors detect the changes of the backscattered reflections R or changes of the losses α induced by the measured quantity. The Rayleigh scattering coefficient depends on temperature, and it can be applied in a temperature sensor (58). Measuring the losses in OTDRs with specially prepared fibers makes it possible to detect the temperature, pressure, liquid leakage, displacements, and other mechanical and chemical measurands. Another possibility is measuring the state of polarization of the backscattered light. Such polarization OTDRs measure changes in birefringence of the fiber, which is sensitive to strain, pressure, magnetic field, electric field, and temperature (59).

Besides the elastic Rayleigh scattering, the inelastic processes of Raman and Brillouin scattering occur in optical fibers. Raman scattering is caused by the molecular vibrations, while Brillouin scattering results from bulk vibrations. Brillouin OTDRs (60,61) and Raman OTDRs have detectors analyzing the spectrum of backscattered light. The ratio of the Stokes and anti-Stokes peaks and the frequency shift in the spectrum can be a source for determining the temperature or strain.

The OTDR allows measurement of changes in the measurand with a spatial resolution approximately equal to 1 m. Another technique, optical-frequency domain reflectometry (OFDR), offers better resolution. It is based on a chirped radio-frequency-modulated light source and on determining the position of light backscattering via the measured frequency.

FIBER-OPTIC SENSOR SYSTEMS

In a simple one-sensor, one-fiber arrangement of a discrete fiber-optic sensor, the optical fiber is largely underutilized as a transmission medium, mostly because the capacity of one information channel is much greater than the information generated by a typical sensor output. Important gains can therefore be made by multiplexing the fiber link by tapping several sensing devices into one passive fiber highway to increase the maximal number of sensors and to establish data telemetry channels. Such a configuration will obviously decrease the installation costs per sensor and at the same time increase the attractiveness of fiber-optic sensing technology for many potential users. Furthermore, industry has a need to install increasing numbers of sensors in surveillance and automation systems in factories, chemical plants, mines, off-shore platforms, aircraft, and spacecraft. Another driving force behind development of multiplexed fiber-optic sensor systems is their close relation to fiber-based LANs; the fact that they utilize the same or similar components will keep their prices low even if the market for industrial fiber-optic sensor systems is only in the early stages of emergence.

Basic Multiplexing Concepts

There are two principal multiplexing concepts that can be applied in designing a sensor system: In one class, a given number of sensors, having only one and the same property modulated by the external measurand field, can be incorporated into a single passive highway loop. Then some kind of multiplexing, addressing, and decoding is added to the system to allow mon-

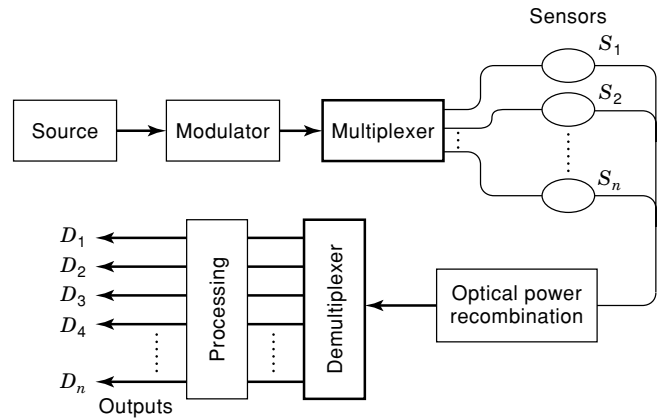


Figure 27. Generalized fiber-optic sensor network.

itoring, detection, and interpretation of a returning output signal from a given sensor at a particular location. A second class involves the so-called orthogonal modulation scheme, where one measurand modulates different optical properties of the sensor. Such a system is capable of simultaneously performing measurements of several different physical parameters.

With the variety of optical fiber sensors currently under development, it does not seem possible yet to characterize the multiplexed system most likely to emerge as the standard, although its topology and parameters will obviously depend upon the type of sensor used, and in almost any multiplexed system at least four basic functions will have to be performed: powering the system at a level adequate for the predesigned power budget of the network, addressing and interrogating (identifying) a sensing device at a chosen discrete location, detecting the measurand-modulated signal of a given sensor, and eventually evaluating and calibrating the acquired individual sensor signals. The topological arrangement of a network, the generalized form of which is shown in Fig. 27, will largely depend on the scheme chosen for sensor addressing and demodulation.

Some of the best-known and most basic network topologies are linear array, ring, reflective/transmissive star, and ladder network. Both passive and active elements such as fiber links, splices, connectors, couplers, polarization controllers, light sources/detectors, modulators, multiplexers/demultiplexers, and sensors are themselves used to implement physical interconnections and components of a multiplexed optical fiber sensor network.

Examples of Discrete Sensor Multiplexing Techniques

Although many different multiplexing schemes and branching network concepts, often very sophisticated and costly, have been extensively investigated and reported (see for instance Refs. 62, 63, 64), industrial applications are usually looking for simpler, lower-key, and cheaper ideas. This short overview of multiplexing techniques starts with such an idea of a simple multiplexing configuration.

Space-Division Multiplexing. Space-division multiplexing (SDM), utilizing separate fiber paths and separate detector-source arrangements for individual sensors, is the easiest method to use and has been already implemented (65). Although it was initially dismissed as a low-key and inelegant

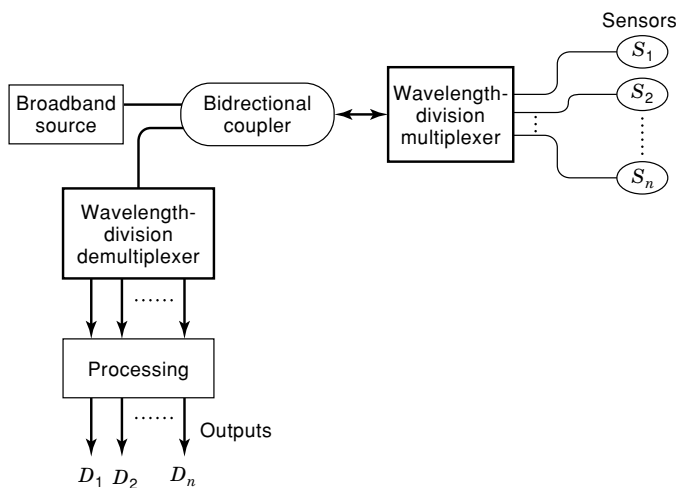


Figure 28. Wavelength-division-multiplexed fiber-optic sensor system.

approach, the rapidly decreasing prices of optoelectronic components (primarily laser sources and detectors) are imposing a reevaluation of the prospects for practical implementation of this method. The power budget of such a system is excellent, crosstalk is nonexistent, and the failure of one channel can usually be tolerated. The method can also be easily combined with a TDM or a WDM scheme (see the two following sub-sub-subsections) at the source–detector terminal unit. Several possible topologies of the SDM method involve common light source with multiple detector array, multiple sources with common detector, single source and single detector with one-to- N and N -to-one switching, and synchronous switching.

Wavelength-Division Multiplexing. The advantage of applying a wavelength-division multiplexing (WDM) scheme (see Fig. 28) for sensor arrays is that it still permits an excellent power budget for the system, as in the case of SDM, while affording savings in the form of a smaller number of necessary fiber links and connections. The light signal sent to each sensor is directed through the appropriate WDM coupling components, which are broadly similar to those designed for fiber communication systems. The most important problem related to this method is obtaining sufficient selectivity of the utilized wavelength division filters; to achieve this, several techniques can be employed, including the very promising spectral splicing technique with a broadband source and narrowband filters such as diffraction or holographic gratings. To further increase the number of multiplexable devices, the technique may obviously be used not only with single-, but also with dual- or multiple-fiber optical links.

An industrial differential-pressure transducer with wavelength-multiplexed power and data channels has been developed (66). It utilizes a resonant element optically excited at one wavelength of 850 nm, the pressure-modulated frequency of which is read out at another wavelength of 633 nm using the same two-wavelength single-fiber link.

Another example of a WDM system involves a digital rotary- and absolute-angular-position sensor utilizing a reflective code plate with ten channels, providing a resolution of 0.35° (67). Two light-emitting diodes with overlapping spectra and a single micro-optic multiplexer–demultiplexer composed of a GRIN rod

lens and a miniature grating are used to disperse the spectrum and recombine the spectral components from each channel after reflection by the code plate. This idea has been proposed for a fly-by-light control of the angular position of a throttle or a fuel control valve on an aircraft.

Time-Division Multiplexing. Time-division-multiplexed (TDM) systems usually require fast and costly electronics and because of this are less attractive for many industrial applications where the cost of installed hardware becomes a dominating factor. In a TDM system, each individual sensor signal can be identified and decoded at the detection end of the setup by arranging different propagation delays for the light signals returning from sensors at different locations. This method can also be coupled with the WDM technique in multiwavelength monitoring systems, with one wavelength affected by the measurand and another used for the reference (68). The method has several important advantages, including the large number of channels, the one-source, one-detector configuration, and equal applicability to both coherent and noncoherent systems. However, usually small optical path differences between the sensors require nontrivial processing. To this end, many complex topologies have already been proposed and reviewed (69). One possible form is a TDM optical passive sensor highway incorporating a commercially available electromechanical switch. Such a system has practically no crosstalk, but a stringent requirement to contain the time-sharing switching sequence within a very short period must be satisfied.

Frequency-Division Multiplexing. The general approach to frequency division multiplexing (FDM) of fiber-optic sensors is to send an amplitude- or frequency-modulated output signal from every sensor in a given network through an assigned frequency channel. The method consists of modulating several light sources by signals of different frequency, or modulating one light source with several frequencies (63) and then combining and separating signals at the detection end of the system employing a multichannel phase-sensitive (usually lock-in-based) detection scheme. One solution is a so-called matrix array method of FD multiplexing of intensity sensors, similar to the one developed by Młodzianowski et al. (70) in a three-sensor configuration. It has much simpler and slower electronics than typical TDM systems, and has a good potential for industrial intensity-modulated sensor multiplexing.

Coherence Multiplexing. Theoretical analysis (71) indicates that numerous WLI sensors with different optical path differences might be interrogated in a serial system by scanning the receiving interferometer. To date, however, a practical realization of only a four-sensor system has been reported (72). There are several factors limiting the number of multiplexed sensors, such as the power budget, the degradation of the contrast of white-light interference (WLI) patterns with increasing number of multiplexed sensors, and the difficulties in proper choice and precise control of optical group imbalances of all sensors in the system in order to assure separation of the signal and noise interference patterns. Furthermore, in many practical applications all multiplexed sensors need to be temperature-compensated. A typical temperature-compensated sensor consists of two equal lengths of highly birefringent (HB) fiber spliced with the polarization axes rotated by 90° , so that its

total group imbalance is close to nil (50). The requirement of temperature compensation conflicts with the principle of coherence multiplexing, which requires significantly different group imbalances between all multiplexed sensors.

Recently an interesting way of overcoming all these difficulties has been reported for sensors based on HB fibers (72). Such sensors are easy to manufacture, especially in that their group imbalances may be controlled simply by measuring the lengths of the fiber components. A system of four serially multiplexed sensors including two temperature-compensated pressure sensors and two temperature sensors has successfully been developed and tested. A modified concept of the temperature-compensated sensor has been adopted for hydrostatic-pressure measurements in order to simultaneously achieve the temperature desensitization of each pressure sensor while maintaining the possibility of individual addressing. The modified pressure sensor consists of three different fiber elements: sensing, compensating, and addressing fibers of lengths L_S , L_C , L_A , respectively. The sensor elements can be made of different types of HB fiber, but their lengths have to satisfy the following condition:

$$K_S^T L_S - K_A^T L_A - K_C^T L_C = 0 \quad (59)$$

where K_S^T , K_A^T and K_C^T are corresponding temperature sensitivities. By choosing the proper lengths and types of fiber for the particular sensor elements it is possible to fulfill the condition of Eq. (59) while setting the total group imbalance ΔR of the sensor at the desired value. In order to avoid overlapping of the noise and the signal interference patterns, the total group imbalance of every additional sensor connected to the serial system has to satisfy the following condition:

$$\Delta R_N = \sum_{i=1}^{N-1} (\Delta R_i + Q_i) \quad (60)$$

where ΔR_i and Q_i are the group imbalance and the operation range of the i th sensor, and the summation is carried out over all sensors already existing in the system.

Examples of Industrial Sensor Systems

As examples of practical implementations, we have selected several application-specific industrial sensor systems reported to have been recently installed in different environments. More information can be found in recent books (7,8) and conference proceedings (see for instance Refs. 54, 68).

Fiber-Optic Stress Monitoring. The fiber-optic stress cell (FOSC) has been developed as a safe and reliable method for embedding and interfacing the FOS (of pressure, load, or stress) in concrete, rock, or soil materials under harsh environmental conditions (73). This simple and practical method combines fiber optics with elements of traditional hydraulic measurement technology based on a compensation method in which stress in the material surrounding the pressure cell is compensated by automatically adjusted pneumatic or hydraulic pressure within it. The assembly of the fiber-optic stress cell is composed of a pressure pad connected to a fiber-optic pressure sensor. The pressure cell is optimized for stress measurements by adjusting the plate surface/thickness ratio to minimize both the measurement error and the influence of the

different modulus of elasticity between the cell and the surrounding material. The cell is attached to a housing containing a fiber-optic polarimetric pressure sensor in a temperature-compensated configuration and equipped with two specially designed pressure leadthroughs to input and output the leading fibers. The connectorized optical leading cables are protected by resistant plastic tubes to withstand the effects of the grouting process during bore-hole installation. Thus, internal pressure induced inside the FOSC, dependent on the stress in the surrounding material, can be directly read out by the fiber-optic pressure sensor and relayed to the remote digital demodulation and measurement unit via long optical cables. A prototype of this device has been successfully installed for load monitoring in a salt mine near Braunschweig, Germany, which is now being used as a nuclear waste disposal site.

Fiber-Optic Thermal Monitoring. An eight-channel fiber-optic sensor network has been installed for thermal monitoring of the stator of a 900 MW turbogenerator (65) by the Bertin Company in France. The system is based on white-light interferometry and uses optical phase demodulation provided by a dual-wavelength passive scheme in a polarimetric configuration (74). The temperature sensor itself is a miniature calcite probe aligned between two polarizers. The eight-channel system is achieved through space-division multiplexing using one LED source at 830 nm per sensor and a common optical and electronic demodulation unit. This unit sequentially interrogates the sensors at a rate of 13 Hz via an 80 m eight-fiber bundle cable and reconstructs the temperatures from the demodulated signals. The accuracy of the system is reported as 0.5°C in the operating range of 20°C to 120°C with accidental rises up to 200°C. The installation is located in the EDF nuclear power plant of Trickastin 1 and is proving successful for high-accuracy temperature monitoring in a harsh environment (high temperature, humidity 93% at 70°C, vibrations at 10 Hz to 100 Hz, and accelerations of 2g).

Fiber-Optic Combustion Monitoring. Real-time pressure measurements on internal combustion engines can provide information for optimum control of engine speed, throttle, air/fuel ratio, and exhaust gas recirculation. Optrand Inc. (USA) is now marketing an intensity-modulated fiber-optic diaphragm sensor integrated within a spark plug, where cylinder pressure flexes the diaphragm and modulates the light reflected from it and coupled back into the input fiber (75). This fiber is metal-coated and temperature-resistant, allowing it to be integrated with the sensor housing and providing a necessary hermetic seal. The sensor is powered from a LED source through a fiber-optic coupler, and the diaphragm is laser-welded to the sensor body. The operating range of this sensor is up to about 7 MPa (1000 psi) with overpressure up to 21 MPa. The range of temperatures is -40° to 550°C, and the frequency range is 0.1 kHz to 15 kHz. The overall accuracy of this system is reported to be better than 5%.

Fiber-Optic Damage Detection Monitoring. Fiber-optic submarine communication cables are exposed to a variety of damage risks from anchoring or fishing equipment. Early detection of such incipient damage can significantly improve the reliability of undersea transmission installations. To satisfy this requirement, a prototype 66 kV XLPE fiber-optic communication cable containing an integrated fiber-optic mechanical damage

detection sensor has been developed in Japan and is now used in real installations (76). The detection part contains four single-mode fibers placed in shallow grooves in the outside part of the cable every 90°. When lateral force is applied to the cable, some of the monitoring fibers become laterally compressed and their transmission losses greatly increase. Monitoring of these losses by the OTDR system means that the increasing risk of damage to the cable can be detected long before any degradation of the actual optical transmission becomes visible, and corrective action can be undertaken in time to save the cable. The system is able to detect cable deformation rates from 10%, whereas the deterioration of communications is normally noticed only at 40% cable deformation.

Fiber-Optic Traffic Monitoring. A fiber-optic airport ground traffic surveillance system has been installed in the Braunschweig airport in Germany to detect and identify vehicles or aircraft movement via monitoring of structural vibrations (77). The system is based on a fiber-optic extrinsic Fabry–Perot microinterferometer cantilever-beam-type vibration sensor. The Fabry–Perot cavity is formed by the air gap between the end of the sensing fiber and a low-reflectivity mirror, and its length depends on the vibrations. The sensing element is powered from a low-coherence SLD at $\lambda = 1300$ nm via a bidirectional coupler. The reflected modulated signal returns through the same coupler and is split into two components before entering a two-wavelength passive quadrature demodulation unit, which converts the two phase-shifted interference signals into output voltages. The sensors are located at taxiway intersections with runways and cemented at the edge of the taxiways without damaging the pavement. Each sensor is connected to the system via one of eight connectorized single-mode optical fibers inside a common optical cable about 1 km long running to the control unit. The reported system has now been in continuous operation for more than two years and has proven successful under all weather conditions.

Fiber-Optic Navigation Monitoring. The fiber-optic gyro (FOG), discussed earlier, is an open-loop ring interferometric rotation sensor and is the first mass-produced FOS for applications in vehicle, missile, or aircraft navigational systems (78). The system is manufactured by Hitachi and is currently installed, among other applications, in luxury automobiles. The sensor is entirely composed of polarization-maintaining fiber, uses an SLD source, and has a CMOS-integrated high-precision signal-processing system based on DSPs. The accuracy of rotation measurement can be as high as 0.01°/h and is stable within a temperature range from –30° to 80°C. The system is mass-produced at a maximum rate of 5,000 units per month and can be used for automotive navigation in planned intelligent transportation systems to overcome heavy traffic congestion or in other applications such as attitude control systems or direction measuring systems.

Fiber-Optic Displacement Monitoring. A WLI system of fiber-optic displacement sensors for monitoring the integrity of civil structures (SOFO) has been developed and implemented in Switzerland (79). It is based on a double all-fiber reflective Michelson interferometer in tandem configuration (80), with the sensing fiber pretensioned and coupled to the measured structure and with a free temperature-compensation fiber. The system is powered by an SLD at 1300 nm, and one arm of the

analyzing Michelson interferometer is terminated by a mobile mirror. The sensor length may vary from 20 cm to 50 m, and the resolution is 2 μ m; the dynamic range for measuring elongation is 1%. The sensors can be connected serially in chains by coherence multiplexing. It has been recently reported that this system has already been installed on four bridges and in several tunnels and dams, with a total of 600 sensors being installed since 1993 (79).

Fiber-Optic Current Monitoring. Commercial optical current transformers based on Faraday rotation in Bi-substituted rare earth iron garnet (BiRIG) materials have been developed by Panasonic in Japan (81) for metering in electric power systems. The optical magnetic field sensor head itself can take the form of a confocal sensor or a miniature fiber-integrated sensor. The polarizer, analyzer, and leading fibers are integrated together within the sensor head, and the intensity-modulated signal is acquired in a polarimetric configuration. It has been reported that the appropriate combination of garnet composition and optical design can yield 1% of sensor accuracy within a temperature range from –20°C to 80°C, with an ac linearity of 1% for input current up to 300 A.

FUTURE OF FIBER-OPTIC SENSORS

During the last 15 years, extensive research has been done on every imaginable fiber-optic sensing mechanism (3), making it rather difficult to come up with an entirely novel and exciting sensing idea. Many of the ideas studied, however, remain mere ideas, with no chance of practical implementation in a real environment. A promising trend is now developing in that more and more research groups are taking approaches that are application-oriented, and consequently such implementation issues as packaging, interconnecting, and above all reliability and multiplexing of fiber-optic sensors and systems are becoming the focus of attention. Research on FOS technology for industrial applications is still, however, in the early stages of development.

Three important aspects have to be addressed simultaneously in the near future for wider successful commercialization of these systems in direct competition with traditional sensing systems. These aspects are: (1) developing accurate, reliable, and cost-effective sensor systems; (2) safe packaging and interfacing of sensors with their working environment; (3) optical multiplexing, signal processing, and remote data telemetering for the installed sensor arrays. Research in these directions will certainly continue to increase, since the technology remains genuinely attractive in many application areas.

BIBLIOGRAPHY

1. G. Cancellieri (ed.), *Single-Mode Optical Fiber Measurement: Characterization and Sensing*, Boston: Artech House, 1993.
2. J. Noda, K. Okamoto, and Y. Sasaki, Polarization-maintaining fibers and their applications, *J. Lightwave Technol.*, **4**: 1071–1089, 1986.
3. P. F. Wysocki et al., Characteristics of erbium-doped superfluorescent fiber sources for interferometric sensor applications, *J. Lightwave Technol.*, **12**: 550–567, 1994.
4. J. Saether and K. Blotekjaer, Noise performance of multiplexed fiber-optic sensor systems with optical amplifiers, *Opt. Rev.*, **4**: 138–144, 1997.

5. C. Tsao, *Optical Fibre Waveguide Analysis*, Oxford: Oxford Univ. Press, 1992.
6. M. J. F. Digonnet and H. J. Shaw, Analysis of a tunable single mode optical fiber coupler, *IEEE J. Quantum Electron.*, **QE-18**: 746–754, 1982.
7. B. Culshaw and J. Dakin (eds.), *Optical Fiber Sensors*, Norwood, MA: Artech House, 1989, Vols. 1, 2.
8. E. Udd (ed.), *Fiber Optic Sensors—An Introduction for Engineers and Scientists*, New York: Wiley, 1991.
9. D. C. Marvin and N. A. Ives, Wide-range fiber-optic strain sensor, *Appl. Opt.*, **23**: 4212–4217, 1984.
10. J. W. Berthold, Historical review of microbend fiber-optic sensors, *J. Lightwave Technol.*, **13**: 1193–1199, 1995.
11. D. Dongalic and M. Zavrsnik, Fiber-optic microbend sensor structure, *Opt. Lett.*, **22**: 837–839, 1997.
12. W. B. Spillman and J. R. Lord, Self-referencing multiplexing technique for intensity modulating sensors, *Proc. SPIE*, **718**: 182–191, 1987.
13. J. W. Berthold et al., Design and characterization of a high temperature pressure transducer, *J. Lightwave Technol.*, **5**: 870–876, 1987.
14. B. G. Grossman et al., Fiber optic pore pressure sensor development, *Transportation Res. Rec.*, **1432**: 76–85, 1994.
15. O. Lumholt et al., Simple low-temperature sensor that uses microbending loss, *Opt. Lett.*, **16**: 1355–1357, 1991.
16. J. B. Freal, C. J. Zarobila, and C. M. Davis, A microbend horizontal accelerometer for borehole deployment, *J. Lightwave Technol.*, **5**: 993–996, 1987.
17. J. D. Weiss, Strain gage, *J. Lightwave Technol.*, **7**: 1308–1318, 1989.
18. D. Varshneya, J. L. Maida, and M. A. Overstreet, Speed sensor for advanced gas turbine engine control, *Proc. SPIE*, **1367**: 181–191, 1991.
19. P. Raatikainen et al., Fiber-optic liquid-level sensor, *Sensors Actuators*, **A58**: 93–97, 1997.
20. G. Betta, A. Pietrosanto, and A. Scaglione, A digital liquid level transducer based on optical fiber, *IEEE Trans. Instrum. Meas.* **45**: 551–555, 1996.
21. W. J. Bock, M. Beaulieu, and A. Domanski, GaAs-based fiber-optic pressure sensor, *IEEE Trans. Instrum. Meas.* **41**: 68–71, 1992.
22. Nortech Fibronic Inc., Québec, Canada, 1997, Catalog.
23. B. Culshaw, J. Foley, and I. P. Giles, A balancing technique for optical fiber intensity modulated transducers, *Proc. SPIE*, **574**: 117–120, 1984.
24. W. J. Bock and T. A. Eftimov, Single- and few mode fiber-optic pressure sensors, in F. Ansari (ed.), *Application of Fiber-Optic Sensors in Engineering Mechanics*, New York: American Society of Civil Engineers, 1993, pp. 252–263.
25. T. A. Eftimov, Resultant mode pattern and polarization in a LP₀₁, LP₀₂ two-mode linearly birefringent optical fibre, *Opt. Quantum Electron.*, **23**: 1143–1160, 1991.
26. A. W. Snyder and J. D. Love, *Optical Waveguide Theory*, London: Chapman and Hall, 1983.
27. W. J. Bock and T. R. Wolinski, Temperature-compensated strain sensor based on polarization-rotated reflection, *Proc. SPIE*, **1370**: 189–196, 1991.
28. W. J. Bock et al., Cross-sensitivity effect in temperature compensated sensors based on highly birefringent fibers, *Appl. Opt.*, **33**: 6078–6083, 1994.
29. G. W. Day, K. B. Rochford, and A. H. Rose, Fundamentals and problems of fiber current sensors, *Proc. Int. Conf. Optical Fiber Sensors OFS-11*, Japan Society of Applied Physics, Sapporo, Japan, 1996, pp. 124–129.
30. G. W. Day and A. H. Rose, Faraday effect sensors: The state of the art, *Proc. SPIE*, **985**: 138–151, 1988.
31. M. Willsch, T. Bosselmann, and M. Bretschneider, Investigation of the vibration sensitivity of different annealed fiber optic current transformers, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, Optical Society of America, 1997, pp. 642–645.
32. J. E. Lenz, C. D. Anderson, and L. K. Strandjord, Magnetic materials characterization using a fiber optic magnetometer, *J. Appl. Phys.*, **57**: 3820–3822, 1985.
33. M. Ko and E. Garmire, Current sensing using bismuth rare-earth iron garnet films, *Appl. Opt.*, **34**: 1692–1696, 1995.
34. N. Inoue and K. Yamasawa, Stabilization of temperature dependence of Verdet constant of Bi-doped garnet and development of high sensitive optical fiber magnetic field sensor, *Electr. Eng. Jpn*, **117**: 1–10, 1996.
35. P. I. Nikitin et al., Fiber-optical magnetic field sensors based on Faraday effect in new materials, *Proc. SPIE*, **1584**: 124–134, 1992.
36. T. Valis, D. Hogg, and R. Measures, Fiber-optic Fabry–Perot strain gauge, *IEEE Photonics Technol. Lett.*, **2**: 227–228, 1990.
37. T. W. Kao and H. F. Taylor, High-sensitivity intrinsic fiber-optic Fabry–Perot pressure sensor, *Opt. Lett.*, **21**: 615–617, 1996.
38. T. Liu et al., Simultaneous strain and temperature measurement using a combined fiber Bragg grating/extrinsic Fabry–Perot sensor, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 40–43.
39. T. S. Priest, G. B. Scelsi, and G. A. Woolsey, Optical fiber sensor for electric field and electric charge using low-coherence, Fabry–Perot interferometry, *Appl. Opt.*, **36**: 4505–4508, 1997.
40. M. Born and E. Wolf, *Principles of Optics*, Oxford: Pergamon, 1986.
41. P. C. Beard and T. N. Mills, Extrinsic optical-fiber ultrasound sensor using a thin polymer film as a low-finesse Fabry–Perot interferometer, *Appl. Opt.*, **35**: 663–675, 1996.
42. K. A. Murphy et al., Quadrature phase-shifted, extrinsic Fabry–Perot optical fiber sensors, *Opt. Lett.*, **16**: 273–275, 1991.
43. N. Fürstenau and M. Schmidt, Fiber-optic extrinsic Fabry–Perot interferometer vibration sensor with two-wavelength passive quadrature readout, *IEEE Trans. Instrum. Meas.*, **47**: to be published.
44. G. Boheim, Fiber-linked interferometric pressure sensor, *Rev. Sci. Instrum.*, **58**: 1655–1659, 1987.
45. Y.-J. Rao and D. A. Jackson, Recent progress in fiber optic low-coherence interferometry, *Meas. Sci. Technol.*, **7**: 981–999, 1996.
46. H.-S. Choi and H. F. Taylor, High performance fiber optic temperature sensor using low-coherence interferometry, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 570–573.
47. S. Chen, A. J. Rogers, and B. T. Meggitt, Electronically scanned optical-fiber Young's white-light interferometer, *Opt. Lett.*, **16**: 761–763, 1991.
48. Y.-J. Rao and D. A. Jackson, Prototype fiber-optic-based ultrahigh pressure remote sensor with built-in temperature compensation, *Rev. Sci. Instrum.*, **65**: 1695–1698, 1994.
49. R. Dandliker, E. Zimmermann, and G. Frosio, Noise-resistant signal processing for electronically scanned white-light interferometry, *Proc. Int. Conf. Optical Fiber Sensors OFS-8*, Monterey, CA, OSA, 1992, pp. 53–56.
50. W. J. Bock, W. Urbanczyk, and M. B. Zaremba, Electronically scanned white-light interferometric strain sensor employing HB fibers, *Opt. Commun.*, **101**: 157–162, 1993.
51. W. J. Bock et al., White-light interferometric fiber-optic pressure sensor, *IEEE Trans. Instrum. Meas.*, **44**: 694–697, 1995.
52. G. Meltz, Overview of fiber grating-based sensors, *Proc. SPIE*, **2838**: 2–22, 1996.

53. H. Kogelnik, Theory of dielectric waveguides, in T. Tamir (ed.), *Integrated Optics*, Berlin: Springer-Verlag, 1979, pp. 66–79.
54. P. Ferdinand et al., Applications of Bragg grating sensors in Europe, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 14–19.
55. J. D. C. Jones, Review of fiber sensor techniques for temperature-strain discrimination, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 36–39.
56. G. A. Sanders et al., Progress in high performance fiber optic gyros, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 116–121.
57. J. P. Dakin (ed.), *The Distributed Fiber Optic Sensing Handbook*, Berlin: Springer-Verlag, 1990.
58. A. H. Hartog, A distributed temperature sensor based on liquid-core optical fibers, *J. Lightwave Technol.*, **1**: 498–509, 1983.
59. M. Tsubokawa, T. Higashi, and Y. Negishi, Mode couplings due to external forces distributed along a polarization maintaining fiber: An evaluation, *Appl. Opt.*, **27**: 166–173, 1988.
60. T. R. Parker et al., Temperature and strain dependence of the power level and frequency of spontaneous Brillouin scattering in optical fibers, *Opt. Lett.*, **22**: 787–789, 1997.
61. T. Horiguchi et al., Development of a distributed sensing technique using Brillouin scattering, *J. Lightwave Technol.*, **13**: 1296–1302, 1995.
62. M. Fahadrousan and R. C. Youngquist, Polarimetric coherence multiplexing using high-birefringence optical-fiber sensors, *Opt. Lett.*, **15**: 786–888, 1990.
63. B. D. Zimmermann et al., Fiber-optic sensors using high-resolution optical time domain instrumentation systems, *J. Lightwave Technol.*, **8**: 1273–1277, 1990.
64. J. Castracane and L. P. Clow, 64 channel fiber optic spectrometer for a dual wavelength interferometric pressure sensor array, *Rev. Sci. Instrum.*, **66**: 3668–3671, 1995.
65. C. Meunier et al., Industrial prototype of a fiber-optic sensor network for the thermal monitoring of the turbogenerator of a nuclear power plant—design, qualification and settlement, *J. Lightwave Technol.*, **13**: 1354–1361, 1995.
66. E. Bois, R. C. Spooner, and B. E. Jones, A hybrid resonant differential pressure transmitter with wavelength-multiplexed power and data channels, in H. J. Arditty, J. P. Dakin, and R. T. Kersten (eds.), *Optical Fiber Sensors*, Springer Proceedings in Physics 44, New York: Springer-Verlag, 1989, pp. 478–483.
67. K. Fritsh, G. Beheim, and J. Sotomajor, Digital angular position sensor using wavelength division multiplexing, *Proc. SPIE*, **1169**: 453–460, 1989.
68. B. Deboux et al., A time division multiplexed multiwavelength optical fiber pH sensor, *Proc. Int. Conf. Optical Fiber Sensors OFS-11*, Sapporo, Japan, 1996, pp. 466–469.
69. S. Koto et al., Multipoint sensing laser Doppler velocimetry based on laser diode frequency modulation, *Proc. Int. Conf. Optical Fiber Sensors OFS-11*, Sapporo, Japan, 1996, pp. 606–609.
70. J. Młodzianowski, D. Uttamchandani, and B. Culshaw, A simple frequency domain multiplexing system for optical point sensors, *J. Lightwave Technol.*, **5**: 1002–1007, 1987.
71. R. H. Wentworth, Theoretical noise performance of coherence-multiplexed interferometric sensors, *J. Lightwave Technol.*, **7**: 941–956, 1989.
72. W. J. Bock and W. Urbanczyk, Multiplexing of white-light interferometric fiber-optic sensors for hydrostatic pressure and temperature measurements, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 479–483.
73. W. J. Bock, W. Urbanczyk, and M. R. H. Voet, Performance and application of fiber-optic pressure cell for direct stress measurements, *Proc. Canadian Inst. Mining Conf.*, Edmonton, Canada, 1996, CD ed.
74. M. Lequime et al., A dual wavelength passive homodyne detection unit for fiber coupled white light interferometers, *Proc. SPIE*, **1267**: 288–293, 1990.
75. M. Włodarczyk and G. He, A fiber-optic combustion pressure sensor system for automotive engine control, *Sensors*, **11**: 35–42, 1994.
76. T. Nishimoto et al., Development of 66 kV XLPE submarine cable using optical fiber as a mechanical-damage-detection-sensor, *IEEE Trans. Power Delivery*, **10**: 1711–1717, 1995.
77. N. Fürstenau et al., Extrinsic Fabry–Perot interferometer vibration and acoustic sensor systems for airport ground traffic monitoring, *IEE Proc. Optoelectron.*, **144**: 134–144, 1997.
78. H. Kajioka et al., Mass-produced I-FOGs for commercial applications, *Proc. Int. Conf. Optical Fiber Sensors OFS-11*, Sapporo, Japan, 1996, pp. 68–71.
79. D. Inaudi, Field testing and application of fiber-optic displacement sensors in civil structures, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 596–599.
80. D. Inaudi et al., Low-coherence deformation sensors for the monitoring of civil-engineering structures, *Sensors and Actuators*, **A44**: 125–130, 1994.
81. N. Itoh et al., Commercial current sensor activity in Japan, *Proc. Int. Conf. Optical Fiber Sensors OFS-12*, Williamsburg, VA, OSA, 1997, pp. 92–95.

WOJTEK J. BOCK
Université du Québec à Hull

MIROSLAW A. KARPIERZ
Warsaw University of Technology

ULTRAVIOLET DETECTORS

The detection of ultraviolet (UV) radiation is of importance in many applications in science, industry, and medicine, such as flame and spark detection, communications, UV dosimetry of industrial processes such as UV-initiated photopolymerization reactions, and sterilization of equipment and water. The UV light spectrum is generally considered to comprise wavelengths from 10 nm to 400 nm. These wavelengths are traditionally subdivided into several ranges, given in Table 1. Photodetection of ultraviolet light within these ranges poses specific challenges in terms of the fabrication and engineering of photodetector devices and the development of appropriate materials. An excellent review (1) of all types of photodetectors, including ultraviolet-sensitive photodetectors, has been published

Table 1. Ultraviolet Light Nomenclature

UV Region	Wavelength Range (nm)
Near UV	400–300
Mid UV	300–200
Far UV	200–100
Extreme UV	100–10
UVA	400–320
UVB	320–280
UVC	280–100
Deep UV	350–190
Actinic UV	315–200
Vacuum UV	200–10

recently. A detailed review (2) of ultraviolet- and X-ray-sensitive detectors that addresses both imaging and nonimaging systems has also been recently published. This article discusses the current commercially available technologies for nonimaging UV photodetection applications. In addition, in view of the fact that UV photodetection is an active and important field of research, emerging technologies are also briefly discussed. The article is divided into two sections, the first addressing UV detection with semiconductor photodiodes and the second covering UV detection with photoemissive phototubes.

SEMICONDUCTOR ULTRAVIOLET PHOTODETECTORS

Semiconductor detectors, used in either the photoconductive or photovoltaic modes, are widely used photoelectric detectors of ultraviolet light. Razeghi and Rogalski (3) have recently published a comprehensive review of ultraviolet-sensitive semiconductor detectors that describes in detail the theory and operation of each different type of detector. Semiconductor detectors are attractive due to their small size, simple operation, and relative low cost, and, for many applications, provide stable, linear detection of ultraviolet light over a wide dynamic range of incident powers. Silicon traditionally has been the most widely used semiconductor material for the fabrication of ultraviolet-sensitive photodetectors; however, its use is limited by radiation-induced aging effects, problems arising from the strong absorption of UV light, and an inability to operate at elevated temperatures. Si photodetectors are also much more sensitive to visible and near-IR light than to UV light. However, silicon wafer processing technology is quite advanced and Si photodetectors have been engineered with features designed to enhance the ultraviolet light sensitivity. The most widely used UV enhanced device structures are junction photodetectors, including, for example, shallow diffused p - n junctions and Schottky barriers. In this section we will discuss the characteristics and relative merits of shallow diffused p - n junction and Schottky barrier photodiodes, fabricated using a number of materials, including silicon and several wider-band-gap materials such as silicon carbide (SiC), III-V nitrides (GaN, AlGaN) and III-V phosphides (GaAsP, GaP). Wide-band-gap materials are better suited in many ways than Si for UV photodetection. For example, wide-band-gap photodetectors exhibit improved radiation hardness, better resistance to temperature and harsh chemicals, and improved wavelength selectivity (visible blind operation). Consequently, the development of wide-band-gap materials and their application in photodetection is an active area of research.

Silicon p - n Junction Photodiode

Photodiodes fabricated using p - n junctions in silicon have been the most widely used photovoltaic detectors for ultraviolet photodetection. Silicon photodiodes exhibit good responsivity, on the order of 0.1 to 1 A/W, over the entire visible and near-IR spectrum. Although the UV response in a standard p - n photodiode is significantly lower than the peak response in the near-IR region, photodiodes are manufactured that exhibit enhanced blue and UV response. Due to the broad spectral response, if visible blind operation is desired, the incident light must be spectrally filtered to remove the visible and near-IR wavelengths. The typical spectral response for a Si p - n junction is shown in Fig. 1, exhibiting a peak at ~ 900 nm and a

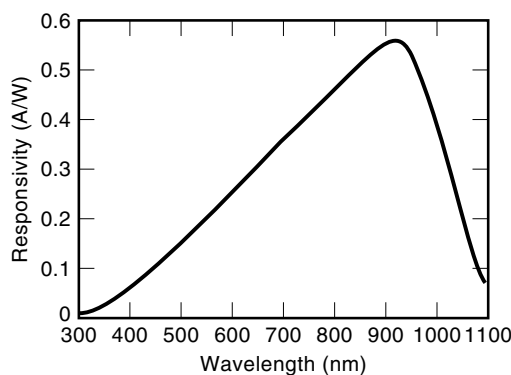


Figure 1. Spectral response of a standard p - n junction Si photodiode.

broad long-wavelength cutoff extending beyond 1100 nm, due to the indirect nature of the Si band gap. The response in the UV region is essentially zero for wavelengths shorter than ~ 300 nm. The construction of a typical silicon p - n junction photodiode is shown in Fig. 2. In this example, a single-crystal, high-purity, n -type silicon substrate is doped with a material, such as boron, to form a thin p layer on one surface. The p layer may be formed by thermal diffusion or ion implantation of the boron to depths of ~ 1 μm or less. The interface between the p -doped Si and the n -doped Si comprises the p - n junction. p - n junction photodiodes operate by the absorption of light in the bulk of the semiconductor. In the figure, light is incident on the surface with the diffused p layer. Since Si has a high absorption coefficient, particularly for blue and UV wavelengths, the p layer at this surface must be thin enough to permit good interaction of the light with the bulk semiconductor near the p - n junction. The absorption of light creates electron-hole pairs that are subsequently separated by the electric field of the p - n junction. The electrons move toward the n layer and holes move toward the p layer. The notation p^+ and n^+ designate heavily doped p and n layers, respectively. Such layers have high conductivity and are used to ensure good electrical contact between the semiconductor and the metal conductor. The open-circuit voltage generated by the photoinduced charge is extremely linear with respect to the intensity of the light. If

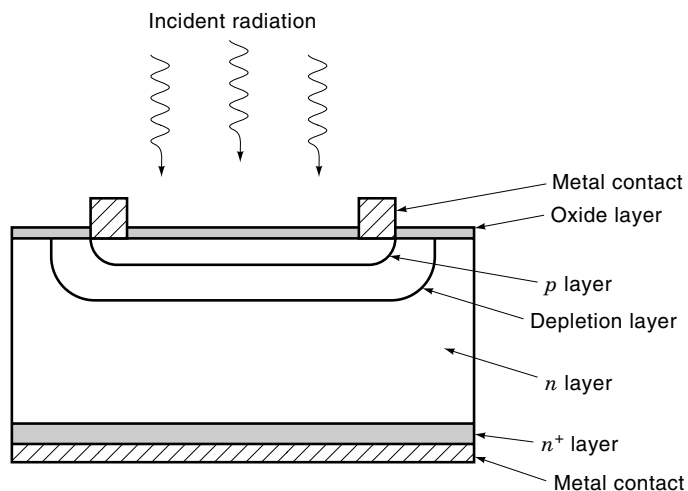


Figure 2. Schematic of a typical p - n junction Si photodiode.

the anode and cathode of the photodiode are connected across a resistive load, this internal charge migration is measured externally as a photocurrent that is similarly found to be linear with the light intensity over a wide dynamic range. The operation of p - n junction photodiodes in this zero-bias photovoltaic mode is used for applications requiring the greatest sensitivity; however, increased speed and linearity can be attained by applying a reverse bias (the photocurrent remains the same). The reverse bias reduces the capacitance of the device, thereby decreasing the RC time constant and improving the temporal response of the photodiode. However, the bias also increases the dark current, resulting in an overall reduction in sensitivity. There is therefore a trade-off between sensitivity and speed in an ordinary p - n junction photodiode.

For the specific application of ultraviolet light detection the p - n junction photodiode must be optimized in a number of ways to enhance the quantum efficiency of the device to UV light and to avoid the loss of carriers to nonradiative recombination processes, particularly at surfaces and interfaces. The quantum efficiency is defined as the number of electron-hole pairs per incident photon that are generated and which are subsequently separated by the electric field of the p - n junction before recombination occurs. High quantum efficiency requires that the incident light be able to penetrate into the bulk of the semiconductor material to generate carriers near the junction. Due to the high absorptivity of Si the penetration depth of light is very small and standard Si photodiodes are fabricated with a relatively thin layer, $\sim 1\ \mu\text{m}$ or less, of doped material on the surface upon which the light is incident. Figure 3 shows the absorption coefficient of Si as a function of wavelength (4). For the wavelength range from 10 nm to 400 nm the absorption length is everywhere less than 100 nm and reaches a minimum value of 3 nm at a wavelength of ~ 300 nm. Thus, the quantum efficiency in the UV region is reduced, even for submicrometer-thick layers, because of the exceptionally strong absorptivity at these wavelengths. Obviously, the thickness of the doped layer must be kept as thin as possible to enhance the UV response. It follows that the shallow p - n junction lies very near the surface of the photodiode. Because of the high doping levels typically used and the close proximity of the charge carriers with the surface, nonradiative recombination of carriers can cause a significant loss of quantum efficiency. In order to minimize the probability of carrier recombination the surface can be passivated with a thin coating of a material such as silicon dioxide. This oxide coating is also intended to protect the photodiode surface from moisture, chemicals, and scratches and serves as

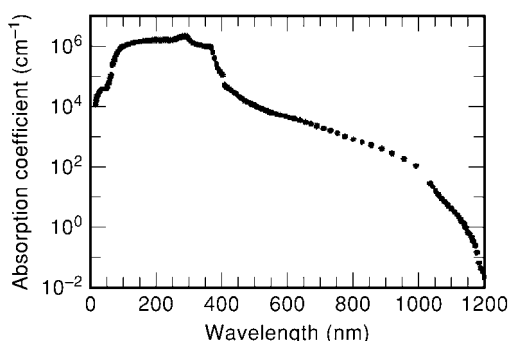


Figure 3. Absorption coefficient of Si as a function of wavelength for UV, visible, and near-infrared wavelengths.

an antireflection coating to reduce incident light losses. Control of the thickness and quality of the oxide coating is important for optimizing the UV response. The quality of this coating and the nature of the interaction at the interface between the coating and the semiconductor are largely responsible for aging effects that have been observed in the UV response of Si photodiodes. This has been a significant problem for Si photodiodes, with extended UV exposure, even at low power, capable of causing substantial degradation in the response (5–7). The quantum efficiency of a Si photodiode decreased (5) from an initial value of 27% to less than 8% when illuminated with 10^{10} photons/(s · mm²) at 124 eV for less than 30 min. UV-enhanced Si photodiodes having superior stability as well as high quantum efficiency (8,9) were manufactured by diffusing arsenic or phosphorus in a thin, highly doped layer on p -type Si. The superior characteristics of the photodiodes were attributed to the n on p construction and careful control of the quality and thickness of the oxide coating. These photodiodes have been further improved more recently (10,11) by nitriding the oxide layer to yield an oxynitride layer that possesses significantly improved radiation hardness and resistance to moisture. Because of the ruggedness of the 60 Å thick oxynitride layer and the virtual absence of nonradiative recombination at the oxynitride/silicon interface, the photodiodes possess high quantum efficiency with exceptional long-term stability. The same photodiodes, with the responsivity calibrated as a function of wavelength in the vacuum UV, are available from the National Institute of Standards and Technology (NIST) as transfer standard detectors. A schematic of the photodiode is shown in Fig. 4 and the wavelength dependence of the responsivity is shown in Fig. 5.

To summarize, UV-enhanced Si photodiodes are widely used for UV photodetection because they can be manufactured using well-established wafer-scale processing methods and they are relatively inexpensive. They are commercially available from a number of manufacturers (12). However, their response in the UV region is much smaller than their response in the visible and near-IR regions so that if solar blind operation is required, then optical filters, such as colored glass filters and dielectric coated band-pass filters, must be used with the photodiode to block the longer wavelengths. In addition, the quantum efficiency in the UV region can vary significantly with wavelength due to the rapidly varying absorption coefficient

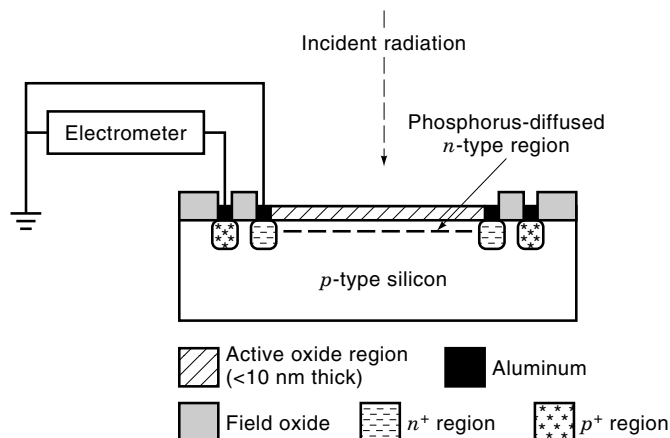


Figure 4. Schematic of a state-of-the-art UV-enhanced Si photodiode. (Courtesy of the Physics Laboratory, National Institute of Standards and Technology.)

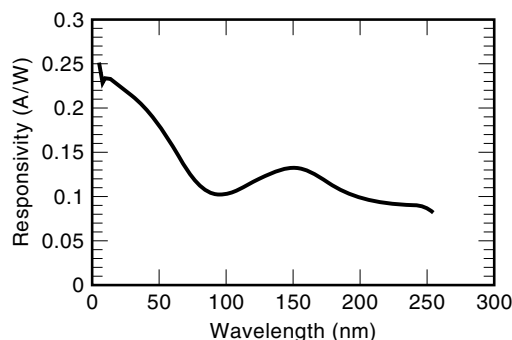


Figure 5. Spectral response of UV-enhanced Si photodiode. (Courtesy of the Physics Laboratory, National Institute of Standards and Technology.)

of Si at UV wavelengths, the related problem of nonradiative recombination losses close to the surface due to the short absorption lengths and because of the finite probability of creating more than one electron-hole pair per UV photon. Finally, when choosing a Si photodiode, the stability of the UV response with age and with accumulated UV exposure is an important consideration.

GaP and GaAsP Schottky Photodiodes

The second semiconductor photodiode architecture that is well suited for UV photodetection applications is the Schottky photodiode. In the diffusion-type p - n junction photodiode, discussed above, the depletion layer is formed between the p -doped and n -doped layers of the semiconductor while in a Schottky photodiode the depletion layer is formed at the interface between the semiconductor and a thin layer of metal, such as gold or platinum. Since the photosensitive region is located nearer to the surface, the Schottky barrier photodiode is capable of superior short-wavelength detection than the p - n junction if the material is highly absorbing. A schematic of a Schottky barrier photodiode is shown in Fig. 6. In this figure a thin layer of gold is deposited on an n -type substrate. The gold layer is thin enough, typically about 10 nm, to permit ultraviolet light to penetrate into the semiconductor layer. Photogenerated electron-hole pairs are collected in the depletion layer near the metal/semiconductor interface and are separated by the depletion layer field, giving rise to an external photocurrent, just as in the case for the p - n junction photodiode. Schottky barrier photodiodes are relatively simple to fabricate compared to diffused junction photodiodes, since the junction is formed by the deposition of a single thin layer of metal on the semiconductor surface. The spectral response of the resulting photodiode depends on the response of the semiconductor. Ultraviolet sen-

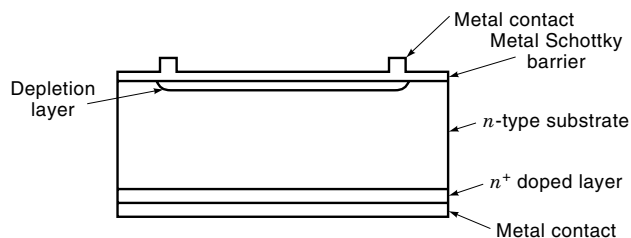


Figure 6. Schematic of a typical Schottky barrier photodiode.

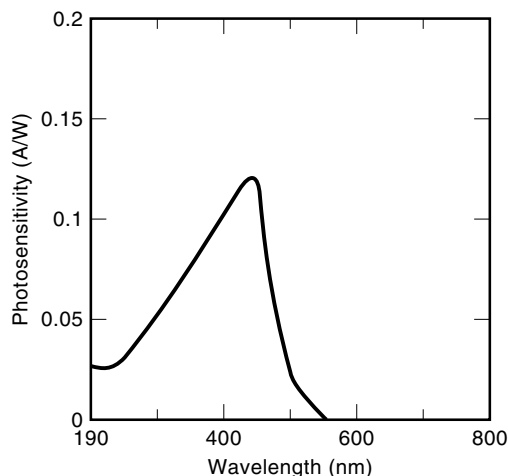


Figure 7. Spectral response of a GaP Schottky barrier photodiode. (Courtesy of Hamamatsu Corporation.)

sitive Schottky photodiodes have been fabricated using Si, however, it was found that the quantum efficiency degraded with exposure in a manner essentially identical to that observed for p - n junction Si photodiodes (5). Commercially available (13) alternatives to Si photodiodes include GaP and GaAsP Schottky-type photodiodes. These photodiodes are manufactured using a thin layer of gold on an n -type semiconductor substrate. The stability of GaP and GaAsP Schottky photodiodes is superior to that of Si photodiodes, and it has been shown that (5) the quantum efficiencies of the GaP and GaAsP Schottky photodiodes remain constant over time with UV irradiation. Both GaP and GaAsP photodiodes have good sensitivity throughout the entire UV region and are more resistant to radiation damage than Si photodiodes. The spectral responsivity of a GaP Schottky photodiode is shown in Fig. 7. Since GaP and GaAsP are wider-band-gap materials than Si, they offer greater selectivity in the wavelength response. GaAsP photodiodes exhibit a peak response at ~ 600 nm and a relatively sharp cutoff at ~ 680 nm, while the peak of the response for GaP is ~ 450 nm, with a cutoff at ~ 550 nm. Although the quantum efficiencies of the GaP and GaAsP photodiodes are lower than the quantum efficiency for Si photodiodes, the spectral responsivities of GaP and GaAsP in the ultraviolet region are smoothly varying functions of wavelength. This is in contrast to the responsivity of Si, which exhibits abrupt excursions throughout the UV region due to multiple photon ionization. The rise time of the temporal response of the GaP and GaAsP Schottky photodiodes is significantly longer than that of a p - n junction photodiode, ranging from $5 \mu\text{s}$ to $50 \mu\text{s}$ for photodiodes having active areas that range from $1 \text{ mm} \times 1 \text{ mm}$ to $10 \text{ mm} \times 10 \text{ mm}$, respectively. This is because the capacitance of the Schottky barrier photodiodes are significantly larger as a result of the extremely thin depletion layer. The surface barrier structure also leads to higher dark currents in the GaP and GaAsP photodiodes compared to Si photodiodes.

Wide-Band-Gap Materials

The semiconductor photodetectors discussed thus far have differed with respect to the substrate material and the nature of the barrier that separates the photogenerated carriers. In all cases the responsivity of the photodiode was greater in the

visible and near-IR regions than in the ultraviolet region. For many applications it is desirable for a photodetector to exhibit high responsivity to ultraviolet wavelengths and negligible responsivity to visible and near-IR wavelengths (greater than 400 nm). This kind of operation is often referred to as visible-blind detection. Solar-blind detection is a class of visible-blind operation having even more stringent wavelength requirements, exhibiting high responsivity to only those wavelengths less than 290 nm. Solar blind operation is impossible to achieve using Si photodetectors without the use of filters such as UV-transmitting-visible-blocking colored-glass filters usually in combination with dielectric-coated band-pass filters. The use of these spectral filtering techniques, however, imposes certain restrictions on the operation of the resulting photodiode-filter assembly. Most band-pass filters are made by combining dielectric stack interference filters with additional colored-glass or metal-film blocking filters. While these band-pass filters offer some flexibility in the specification of the center wavelength (or edge wavelength) and the width of the band, the overall transmission off-band cannot be reduced to a value less than 10^{-4} without seriously degrading the transmission within the band. In addition, the transmission characteristics of interference-type filters are extremely angle dependent and often can accept only a small field of view. Colored-glass filters by themselves are not capable of rejecting both visible and near-IR wavelengths and their transmission in the ultraviolet degrades as the glass ages. For these reasons the field of solar-blind photodetection is an active area of research and development. In particular, the development of photodetectors manufactured using wide-band-gap materials (direct and indirect) offers the potential for developing true solar-blind detectors that have negligible visible and near-IR response. In addition to superior spectral selectivity, wide-band-gap materials exhibit superior physical properties compared to Si, such as higher thermal conductivity, better radiation hardness and durability, and greater resistance to dielectric breakdown. These properties permit the use of wide-band-gap photodetectors in environmentally demanding applications that require resistance, for example, to high temperature, high voltage, or high incident power. The wide-band-gap materials that have received the greatest scrutiny for UV photodetector applications are silicon carbide (SiC), III-V nitrides, such as gallium nitride (GaN), and diamond. Silicon carbide and the III-V nitrides are discussed later. Diamond is potentially an ideal solar-blind detector because of its wide band gap (5.5 eV), low electrical conductivity, and chemical inertness. Natural diamond has been used previously, with significant restrictions and limitations, as a photoconductive detector for ultraviolet and ionizing radiation. More recently, the growth of polycrystalline diamond films by chemical vapor deposition methods has raised significant anticipation for the development of diamond-based UV detectors. However, high-quality undoped and *n*-doped diamond films have proven difficult to prepare. Structural defects in the polycrystalline films are responsible for defect states that can generate an unwanted subgap photoresponse. Diamond UV detectors will not be discussed in detail and the interested reader is referred to the literature (14–17) for the latest developments in diamond materials research and diamond UV detector device development.

Silicon Carbide. Hexagonal, crystalline silicon carbide (6H-SiC) is a wide-band-gap (2.9 eV at 300 K) semiconduc-

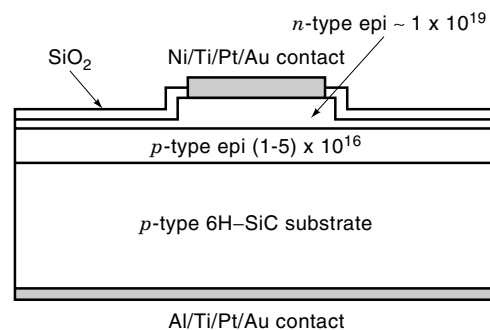


Figure 8. Schematic of a typical 6H-SiC photodiode. [Courtesy of Cree Research, Inc. Reprinted with permission from *Phys. Status Solidi A*, **162**, 481, 1997. Copyright 1997 Wiley-VCH Verlag.]

tor that absorbs UV light, but is essentially transparent to IR and visible light. The strength of the Si-C bond (bond energy ~ 5 eV) provides the material with superior resistance to high temperatures, chemicals, and mechanical shock. The width of the band gap results in low dark currents, even at elevated temperatures, thereby permitting sensitive photodetection under widely varying environmental conditions. While the device fabrication techniques for SiC materials are not as mature as those for Si fabrication, recent progress has yielded good-quality, bulk 6H-SiC single crystals that are used as substrates for the subsequent deposition of epitaxially grown *p*-type and *n*-type layers (18). Both *p-n* junction and Schottky barrier SiC photodiodes have been reported (19–23). High-quality *p-n* junction SiC photodiodes are commercially available as a chip (18,21) or packaged within a fixture (24,25). A schematic of the photodiode is shown in Fig. 8. A wafer of *p*-type 6H-SiC is used as the substrate and an epitaxial junction is fabricated by first growing an aluminum doped *p*-type layer followed by a nitrogen-doped *n*-type layer. The responsivity of the photodiode depends on the degree of doping and the thickness of these layers. A mesa is patterned and etched using reactive ion etching. The typical size of the mesa varies from 0.09 mm² to 9 mm². A 50-nm-thick silicon dioxide layer is used to passivate the surface. A cross-shaped Al contact covers less than 2% of the area of the mesa. A scanning electron microscope (SEM) micrograph of the SiC photodiode is shown in Fig. 9. Although the absorption length of UV light in SiC is greater than it is in Si, the responsivity of the SiC photodiode is improved by limiting the thickness of the *n*⁺ layer to ~ 0.1 μ m. The responsivity and quantum efficiency of the SiC photodiode are shown in Fig. 10 as a function of wavelength at several temperatures. The spectral response curves are smoothly varying functions of UV wavelength from 200 nm to 400 nm, exhibiting peaks at ~ 275 nm to ~ 300 nm, with corresponding maximum quantum efficiencies ranging from 80% to 100%. The response curves shift to lower energy with temperature due to the shift in the band gap with temperature from ~ 3.03 eV at 223 K to ~ 2.88 eV at 623 K. The dark current, as expected, is quite low in these photodiodes. Figure 11 shows the reverse-bias dark-current density versus voltage as a function of temperature for a SiC photodiode having an active area of 0.04 cm². At a typical operating bias of -1.0 V, the dark current at 473 K is $\sim 10^{-11}$ A/cm². Extrapolated to room temperature, this corresponds to an extremely low dark-current density of $\sim 2 \times 10^{-17}$ A/cm² at a reverse bias of -1.0 V. Comparing the performance of the SiC photodiode to a

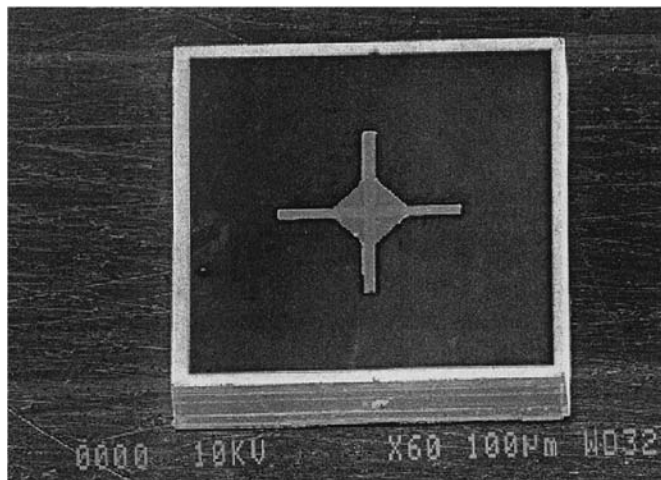


Figure 9. SEM micrograph of a typical 6H-SiC photodiode. [Courtesy of Cree Research, Inc. Reprinted with permission from *Phys. Status Solidi A*, **162**, 481, 1997. Copyright 1997 Wiley-VCH Verlag.]

UV-enhanced Si photodiode, we find the SiC possesses a number of advantages. First, the response of SiC peaks near 300 nm and falls dramatically after 400 nm, while that of Si peaks near 900 nm and extends to ~ 1200 nm. Therefore, for many applications, blocking filters are not required to eliminate unwanted visible and near-IR wavelengths in SiC while they are critical in Si. The responsivity of the SiC photodiodes is ~ 0.2 A/W to ~ 0.25 A/W near 300 nm, about the same or even greater than that of UV-enhanced Si photodiodes at the same wavelength. SiC photodiodes can operate with low noise at significantly higher temperatures than Si photodiodes and provide greater

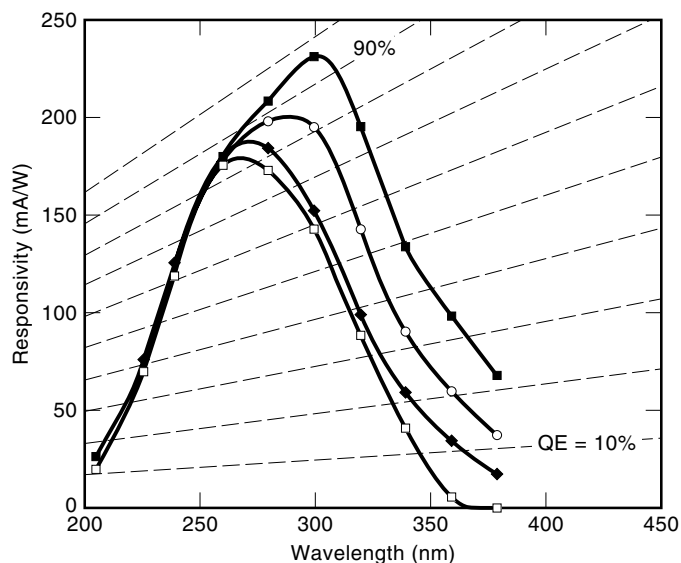


Figure 10. Spectral response of a typical 6H-SiC photodiode at several temperatures. Open squares: 223 K; diamonds: 300 K; circles: 498 K; and solid squares: 623 K. The peak quantum efficiency (QE) varies from 82% to 96% as a function of temperature. [Courtesy of Cree Research, Inc. Reprinted with permission from *Phys. Status Solidi A*, **162**, 481, 1997. Copyright 1997 Wiley-VCH Verlag.]

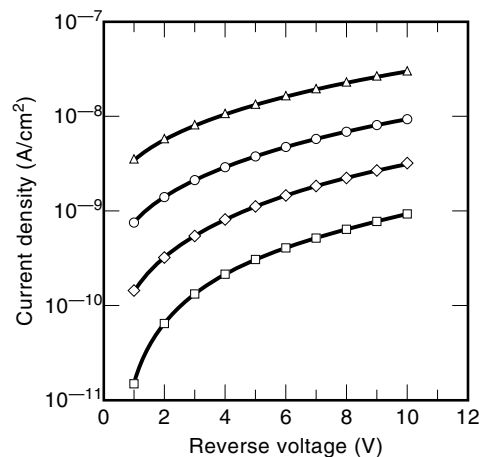


Figure 11. Dark current density of a typical 6H-SiC photodiode as a function of reverse-bias voltage at several temperatures. Squares: 473 K; diamonds: 523 K; circles: 573 K; and triangles: 623 K. [Courtesy of Cree Research, Inc. Reprinted with permission from *Phys. Status Solidi A*, **162**, 481, 1997. Copyright 1997 Wiley-VCH Verlag.]

long-term radiation resistance and stability than do Si photodiodes. (The ability to operate at high temperatures depends also on the limitations of the package that houses the semiconductor material and provides for electrical connections.) Recently, a thin-film photodetector utilizing amorphous SiC was reported (26). The photodetector was a $p-i-n$ structure made of hydrogenated amorphous Si and SiC in a sandwich configuration on a glass substrate. A schematic of the photodetector is shown in Fig. 12. Photogeneration of carriers due to UV light absorption occurred only in the α -SiC:H p layer, so the device exhibited good wavelength selectivity in rejecting visible and near-IR light. The primary advantage of this technology is the potential for fabricating low-cost, large-area arrays of photodiodes on a wide variety of substrates.

III-V Nitrides. Wide-band-gap materials that have been intensively studied in recent years are the III-V nitrides, including, in particular, gallium nitride (GaN), aluminum nitride

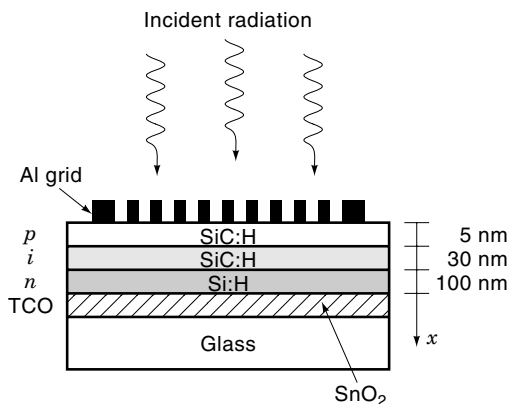


Figure 12. Schematic of thin film SiC photodetector fabricated using amorphous SiC and amorphous Si on a glass substrate. TCO = transparent conductive oxide film. [Reprinted with permission from *IEEE Trans. Electron Devices*, **43**, 1351, 1996. © 1996 IEEE.]

(AlN), and alloys of the two, $\text{Al}_x\text{Ga}_{1-x}\text{N}$ ($0 \leq x \leq 1$). GaN and AlN are direct-gap materials with band gaps of 3.4 eV and 6.2 eV, respectively, at 300 K. The band gaps of AlGaN alloys are also direct and can cover the range from 3.4 eV to 6.2 eV continuously, depending on the relative concentrations of Al and Ga. Therefore, optical devices developed using these materials will have cutoff wavelengths ranging from 365 nm to 200 nm. Devices of interest include light-emitting diodes and laser diodes in addition to UV photodetectors. The electronic properties of GaN provide for superior solar-blind wavelength selectivity, high dielectric breakdown thresholds, high thermal conductivity, and excellent mechanical hardness. These features offer the potential for operation of solar-blind GaN UV photodetectors without using long-wavelength cutoff filters. Also, these photodiodes may be operated under extreme environmental conditions such as high temperature and high humidity. The realization of such devices has been impeded because of the difficulty of fabricating high-quality intrinsic and *p*-doped materials. For this reason GaN manufacturing technology is still considered an art and is much less mature than that for Si or SiC. Since there is no suitable material that will lattice match it, GaN films are usually grown at high temperature on sapphire substrates. Metal-organic chemical vapor deposition and molecular beam epitaxy techniques can yield high-quality films on sapphire substrates. These have been used to demonstrate exceptionally good solar-blind UV photodetection. The first demonstration of UV photodetection using single-crystal GaN films was reported by Khan and coworkers (27). A photoconductive detector using interdigitated electrodes (due to the extremely high resistivity of GaN) exhibited excellent responsivity for wavelengths shorter than 365 nm and very little responsivity for wavelengths longer than 365 nm, with an extremely sharp cut-off due to the direct gap. This device had a very high photoconductive gain (responsivity > 1000 A/W), but had a long response time (order of ms). It has been found that both the responsivity and temporal response of photoconductive GaN detectors are functions of the incident power (28). This is the result of a competition between distinct recombination processes that involve deep traps within the GaN. More recently, a vertical geometry, transparent GaN Schottky barrier detector that uses front illumination has been described (29). A schematic of the detector is shown in Fig. 13. The vertical geometry of this detector resulted in efficient photocarrier collection and in a substantially improved response time (118 ns for a 50 Ω load). The spectral response of the Schottky photodiode for a -5 V bias is shown in Fig. 14. The response is relatively constant, ~ 0.15 A/W to

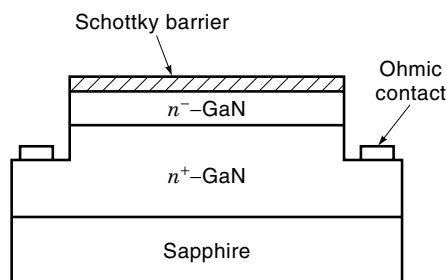


Figure 13. Schematic of vertical geometry, transparent GaN Schottky barrier photodiode. [Courtesy of APA Optics, Inc. Reprinted with permission from *Appl. Phys. Lett.*, **70**, 2277, 1997. Copyright 1997 American Institute of Physics.]

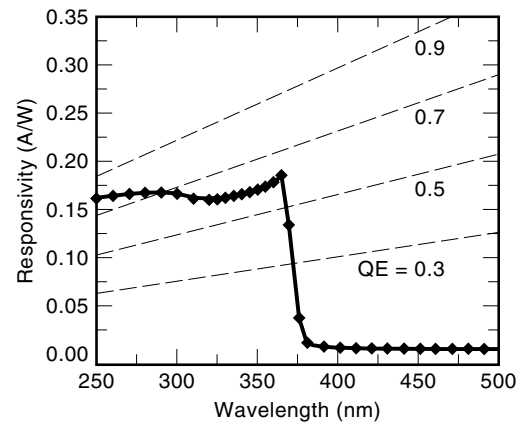


Figure 14. Spectral response of GaN Schottky barrier photodiode. [Courtesy of APA Optics, Inc. Reprinted with permission from *Appl. Phys. Lett.*, **70**, 2277, 1997. Copyright 1997 American Institute of Physics.]

0.18 A/W, from 250 nm to the band edge at 365 nm. Although GaN photodetectors are available commercially (30) the fabrication of GaN epitaxial films and the development of photovoltaic and photoconductive detectors is still a topic of intense current research (31–36).

PHOTOEMISSIVE ULTRAVIOLET PHOTODETECTORS

In a semiconductor photodetector ultraviolet light is absorbed within the semiconductor, producing electron-hole pairs that are then separated by an electric field, giving rise to the photoconductive and photovoltaic effects. In a photoemissive photodetector, however, ultraviolet light ejects photoelectrons from the surface of a photocathode, generating a photoelectron current when the photoelectrons are collected by an anode at a positive bias with respect to the cathode. The photoelectron current is proportional to the intensity of the incident ultraviolet light. The spectral response of the detector depends largely on the work function of the photocathode material. Vacuum photodiodes are relatively simple photoemissive detectors that can have subnanosecond time resolution. They are well suited for applications that require the measurement of relatively intense, fast UV light signals (37). A schematic of a windowless aluminum oxide photoemissive photodiode (38) is shown in Fig. 15. The photocathode is manufactured by evaporating a thin film of aluminum onto a quartz disk and oxidizing the surface to yield a thin layer (~ 15 nm) of aluminum oxide. The anode is a stainless-steel ring that is assembled with the photocathode in a Teflon body. The response of this photodetector is quite solar blind, exhibiting a peak at ~ 70 nm and falling to essentially zero beyond 150 nm. The response at the peak is 0.16 electrons/photon, reflecting the fact that while these photodetectors can have high quantum efficiency there is no internal gain.

Photomultiplier Tubes

The primary photoelectrons produced by incident photons at the photocathode can be multiplied to provide a photoemissive detector with high gain. Photomultiplier tubes (PMTs) are photoemissive detectors with very high gain that provide the most

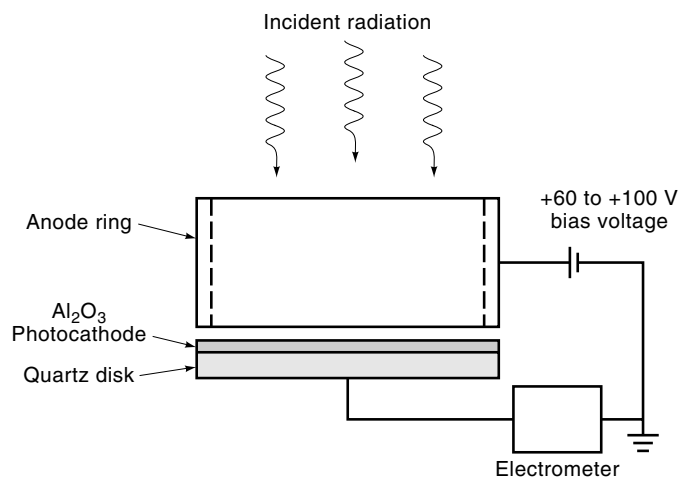


Figure 15. Schematic of windowless aluminum oxide photoemissive photodiode. [Courtesy of the Physics Laboratory, National Institute of Standards and Technology.]

sensitive method for measuring low light levels (39). The principle elements of a PMT are shown schematically in Fig. 16 and include a photocathode, an electron multiplier, and an anode, all of which are contained in a vacuum envelope. A photon impinges upon the photocathode of a PMT, causing a photoelectron to be ejected and accelerated toward the electron multiplier or dynode chain. The initial photoelectron impacts the surface of the first dynode, causing the emission of several secondary electrons, which in turn cascade through the dynode chain resulting in gains of 10^5 to 10^6 secondary electrons per photoelectron. The resulting electron current is detected at the anode. Depending on the rate at which the incident photons arrive, the output of the PMT may appear as short bursts of voltage pulses for low light levels or as a dc current for higher light levels. The most sensitive method for the measurement of low light levels ($N < 10^8$ photons/s) is single photon counting. This method counts the number of photoelectron-generated pulses that arrive at the anode during a preselected time interval. A voltage discriminator is used to discriminate between voltage pulses that originate from the photocathode (signal) and voltage pulses that originate from within the dynode chain (noise). For higher light levels, current measurement methods are used. There are a variety of types of electron multipliers that are used in PMTs including several types of dynode chains, wire mesh dynodes, and microchannel plates. Specific dynode structures are chosen for particular applications, for example, a linear focused dynode chain provides fast response times and

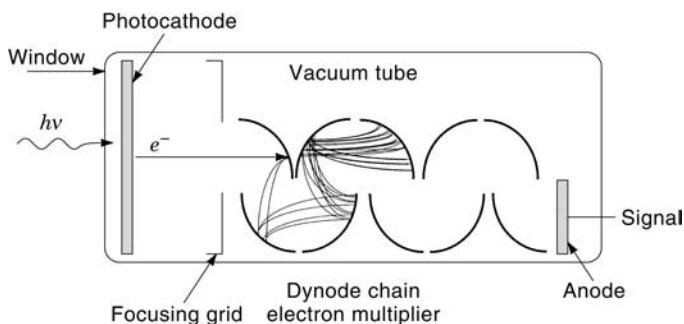


Figure 16. Schematic of a typical photomultiplier tube.

is particularly useful for pulsed measurements, while a box and grid type dynode chain provides excellent photon collection efficiency (40). Wire mesh and microchannel plate electron multiplier structures may be used for position-sensitive detection applications. These types of electron multipliers are particularly insensitive to external magnetic fields, and can be used in environments that are not suitable for most other PMTs. For the specific application of UV light detection, the wavelength response of a PMT is dependent on a number of factors including the window material used to seal the vacuum tube and the composition of the photocathode material. Windows composed of MgF_2 provide transmission down to 105 nm. Other UV-transmitting window materials include LiF, sapphire, and synthetic fused silica, with minimum transmission wavelengths of 115 nm, 145 nm, and 160 nm, respectively. Commercially available solar-blind, UV-sensitive PMTs utilize photocathode materials such as CsI, CsTe, KBr, and RbTe. The spectral response for a photomultiplier manufactured using a CsI photocathode and a magnesium fluoride window (13) extends from 115 nm to 195 nm, with the peak of the response occurring at ~ 120 nm. The spectral response for a photomultiplier manufactured using a CsTe photocathode and a synthetic silica window (13) extends from 160 nm to 320 nm, with the peak of the response occurring at ~ 200 nm. The short-wavelength cutoff for these photomultiplier tubes is determined by the absorption coefficient of the window material, while the long-wavelength limit is determined by the work function of the photocathode material. These UV-sensitive photocathodes are particularly useful for solar-blind detection applications as they are practically unresponsive to visible light wavelengths and can often be used without filters to block background light. Solar-blind PMTs are commercially available from a number of suppliers (12). Other photocathode materials are available with good UV sensitivity, such as bi-alkali, S-20, and S-11 photocathodes, but these are not visible-blind photocathodes, and appropriate wavelength discrimination using filters or monochromators is required to exclude visible light. Generally, although PMTs do not have the high quantum yield of semiconductor detectors, their large gain and low noise are necessary for very-low-light-level applications. Traditionally, the expense, mechanical fragility, susceptibility to magnetic fields, and high-voltage power requirements of PMTs have been important considerations in determining their use. Advances in PMT technology are challenging these traditional considerations as miniaturized PMTs, and PMT modules provide exceptional performance in small, rugged packages (13).

Gas-Filled Phototubes

The gas-filled phototube is another type of photoemissive photodetector that is used for solar-blind ultraviolet light detection applications. In these detectors a dc gas discharge is maintained by a small discharge current from a metal photocathode to a wire anode. When the gas discharge is illuminated with very low levels of ultraviolet light, gas breakdown occurs leading to an increase in the discharge current. The bias current is set near the breakdown condition and the response is nonlinear, exhibiting saturation at relatively low (nanowatt) power. Commercially available gas-filled phototubes (41) have a solar-blind spectral response (185 nm to 260 nm) and are used in flame detectors and fire alarms. Improved gas-filled phototubes have been described more recently (42) that operate without

gas breakdown and are significantly more sensitive than the gas-discharge tubes. These phototubes operate in the prebreakdown regime and detect light by an excited-state photoionization mechanism rather than glow discharge. The responsivity in the ultraviolet region is quite high, 6 A/W at 200 nm using molybdenum electrodes and argon gas, due to the internal gain. The response of these phototubes ranges from 190 nm to 270 nm.

BIBLIOGRAPHY

1. P. R. Norton, Photodetectors, in M. Bass (ed.), *Handbook of Optics, vol. 1, Fundamentals, Techniques and Design*, 2nd ed., New York: McGraw-Hill, 1995.
2. G. R. Carruthers, Ultraviolet and x-ray detectors, in R. W. Waynant and M. N. Ediger (eds.), *Electro-Optics Handbook*, New York: McGraw-Hill, 1994.
3. M. Razeghi and A. Rogalski, Semiconductor ultraviolet detectors, *J. Appl. Phys.*, **79**: 7433–7473, 1996.
4. D. E. Aspnes, Optical functions of intrinsic Si: Table of refractive index, extinction coefficient and absorption coefficient vs energy (0 to 400 eV), in *Properties of Silicon, EMIS Datareviews Series No. 4*, London: Institution of Electrical Engineers, 1988.
5. M. Krumrey et al., Schottky type photodiodes as detectors in the VUV and soft x-ray range, *Appl. Opt.*, **27**: 4336–4341, 1988.
6. N. M. Durant and N. P. Fox, Evaluation of solid-state detectors for ultraviolet radiometric applications, *Metrologia*, **32**: 505–508, 1995/96.
7. E. Tegeler and M. Krumrey, Stability of semiconductor photodiodes as VUV detectors, *Nucl. Instrum. Meth. Phys. Res. A*, **282**: 701–705, 1989.
8. R. Korde and J. Geist, Stable, high quantum efficiency, UV-enhanced silicon photodiodes by arsenic diffusion, *Solid State Electron.*, **30**: 89–92, 1987.
9. R. Korde and J. Geist, Quantum efficiency stability of silicon photodiodes, *Appl. Opt.*, **26**: 5284–5290, 1987.
10. R. Korde, J. S. Cable, and L. R. Canfield, One gigarayd passivating nitrided oxides for 100% internal quantum efficiency silicon photodiodes, *IEEE Trans. Nucl. Sci.*, **40**: 1655–1659, 1993.
11. Product literature, International Radiation Detector, Inc. 2527 W. 237th St, Unit B, Torrance, CA 90505.
12. See the *Laser Focus World Buyers Guide* or *Photonics Buyers Guide* for comprehensive lists of vendors.
13. Hamamatsu Corp., 360 Foothill Road, Bridgewater, NJ 08807.
14. R. D. McKeag et al., Photoconductive properties of thin film diamond, *Diamond Relat. Mater.*, **6**: 374–380, 1997.
15. S. Salvatori et al., Photoelectrical characteristics of diamond UV detectors: Dependence on device design and quality, *Diamond Relat. Mater.*, **6**: 361–366, 1997.
16. E. Pace et al., Electrooptical properties of diamond thin films as UV photodetectors, *Nucl. Instrum. Meth. Phys. Res. A*, **387**: 255–258, 1997.
17. G. Popovici et al., Diamond ultraviolet photovoltaic cell obtained by lithium and boron doping, *J. Appl. Phys.*, **81**: 2429–2431, 1997.
18. Cree Research Inc., 2810 Meridian Parkway, Durham, NC, 27713.
19. D. M. Brown et al., Silicon carbide UV photodiodes, *IEEE Trans. Electron. Devices*, **40**: 325–333, 1993.
20. J. A. Edmond, H.-S. Kong, and C. H. Carter, Jr., Blue LEDs, UV photodiodes and high-temperature rectifiers in 6H-SiC, *Physica B*, **185**: 453–460, 1993.
21. J. Edmond et al., 6H-Silicon carbide light emitting diodes and UV photodiodes, *Phys. Status Solidi A*, **162**: 481–491, 1997.
22. C. Frojdh et al., UV-sensitive photodetectors based on metal-semiconductor contacts on 6H-SiC, *Phys. Scripta*, **T54**: 169–171, 1994.
23. R. G. Verenchikova et al., *Sov. Phys. Semicond.*, **26**: 565–568, 1992.
24. Boston Electronics Corporation, 72 Kent St., Brookline, MA 02146.
25. Laser Components, Inc., 70 W. Barham Ave., Santa Rosa, CA 95407.
26. D. Caputo et al., Solar-blind UV photodetectors for large area applications, *IEEE Trans. Electron. Devices*, **43**: 1351–1355, 1996.
27. M. A. Khan et al., High-responsivity photoconductive ultraviolet sensors based on insulating single-crystal GaN epilayers, *Appl. Phys. Lett.*, **60**: 2917–2919, 1992.
28. F. Binet et al., Mechanisms of recombination in GaN photodetectors, *Appl. Phys. Lett.*, **69**: 1202–1204, 1996.
29. Q. Chen et al., Schottky barrier detectors on GaN for visible-blind ultraviolet detection, *Appl. Phys. Lett.*, **70**: 2277–2279, 1997.
30. APA Optics, Inc., 2950 NE 84th Ln., Minneapolis, MN 55449.
31. J. M. Van Hove et al., Ultraviolet-sensitive, visible-blind GaN photodiodes fabricated by molecular beam epitaxy, *Appl. Phys. Lett.*, **70**: 2282–2284, 1997.
32. R. D. Vispute et al., Growth of epitaxial GaN films by pulsed laser deposition, *Appl. Phys. Lett.*, **71**: 102–104, 1997.
33. J. C. Carrano et al., Very low dark current metal-semiconductor-metal ultraviolet photodetectors fabricated on single-crystal GaN epitaxial layers, *Appl. Phys. Lett.*, **70**: 1992–1994, 1997.
34. D. Walker et al., $\text{Al}_x\text{Ga}_{1-x}\text{N}$ ($0 \leq x \leq 1$) ultraviolet photodetectors grown on sapphire by metal-organic chemical-vapor deposition, *Appl. Phys. Lett.*, **70**: 949–951, 1997.
35. M. J. Malachowski and A. Rogalski, GaN ultraviolet photodiodes-photoresponse modelling, *J. Tech. Phys.*, **38**: 65–72, 1997.
36. M. Smith et al., Room temperature intrinsic optical transition in GaN epilayers: The band-to-band versus excitonic transitions, *Appl. Phys. Lett.*, **71**: 635–637, 1997.
37. C. V. S. Rao et al., Vacuum photodiode detectors for broadband vacuum ultraviolet detection in the Saha Institute of Nuclear Physics Tokamak, *Rev. Sci. Instrum.*, **68**: 1142–1148, 1997.
38. National Institute of Science and Technology, Physics Laboratory, Gaithersburg, MD.
39. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., New York: Wiley, 1989.
40. E. J. Lerner, Photomultiplier tubes offer high-end sensitivity, *Laser Focus World*, 87–96, July 1996.
41. UV TRON flame detectors, Hamamatsu Corp., 360 Foothill Road, Bridgewater, NJ 08807.
42. M. Cohen and N. S. Kopeika, A near UV envelope detector in the prebreakdown regime based on photoionization of excited gas atoms, *Meas. Sci. Technol.*, **5**: 540–547, 1994.

ALAN L. HUSTON
Université du Québec à Hull
BRIAN L. JUSTUS
Naval Research Laboratory

CHEMICAL SENSORS

Chemical sensors are devices that allow for the detection and measurement of individual chemical components, usually from a complex mixture. In practice, the sensor is always part of a larger system designed for the purpose of providing process or product information. This information might be needed for a control loop or quality assurance procedure. The mixture is typically a liquid or gas, but may also be a solid or semisolid

substrate. For example, the pH of aqueous solutions in the presence of dissolved salts can be monitored by an electrochemical device called a glass pH electrode or pH sensor. This chemical sensor responds specifically to hydrogen ions over a concentration range spanning approximately 10 or more decades. The pH reading can then be used to regulate the pH level by the addition of acid or base to the process stream. As another example, gas sensors are available that use optical properties to track oxygen concentration over a wide range of temperatures and pressures. In automobile engines, this might be employed to adjust the fuel–air mixture for better engine performance. Many other examples can be cited since there are potentially as many different types of chemical sensors as there are different molecules. However, a few common sensing strategies and operating principles can be formulated that tie them neatly together. It is logical, therefore, to focus on these strategies and use a few well-developed chemical sensors to illustrate the approaches.

IDEAL CHEMICAL SENSORS

It is convenient to define an ideal chemical sensor as one that operates in an instantaneous, reversible, and absolutely selective fashion without disturbing the composition or other properties of the sensing fluid. In the end, it provides an electrical signal that is precisely calibrated to the composition information. In practice, few of these characteristics are actually achieved; nevertheless, commercial chemical sensors are available with 1994 worldwide market values approximating \$500 million (1). A similar market exists for biosensors (see Biomedical sensors), which we distinguish here from the term *chemical* sensors simply by the method in which the sensing is carried out. If no biological components are used to carry out the sensing, it will be called a chemical sensor regardless of its intended use. The relative importance of the various attributes found in an ideal sensor are dependent on the application in which the sensor will function. Unfortunately, this also tends to cloud the picture of what is meant by a sensor. Two simple examples are given to better highlight the definition.

First, consider a simple device that monitors pH (a measure of the acidity) of a sample using a paper “test strip,” also called pH paper. The person carrying out the measurement would wet the paper by immersing it in a sample of fluid and then examine the color of the strip after a short time period, usually by placing it next to a calibrated color scale supplied with the container that held the test strip. The device gives the pH with a certain degree of accuracy, but it is not a chemical sensor since it lacks features that are important for its utility in a process or product monitoring environment. Its biggest failing is that the sensor provides no direct signal in the absence of the operator. It also has a single-use feature that is not reversible, and it would be more appropriately called a pH measurement device. Therefore, a better definition should include the point that chemical sensors are measuring devices that provide signal output in a continuous or nearly continuous fashion.

In a second example, consider a fluid where we extract a sample and pass it to a sophisticated gas chromatograph (GC) coupled with a mass spectrometer (MS). The gas chromatograph, for our illustrative purpose here, is capable of exquisite chemical separation, and the mass spectrometer is capable of equally

exquisite identification of the separated components. The resulting signal could be used to provide detailed composition information of the fluid and can be subsequently coupled to a control system to better regulate the process being monitored; and, in fact, many chemical production facilities operate with GC/MS devices, as they are popularly known. This is also not what we intend to call a chemical sensor, however, since it is quite simply too “big.” More precisely, the need for sampling the fluid stream and then characterizing it using bench-top analytical instrumentation is usually an undesirable step for process control. The lag times introduced between sampling and subsequent control action, the cost of the equipment, and the inability to position an arbitrary number of such devices at or near the process line severely limit the utility of such otherwise sophisticated approaches. In fact, such steps are often considered for replacement whenever a more practical device, like a chemical sensor, is introduced. Therefore, we amend our definition again by saying that the sensor should provide rapid output of the signal to the process in a cost-effective manner.

A somewhat cumbersome, but inclusive, working definition can now be stated: a chemical sensor is a device that provides quantitative information on the composition of a fluid mixture in a reversible, minimally invasive, selective, rapid, continuous, and cost-effective manner.

In examining different examples of chemical sensors, we will see that many of the designs are refinements and adaptations of strategies that have been known to work effectively in the analytical laboratory. For example, a pH sensor could be developed by placing pH indicators, originally present in the pH paper just considered, on the end of an optical fiber integrated with a compact light source and detector arrangement to overcome the previously cited limitations. Additionally, gas separation steps employing membranes instead of chromatographic operations could be used together with selective electrochemical electrodes to eliminate the unfavorable attributes of the gas chromatograph setup. Nowadays, even miniaturized GC/MS devices operating on a millimeter-scale device (like an analytical lab on a chip) are being envisioned to function as future chemical sensors. The many advantages of microfabrication techniques are being increasingly exploited to produce sophisticated, reproducible, and inexpensive sensing systems. Some of these sensors have no equivalence to any laboratory protocol, and they function as truly unique devices.

GENERAL FEATURES OF CHEMICAL SENSORS

Three general features can be identified that make up a chemical sensor. These include a selective surface, a transducer, and a processor as depicted in Fig. 1, along with some examples of different materials and techniques that can be used for each component. The selective surface allows for a specific interaction with the target analyte, where analyte refers to the compound whose composition or presence is being sought. The surface can be modified by addition of ligands or catalysts that provide specific recognition features. The transducer tracks the association of the surface with the analyte. For example, the transducer may be a polymeric material that changes its electrical resistance with the incorporation of the analyte. Alternatively, the transducer may change color or modify the transmittance of light passing through it. Some signal detection, processing, and amplification are subsequently carried out to generate a

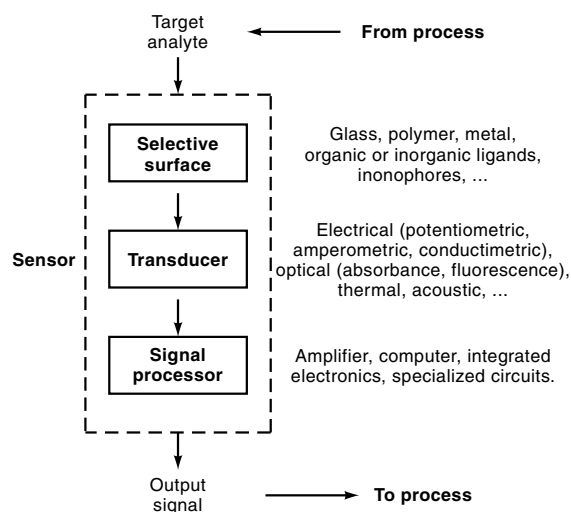


Figure 1. The basic features of a chemical sensor and its interaction with a process system. Representative examples of the three basic components found in any sensor are listed. The target analyte refers to the particular chemical species whose composition information is being sought.

practical output signal used by the process controller. The final signal coming from the sensor is invariably electronic so that it can be conveniently interfaced to standard controllers or data loggers.

The chief limitations of many chemical sensors are associated with the selectivity of the response to the target analyte compared with the response seen for potential background or contaminating compounds. In a constant background there is little need for selectivity, but in a more typical sensing situation the background concentration changes in unknown ways due to the presence of similar chemical components or "active" species as well as temperature and pressure fluctuations. A typical interferent in many gas-sensing situations is the background moisture or humidity level that is constantly undergoing variations.

The main types of chemical sensors developed to date rely primarily on electrical or optical transduction technologies. To a lesser degree, physical transduction processes such as displacement tracking, due primarily to pressure- or temperature-mediated effects, may also be used, but it is often the case that the operation of such devices rely, again, on electrical or optical signal modulation. Within the electrical transducers, it is convenient to divide them into amperometric, potentiometric, or conductimetric sensors depending on whether the measured

signal is the current, voltage, or conductance, respectively. Optical sensors include those that monitor the light absorption, refractive index, polarization, or fluorescence of a transducer, perhaps at many different wavelengths and using continuous or variable light input schemes. The class of electrochemical sensors dominates the worldwide chemical sensors market and, for example, accounted for 88% of the sales share in Japan for all chemical sensors in 1989 (2).

SENSOR TYPES

Electrochemical sensors include the familiar wired electrodes as well as metal oxide semiconducting devices, fuel cells, ion-selective field effect transistors (ISFETs), interdigitated electrodes, high-temperature solid-electrolyte systems, and chemiresistors. Optical sensors include so-called optodes and surface plasmon resonance systems that operate as fiber or planar structured devices. A current trend in many chemical sensors is toward miniaturization and microfabrication that can impart novel properties to sensors, thereby affecting their selectivity, sensitivity, and stability. This introduction does not explore all the latest research variations and innovations but focuses on devices that are commercial or near commercial. Recent research results can be found in specialized journals devoted to chemical sensing and analytical chemistry.

Electrochemical Sensors

Table 1 lists the three main types of electrochemical sensors. They are explored in separate sections below, with specific examples that illustrate the important features. Not all sensors can be exactly categorized in this way since they represent only limiting behaviors. Nevertheless, such distinctions are routinely used and widely adopted in the chemical sensor literature.

Amperometric Sensing. An amperometric chemical sensor is composed, at a minimum, of two metal electrodes maintained at a constant potential (voltage) difference. The current flowing in a closed circuit (which is established and monitored by suitable electronics) through the two electrodes is measured and indicates the presence of "electroactive" species. The relationship between the moles, N , of electroactive species reacted (either oxidized or reduced) and the charge passed through the sensor, Q , is known as Faraday's law:

$$N = Q/nF \quad (1)$$

where n is the number of electrons transferred per mole and

Table 1. Electrochemical Sensor Transduction Mechanisms

Transducer	Mechanism	Signal	Example Sensors
Amperometric	Electron charge transfer reaction (Faraday's law)	Current; due to production or consumption of electroactive species	Oxygen (Clark electrode), some hydrocarbons, hydrogen peroxide, fuel cells, high-temperature limiting-current sensors
Potentiometric	Multiphase equilibrium (Nernst equation)	Potential; due to distribution of ionic species across phases	Glass pH electrode, ion selective electrodes, gas sensors (Severinghaus CO ₂ electrode), some high-temperature solid-state sensors
Conductimetric	Variation of resistance with composition	Conductance or resistance; due to changes in resistive elements with absorption-adsorption events	Metal-oxide-semiconductor gas and humidity sensors, tin oxide gas sensors, conducting polymer humidity and gas sensors, broad specificity sensors

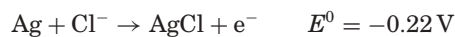
F is the Faraday constant ($F = 96487 \text{ C/mol}$). By monitoring the charge over time, which is the current, the amount of electroactive compounds can be deduced. To operate as a true sensor, however, some selectivity is required, and this can be achieved by the choice of a suitable electrical potential, by the introduction of a catalytic reaction step, and/or through permselective membranes. For example, a common method for the detection of dissolved oxygen in liquids is through the use of a so-called polarographic or Clark electrode. In this chemical sensor, the reduction of oxygen is followed at constant potential by monitoring the current passing between a working Pt (platinum) electrode and a Ag/AgCl (silver/silver chloride) reference electrode. At steady state, the current is proportional to the dissolved oxygen concentration. This is often reported as an equivalent partial pressure corresponding to the oxygen concentration in a gas phase that would be in equilibrium with the actual liquid-phase concentration.

Polarographic Sensors. In order to see how this chemical sensor functions and to draw analogies with many other equivalent systems, we examine the individual components of the device in greater detail. Figure 2 shows the salient features of a Clark electrode. First, in order to achieve some specificity to oxygen, a catalytic reaction is carried out at the cathode. The cathode is typically made of platinum, coated onto a wire or some other underlying support, that catalyzes the following electrochemical reaction,



where E^0 is the standard electrode potential at 25°C for the reaction as written measured against the standard hydrogen electrode. The hydrogen electrode is further described in the next section on potentiometric sensing. The actual reference electrode often used in a Clark oxygen sensor, however, is not the hydrogen electrode, but rather a Ag/AgCl electrode at which

the matching reaction occurs:

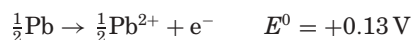


The net effect is the reduction of oxygen accompanied by the transfer of electrons at a standard electrode potential difference of about 600 mV. Second, the cathode is separated from the sensing fluid by the presence of an oxygen permeable membrane, usually made of a thin silicone polymer. This allows for oxygen transfer, but it eliminates many other electroactive polar and nonlipophilic compounds that would interfere if they were to be reduced at the cathode surface. The membrane also protects the sensor from fouling due to nonreactive adsorption that can severely degrade its performance. Finally, the anode is maintained at fixed reference conditions by establishing a constant chloride ion concentration. In practice, this is achieved by immersing the reference anode in a fixed KCl salt solution that is separated from the sampling fluid. The actual potential created between the reference and platinum electrodes confers further selectivity to the sensor in that it favors the reduction of oxygen.

The cathode in this example is often called the working electrode and has a surface area much smaller than the anode, or reference (also called the auxiliary) electrode. This ensures that the working electrode limits the current flow in the circuit and, hence, that the current measurements will only reflect reaction steps at the cathode (i.e., reactions with oxygen). The actual working potential is often greater than 600 mV to drive the oxygen reduction as much as possible. If the voltage is too high, however, undesirable electrolysis and corrosion steps can also take place that interfere with the measurement and limit the life of the sensor. A usual choice for the polarographic oxygen sensor (Clark electrode) is around 600 mV to 900 mV, where the working Pt electrode is made negative with respect to the auxiliary electrode. The two electrodes are often housed in an integrated package that provides a compact arrangement of all the necessary parts and allows the electric circuit to be completed on the sensing side of the permeable membrane.

Variations on this strategy for oxygen sensing have been developed that illustrate some alternative detection and sensing schemes. For instance, if a third electrode is introduced, called a "counter electrode," the current measurement can be carried out through a low-impedance circuit path that avoids driving large currents through the high impedance reference electrode (3). This has the advantage of maintaining long-term sensor calibration as well as prolonging the life of the sensor by avoiding corrosion reactions. The typical electric configurations for working with two- and three-electrode designs are shown in Fig. 3.

Galvanic Sensors. Alternatively, the reference electrode can be replaced by a base metal such as lead or zinc. The reaction at the anode (in the case of a lead auxiliary electrode) is now summarized as



This oxidation reaction obviously consumes the electrode over time, but nevertheless maintains a fixed potential that is sufficient to drive the reduction of oxygen. Usually, a silver cathode is used in combination with the base metal anode. Some simplification is introduced since no external potential is required and a direct measurement of the current is proportional to the

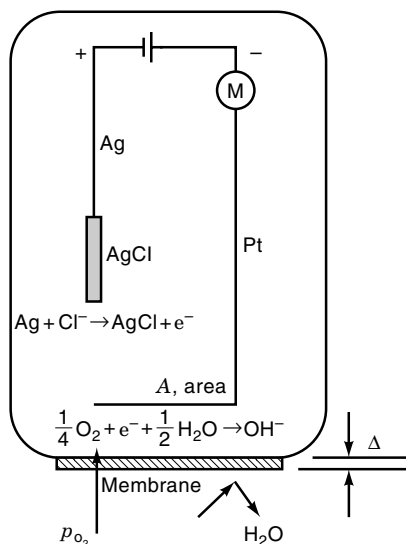


Figure 2. Features of an amperometric sensor for detecting oxygen (Clark electrode). The reactions occurring at the two electrodes are separated from the bulk fluid by an oxygen-permeable, nonporous membrane.

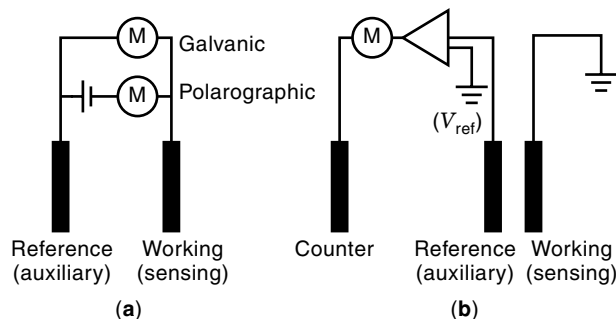
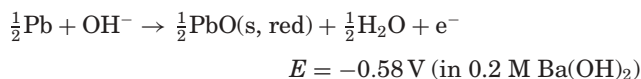


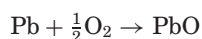
Figure 3. Electrochemical sensor configuration for amperometric sensors in two-electrode (a) and three-electrode (b) designs.

oxygen consumed. Such a design is referred to as a galvanic oxygen sensor, and a two-electrode configuration is shown in Fig. 3. In general, polarographic chemical sensors impose a fixed potential between the reference and sensing electrodes while galvanic chemical sensors do not. In both cases, however, the current is measured, and they are collectively referred to as amperometric sensors.

Fuel Cell Sensors. Finally, a “fuel cell” oxygen sensor can be constructed that contains the lead and working electrodes immersed in a basic electrolyte. Here, the anode generates a potential due to the following reaction:



In combination with the cathodic reaction, a net overall stoichiometry can be written as



which does not involve any of the electrolyte species. This is one reason it is often called a fuel cell sensor even though the generation of electricity is used only to obtain an appropriate signal. Again, a direct measurement of the current can be calibrated with the oxygen concentration.

Diffusion-Limited Current. Referring back to the polarographic oxygen electrode, it is instructive to consider the limiting case where the cathodic reaction is sufficiently fast that the oxygen concentration at the cathode surface goes to zero. As depicted in Fig. 2, the membrane will now generally limit the transport of oxygen from the measurement fluid to the electrode surface. Sensors that function under such conditions exhibit the maximum possible signal, all other things being equal, and are said to operate under diffusion-limited conditions. Such an operating protocol also allows for less stringent control over the applied potential. In fact, the potential can be allowed to fluctuate over small ranges without any dramatic effects on the sensor response since the applied potential is already sufficient to reduce all the oxygen at the electrode surface.

The mass transfer through the membrane satisfies Fick’s second law:

$$\frac{\partial p}{\partial t} = D_{\text{O}_2} \frac{\partial^2 p}{\partial x^2} \quad (2)$$

where p is the (equivalent) oxygen partial pressure in the membrane and D_{O_2} is the oxygen diffusion coefficient in the membrane. If we consider a step change in the bulk oxygen concentration from 0 to a value p_{O_2} , the measured current at the cathode surface can be found by integrating the above equation. The solution is

$$i(t) = \frac{4FAP_m}{\Delta} p_{\text{O}_2} \left[1 + 2 \sum_{n=1}^{\infty} (-1)^n e^{-n^2 \pi^2 D_{\text{O}_2} t / \Delta^2} \right] \quad (3)$$

The factor 4 in Eq. (3) reflects the transfer of four electrons per mole of oxygen reduced and A is the cathode surface area, as shown in Fig. 2. The other parameters include the membrane permeability, $P_m = D_{\text{O}_2} S_m$, where S_m is the solubility of oxygen in the membrane. The membrane thickness is Δ and the oxygen diffusivity is D_{O_2} . At sufficiently long times, the current is seen to be directly proportional to the oxygen partial pressure. This is the reading that would be measured at steady state, and the linearity of the steady-state current response with the oxygen partial pressure is a direct consequence of the diffusion-limited nature of the sensor. Furthermore, the time constant for the sensor can be characterized by the dominant exponential term and is conveniently represented as

$$\tau = \frac{\Delta^2}{\pi^2 D_{\text{O}_2}} \quad (4)$$

For a 0.1 mm thick membrane with an oxygen diffusivity of $10^{-6} \text{ cm}^2/\text{s}$, the time constant will be approximately 10 s. Other factors such as poor mixing in the electrolyte solution as well as slow mixing in the external sensing fluid will cause the actual sensor response to be longer. If it is necessary to track oxygen concentration on a much shorter time scale, thinner and more permeable membranes would be needed, or mixing in the external as well as internal regions should be introduced.

Selectivity. Other species besides oxygen can also be detected by analogous procedures. In some cases, a preliminary reaction step is introduced that converts the target compound into oxygen or consumes oxygen at a rate proportional to the concentration of the intended analyte. This allows the sensor just described to be used directly. Many biosensors (see Biomedical sensors) have been developed with this strategy in mind where an enzymatic reaction generates or consumes oxygen. In other cases, appropriate selection of the electrode potential can provide some selectivity. For instance, an amperometric sensor for ethylene (C_2H_4) has been reported (4) that oxidizes ethylene on a gold working electrode (anode). In this case, the gold was deposited on a membrane in order to generate high surface areas, rather than using a membrane to cover a typical wire electrode. The electrochemical sensor configuration was operated with an $\text{Hg}/\text{Hg}_2\text{SO}_4$ reference electrode (cathode) as well as a gold counterelectrode, and a linear current was observed for gas-phase ethylene concentrations between 0 ppm and 500 ppm (1 ppm = 1 “part per million” = 1 molecule or mole per million molecules or moles of gas). However, this sensor exhibits some cross-sensitivity to other electroactive species as shown in Table 2. This indicates that the selectivity of the sensor is such that other compounds can simultaneously contribute to the amperometric response. In this case, the response is evaluated at 10 ppm (or higher) for a variety of gases and is normalized to the response seen for 10 ppm ethylene. Such a ratio is a convenient

Table 2. Selectivity Profile of Ethylene Sensor

Gas Measured (concentration is 10 ppm, unless indicated)	Relative Response (ethylene = 1; 10 ppm)
Acetaldehyde (100 ppm)	0
Acetylene	7.4
CO (100 ppm)	0
Ethanol (100 ppm)	0
NO	1.8
NO ₂	0.3
SO ₂	2.5

measure of the selectivity. The sensor was found to be more selective for some compounds when the response was compared with that seen for ethylene. This implies that it would, in fact, be a better sensor if the target compound had been chosen to be, in this case, acetylene or sulfur dioxide. The consequence of this on the usefulness of the sensor for ethylene detection, however, depends on the intended application. In general, selectivity considerations must always be considered from the point of view of known interferents and expected interferent levels. These factors are prone to change when the application environment changes, and sensors that function well in one situation may turn out to be completely unacceptable in other applications. If it is not likely that known interferents will be present in the sampling stream, the response to those components may be of little consequence.

Potentiometric Sensing. A potentiometric chemical sensor is composed of many of the same parts as an amperometric sensor, except that voltage is measured under conditions of a small, often negligible, current. The most important feature in a potentiometric measurement is that equilibrium conditions are established between electroactive species in solution and at the electrode. This is in contrast to the diffusion-limited condition just discussed for amperometric measurements and leads to relations between voltage and concentration that are logarithmic rather than linear.

Nernstian (Equilibrium) Behavior. The general way to describe the performance of an ideal potentiometric chemical sensor is to refer to it as Nernstian. The derivation of the Nernst equation can be found in many introductory chemistry texts, so we will only outline the important features here for a particular system (see Electrochemical Electrodes). In particular, consider the silver-(porous) silver chloride (Ag/AgCl) electrochemical reaction described previously, but now as a reversible “half-cell” reaction:



The equality of electrochemical potentials for all reactants and products defines the equilibrium state of the system. Under ideal conditions, the electrochemical potential, μ_A^* , of a particular species in solution at a concentration $[A]$ is found from the thermodynamic relation (5)

$$\mu_A^* = \mu_A^0 + RT \ln[A] + zF\phi_A \quad (5)$$

where the brackets denote concentrations, R is the gas constant, T is the absolute temperature, z is the electric charge associated with the ion A , and ϕ_A is the electrode potential.

For any phase consisting of only a pure component, such as the metallic silver wire or the porous silver chloride coating, the electrochemical potential will be a constant at a specified temperature. Thereby, the Ag/AgCl equilibria (referred to also as a “couple”) can be expressed in terms of the chemical and electrical potentials as

$$\mu_{\text{Ag}}^0 + (\mu_{\text{Cl}^-}^0 + RT \ln[\text{Cl}^-] - F\phi_{\text{Cl}^-}) = \mu_{\text{AgCl}}^0 + (\mu_{\text{e}^-} - F\phi_{\text{e}^-}) \quad (6)$$

If we rearrange this equation, collect all terms that are at most a function of the temperature, and introduce $\Delta\phi = \phi_{\text{Cl}^-} - \phi_{\text{e}^-} = \phi_{\text{S}} - \phi_{\text{M}}$ to emphasize that the potential is determined as the difference between the solution and the metal, we obtain

$$\Delta\phi = E^0 - \frac{RT}{F} \ln[\text{Cl}^-] \quad (7)$$

The standard electrode potential is tabulated in many reference books (6) and was used previously in discussing amperometric reactions. It is seen that the potential at a fixed temperature will be a function of only the chloride ion concentration. In a saturated solution containing salt crystals, this is easily maintained at a fixed value and explains why this type of electrode is often used as a reference. The generalization of this expression to any number of soluble species participating in a half-cell reaction can be done using the same sequence of steps. The reaction, of course, must be properly balanced with respect to the stoichiometry as well as the charge.

If an experiment is set up to actually measure the potential predicted by the above equation, it will always be necessary to introduce a second wire (electrode) into the solution to complete the measurement circuit. This second wire is problematic since it can now establish its own potential difference in the solution under investigation and it will generally be impossible to isolate the individual contributions. For this reason, the H^+/H_2 couple has been arbitrarily chosen as a particular half-cell reaction to have a potential of zero (at 25°C). As a result, all measurements comprising a closed circuit with this electrode, called the standard hydrogen electrode (SHE), can be assigned unique values of the standard electrode potential.

For any two electrodes, the net resulting electrical potential will be found from the differences in their half-cell potentials. A cell made of any two half-cells that is balanced with respect to the production and utilization of electrons is called a “formal” cell, and the net reaction is called the formal cell reaction. The production of PbO from metallic lead and oxygen in a fuel cell oxygen sensor is a formal cell reaction, although it is not at equilibrium. It is convenient to summarize the components as well as the physical arrangement of a formal cell by using the following type of cell diagram:



This indicates that the potential difference, for this example, between a mercury electrode (commonly called a calomel electrode) and the silver electrode will be measured. The single vertical bars indicate a different solid or liquid phase that participates in the equilibria. In this case, both electrodes share the same aqueous phase, but this is not usually possible when sensing is to be carried out—since we would already need to know the composition in order to predict the reference potential. For liquids, porous barriers are often introduced that allow

for ion migration, but otherwise limit the bulk flow of aqueous solutions.

From the equation developed previously, we can write the electrode potentials for each half-cell and simply sum them. This eliminates the solution potential from the equations and leads to the Nernst equation, which in general will be expressed as

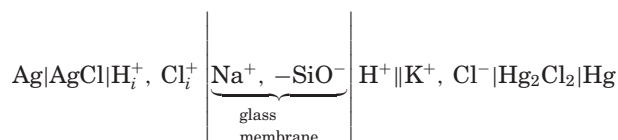
$$\Delta\phi = \Delta E^0 + \frac{RT}{F} \ln \frac{\prod a_{R_i}^{v_i}}{\prod a_{P_j}^{v_j}} \quad (8)$$

Here we have also introduced the activities in place of the concentrations to make the result applicable to nonideal solutions as well. The exponents v_i and v_j are the stoichiometric coefficients (always positive numbers) for those reactants and products, respectively, that participate as solution- or gas-phase species. The activity, a_i , is related to the concentration by the so-called activity coefficient,

$$a_i = \gamma_i [i] \quad (9)$$

Correlations for the activity coefficient are expressed in terms of the Debye–Hückel limiting law or its extension (7). Components that participate in the reaction as pure phases (like solids) or in great excess (like water) do not explicitly show up on the right hand side of Eq. (8). Experimentally, the measurement of electric potential together with the Nernst equation is often used to determine the activity coefficients of salt solutions.

Glass Electrode and the pH Sensor. By clever choice of electrode reactions and salt solutions, it is possible to construct chemical sensors for various ions that operate on the principles just described. However, poor sensor selectivity often limits such devices to a few specialized situations where the background interferences are fixed. Membranes, however, can offer a means to impart selectivity to potentiometric sensors in much the same way they were used for constructing the Clark electrode. One of the most widely used potentiometric sensors is the glass electrode, or pH sensor, which incorporates a glass membrane to limit the response to being dependent only on the hydrogen ion concentration or, more exactly, on the hydrogen ion activity. Figure 4 shows the construction of a pH sensor. With a suitable reference electrode, say a calomel electrode, the components of the cell can be summarized as



In fact, both the silver/silver chloride and the calomel electrodes can be considered as reference electrodes, and the potential-determining reactions are all confined to the glass membrane. The equilibration of hydrogen ions on the external side (surface) of the glass membrane and the external solution (this is the solution whose pH is sought) is due to an ion exchange with sodium ions in the special sodium/calcium-silicate glasses used. The anionic binding sites are fixed, and at high concentration

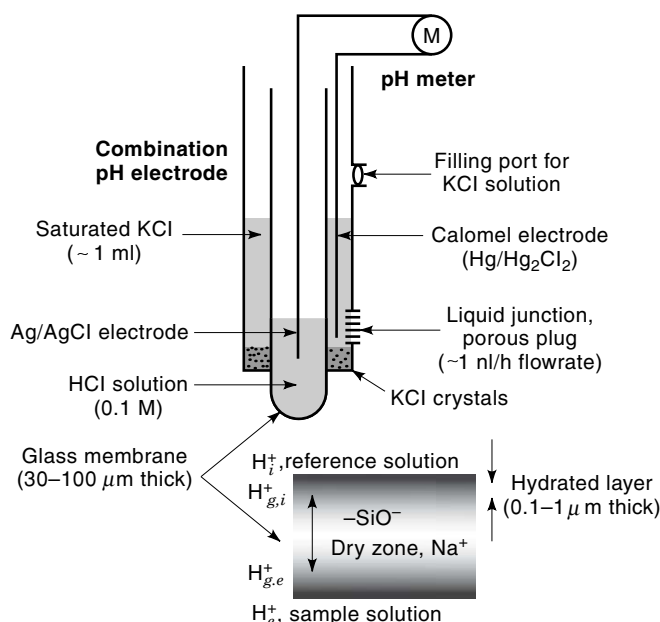


Figure 4. The glass electrode (pH sensor). A combination electrode is shown. The glass membrane region is expanded at the bottom of the figure to highlight the notation used in the text.

of sodium ions this leads to the relation

$$\phi_e - \phi_{g,e} = -\frac{RT}{F} \ln \frac{H_e^+}{H_{g,e}^+} \quad (10)$$

where the notation $H_e^+ = a_{H^+}$ is used to indicate the (external) solution activity. The subscript g denotes activities evaluated at the glass surface. At the inside surface of the glass membrane, an exactly analogous relation holds,

$$\phi_i - \phi_{g,i} = -\frac{RT}{F} \ln \frac{H_i^+}{H_{g,i}^+} \quad (11)$$

and, by adding the equations and rearranging, one obtains

$$-\frac{RT}{F} \ln \frac{H_e^+}{H_i^+} = \left[\frac{RT}{F} \ln \frac{H_{g,e}^+}{H_{g,i}^+} + (\phi_{g,e} - \phi_{g,i}) \right] + (\phi_e - \phi_i) \quad (12)$$

The term in square brackets reflects equilibria occurring within the glass membrane but confined to a small region on each surface. It has been estimated that the exchange surface consists of a hydrated layer on the order of 0.1 μm to 1 μm thick, whereas the overall thickness of the glass is typically about 100 μm . Only sodium ions transport charge through the dry glass membrane (no hydrogen ions are transferred through the glass), and they also participate in the surface equilibria. Although not readily apparent, this leads to conditions such that the term in brackets is essentially constant.

The two reference electrodes monitor the internal and external potentials. The last step in developing a sensor is to maintain a constant internal hydrogen ion concentration. This is easily done by filling the inside of the glass bulb shown in Fig. 4 with a solution of, say, 0.1 M HCl since neither hydrogen ions nor chloride ions permeate the glass. Collecting all the

constant terms as ϕ' and using activities, we obtain

$$(\phi_e - \phi_i) + \phi' = -\frac{RT}{F} \ln a_{H^+} = \frac{2.303 RT}{F} \text{pH} \quad (13)$$

The standard definition of the pH has been used:

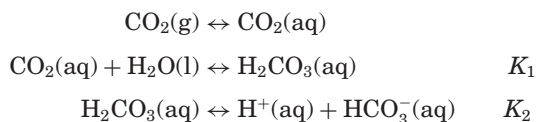
$$\text{pH} \equiv -\log_{10} a_{H^+} \quad (14)$$

At 25°C, the equation above shows that for every unit change in pH the recorded potential should change by 59 mV ($R = 8.313 \text{ J/K mol}$, $T = 293 \text{ K}$). The selectivity, measured at equal concentrations, of the ion-exchange reaction is always greatly in favor of the hydrogen ion over other competing cations. At low pH and high sodium concentrations, however, the nature of the cation exchange mechanism shows that it will also be a function of the external sodium ion concentration. This usually limits pH sensors to values less than pH = 11. Special "high-pH" glass made of Li_2O (in place of $\text{Na}_2\text{O}/\text{SiO}_2$) can extend the range a few units higher, however.

In practice, the pH sensor is often designed as a "combination electrode" in which the calomel electrode is placed around the internal Ag/AgCl electrode as shown in Fig. 4. The calomel electrode must be exposed to a constant (usually saturated) chloride ion concentration, but it must also have continuous ionic contact with the analyte solution in order to allow for a complete electric circuit. A small porous plug or capillary channel, referred to as a liquid junction and indicated in a cell diagram by a vertical double bar, is used with a very small leakage flow rate to accomplish this. Furthermore, by choosing KCl as the salt, the anion and cation molar conductivities are such that very little additional potential is generated at the liquid junction.

A commercial pH meter includes the glass electrode, a high-impedance voltmeter, and some form of temperature compensation and calibration functions. Temperature compensation is usually built-in according to the predictions of the Nernst equation, and calibration protocols with buffers of known pH are established to account for asymmetry potentials in the glass membrane, reference electrode drift, and nonzero liquid junction potentials. The asymmetry potential arises from surface defects and strain due to production or aging of the glass bulb as well as surface adsorption phenomena. With proper care and operation, a pH meter is almost unmatched among all chemical sensors in its sensitivity, selectivity, dynamic range, and ease of operation.

Gas Sensors. The extension of this sensor to other chemical species or ions can be accomplished in two ways. On the one hand, a reversible reaction can be introduced that selectively generates or consumes hydrogen ions in proportion to the target analyte. On the other hand, membranes with selective permeability to other ions can be used. For the former situation, gas-sensing electrodes have been fabricated that rely on establishing an acid-base equilibrium in a solution exposed to the gas and also contacting a pH-sensing electrode. For example, if CO_2 is allowed to equilibrate in an aqueous solution, the following equilibria will be established:



This leads to the relation

$$p\text{CO}_2 = K a_{H^+} a_{\text{HCO}_3^-} \approx \kappa a_{H^+} \quad (15)$$

where the second equality depends on a constant concentration (activity) of HCO_3^- ion. This is readily achieved by using a 10 mM solution of sodium bicarbonate (NaHCO_3) in the aqueous phase. In practice, the bicarbonate ion is often integrated into the reference electrode that further communicates with the external solution through a gas-permeable membrane. It is also possible to wrap the membrane around the bulb of regular combination pH electrode so that the space between the membrane and the electrode sequesters a carbonate buffer solution. A combination pH electrode is often used in either of these situations since the entire sensing circuit can be placed behind the membrane. Levels as low as 10^{-5} M CO_2 can be detected. One of the most popular uses of this type of electrode is in biomedical applications to monitor arterial blood gases, and in such situations it is often referred to as a Severinghaus electrode.

The gas-sensing principle can be readily adapted to other gases such as ammonia (NH_3). In this case, a 0.1 M solution of ammonium chloride (NH_4Cl) is used in place of the bicarbonate salt. Through a series of equilibrium reactions, the ammonia gas solubility is directly related to the solution pH, and levels of NH_3 down to $1 \mu\text{M}$ can be detected. Table 3 lists some additional gas-sensing strategies for other species that can be coupled to the pH electrode. Other ion electrodes, mentioned below, have

Table 3. Gas Sensors Based on pH Electrodes

Gas Analyte	Equilibria, ($\text{p}K_a$; 25°C)	Lower Limit (μM)	Reference Solution, Sample Condition
Ammonia, NH_3	$\text{NH}_4^+ \leftrightarrow \text{NH}_3 + \text{H}^+$ (9.40)	1	0.01 M NH_4Cl , pH > 11
Acetate, HOAc (Ac = $\text{CH}_3\text{CO}-$)	$\text{HOAc} \leftrightarrow \text{H}^+ + \text{OAc}^-$ (4.76)	1000	0.1 M NaOAc, pH < 2
Carbon dioxide, CO_2	$\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{H}^+ + \text{HCO}_3^-$ (6.38; using $(\text{CO}_2 + \text{H}_2\text{CO}_3)$ in denominator)	10	0.01 M NaHCO_3 , pH < 4
Chlorine, Cl_2	$\text{Cl}_2 + \text{H}_2\text{O} \leftrightarrow 2\text{H}^+ + \text{ClO}^- + \text{Cl}^-$ (multiple equilibria)	5000	H_2SO_4 , pH < 2
Ethanol amine, Et_2NH (ET = $\text{CH}_3\text{CH}_2\text{O}$)	$\text{Et}_2\text{NH}_2^+ \leftrightarrow \text{Et}_2\text{NH} + \text{H}^+$ (9.01)	1	0.01 M $\text{Et}_2\text{NH}_2\text{Cl}$, pH > 11
Nitrogen dioxide, NO_2	$2\text{NO}_2 + \text{H}_2\text{O} \leftrightarrow \text{NO}_3^- + \text{NO}_2^- + 2\text{H}^+$ (1.4–3.2; multiple equilibria)	0.5	Citrate buffer, no SO_2
Sulfur dioxide, SO_2	$\text{SO}_2 + \text{H}_2\text{O} \leftrightarrow \text{H}^+ + \text{HSO}_3^-$ (1.89; includes dehydration)	1	0.01 M NaHSO_3

Adapted from J. Janata, *Principles of Chemical Sensors*, New York: Plenum Press, 1989, pp. 150–151.

also been coupled to gas sensing in analogous fashion. Many enzyme electrodes work by further coupling enzymatic reaction steps in the membrane to eventual production or depletion of hydrogen ion, carbon dioxide, or ammonia. This allows for the detection of species that are not directly amenable to potentiometric analysis.

Ion-Selective Electrodes. As mentioned above, the alternative strategy for extending the concept of the glass electrode to other species is to develop membranes that are selective to the species of interest. Glasses have been developed for cations other than hydrogen, although the selectivity is generally much poorer. It is also possible to use (supported) liquid membranes or polymer films for ion sensing by incorporating ion-exchange sites that selectivity bind different target ions. Such ion-selective electrodes (ISEs) are widely used and commercially available. An instructive example is the sensing of potassium cations (K^+) carried out with a diphenylether film containing the ionophore valinomycin. Valinomycin is a cyclic peptide that binds to potassium with strength three to four orders of magnitude greater than that of competing cations such as sodium or ammonium. If the solution containing HCl in the pH sensor design is replaced by KCl and the glass membrane is replaced by the valinomycin membrane, the potential measured with a suitable reference electrode will be, again, 59 mV per decade change in the potassium ion activity. Such sensors have been used as potassium-selective microelectrodes to probe the ion concentration inside single living cells.

When the selective properties of the membrane are not sufficient to reject competing chemical species, the response often deviates from Nernstian behavior as shown in Fig. 5. At low target analyte concentrations, background ions can contribute to the signal through the generation of additional potentials. The exact analysis of such situations is complex and often approximated by the Nikolskij–Eisenman equation:

$$E = \Delta\phi = \phi_{\text{constant}} + \frac{2.303 RT}{z_x F} \log \left(a_x + \sum_i K_{x,i} a_i^{z_x/z_i} \right) \quad (16)$$

The selectivity coefficient, $K_{x,i}$, is defined to represent the contribution to the sensor response for each interfering ion and can be evaluated independently. Note that this equation shows the correct limiting behavior when the solution contains only a single ion, whether that ion is the target compound or an interfering one. Nonlinear calibrations can be used if the background is constant. Figure 5 also shows how the limit of detection is defined for ISEs.

Conductimetric Sensing. The electrical resistance change that accompanies the interaction of a target analyte with a conductive layer (typically a polymer or ceramic) held between two electrodes can also be used to construct a sensor. Such devices are generally known as conductimetric or resistive sensors. They are one of the simplest types of sensor to construct and have been widely used for the past 30 years, predominantly in gas-sensing applications. At the same time, however, they are among the least understood in terms of their selectivity properties. For this reason, many of the applications have been discovered by empirical modifications to various selective layers and to electrode physical design. The result is that conductimetric sensors often require detailed calibration with a variety of potential interferents in order to carefully define the

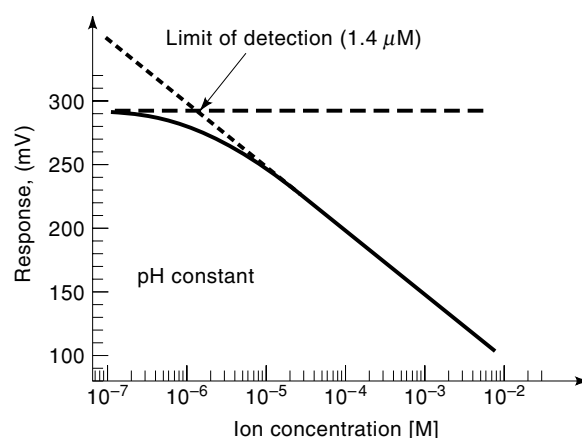


Figure 5. Deviation of ISEs from Nernstian behavior at low analyte concentrations. The deviation is due to presence of interfering ions. The IUPAC definition of the limit of detection (LOD) for ISEs is indicated by the intersection of the two straight-line extrapolations.

selectivity and be commercially useful. Nevertheless, these sensors are generally inexpensive, and the perceived disadvantages are easily outweighed.

The electrodes in a conductimetric sensing circuit contact the resistive element in such a way that direct-current (dc) or alternating-current (ac) measurements can be taken with the resistive layer exposed to the sensing fluid. Adsorption of the analyte occurs both at the surface and within the bulk of the material, which leads to changes in the electrical resistance. Such measurements can usually be done very accurately. Two particular types of materials have found the widest application in their use as selective layers: metal oxide semiconductors (MOSs) and conducting polymers (CPs).

Metal Oxide Semiconductors. At elevated temperatures (200°C to 600°C) in an oxygen atmosphere, oxygen ions (O^{2-}) are present on the surface of various metal oxides. These ions are formed by extraction of electrons from the oxide material. This electron depletion along with the surface morphology and microstructure will modulate the conductivity. If tin dioxide (SnO_2) is used, for example, the adsorbed oxygen ions will react with a variety of reducing gases to decrease the material resistance. The metal oxide functions as an *n*-type semiconductor in this case, and the underlying solid-state physics has been extensively investigated. The change in resistance can be quite dramatic, and linear responses have been observed over several decades of reducing gas concentration.

The requirement for high temperatures and oxygen (or other donor molecules such as water) leads to their predominance as gas sensors. The high temperature requirement could be met by the environment in which the sensor operates, but is often satisfied by integrating compact electrical heating elements into the design so they can function at room temperature conditions. The heating requirement necessitates a power consumption that is typically on the order of 10 mW. The tin oxide gas sensor (TGS) is the most popular MOS sensor and is produced by a variety of manufacturers. Some representative performance data are shown in Fig. 6 for Scimarec (Tokyo, Japan) gas sensors where the resistance is plotted as a function of gas concentration. Gases with resistance ratios much less than 1 will be easily sensed in a background of clean air.

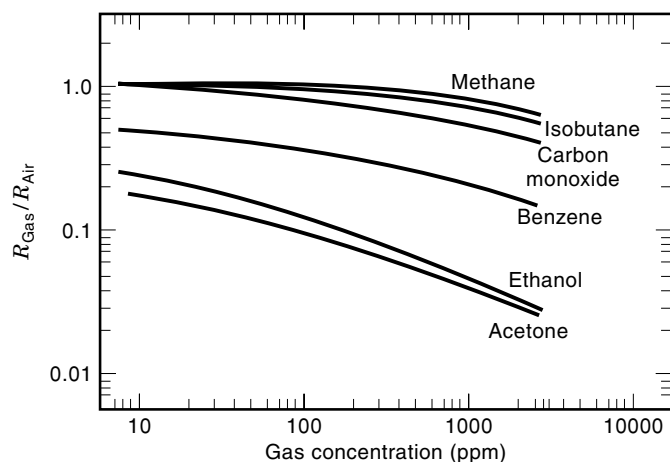


Figure 6. MOS sensor resistance ratios for target gases and for clean air as a function of the target gas concentration. [From Scimarec Co., Ltd. (Tokyo, Japan) specifications sheet (AF-63 sensor), with permission.]

The range of concentrations from 1 ppm to 1000 ppm is typical of the performance of MOS sensors. It is also seen that different sensor types have differing selectivities. The mechanism for imparting selectivity to the sensor is often not understood at a level to allow for rational design. Nevertheless, selectivity is known to depend on the addition of dopants that promote catalytic reaction steps and affect the semiconductor charge layer. The fabrication method can also change the MOS film thickness as well as the grain boundary profile that results in variable resistance changes. The particular choice of metal oxide (SnO_2 , In_2O_3) or the use of mixed metal oxides ($\text{SnO}_2 + \text{Fe}_2\text{O}_3$) will change the selectivity. The heating cycle can also be adjusted in some cases to dynamically manipulate selectivity. In general, proprietary fabrication and processing

steps distinguish one manufacturer from another. A basic dc circuit diagram that shows how an ethanol MOS sensor (Scimarec AF63) can be integrated with signal processing steps is seen in Fig. 7. The absolute ethanol level is replaced by readings corresponding to ethanol ranges such as low, medium, and high. Some additional features include ready, reset, power, and low battery indicators that make for a practical device.

Conducting Polymers. The use of a conducting polymer in place of an MOS has been recently investigated for many liquid and gas phase applications. The CP films can be made of polymers formed from acetylene, thiophene, pyrrole, furan, phenylene, quinoline, and aniline monomers, as well as their substituted analogs. The key to producing usable sensors is to formulate conditions such that films are easy to process and manipulate. Polypyrrole, polythiophene, and polyaniline, as well as their derivatives, are generally easier to process into thin films than other CPs. Manipulation of polymer properties is done by addition of plasticizers and other constituents either before or after the polymerization process. The films are usually placed between two Pt or Au (gold) electrodes arranged as an interdigitated structure over an inert substrate such as silica. The sorption of vapors or liquid phase components causes swelling of the polymer films and/or modification of the charge-carrier complexes along the polymer backbone that results in either decreases or increases in the resistance. As with MOS devices, extensive calibrations are usually required.

The chemical specificity of CPs is manipulated to a certain extent through the additives used. It is also possible to covalently couple high-affinity ligands to the polymer, and this has been demonstrated with antibodies. Antibodies are biological molecules that have a strong affinity for other proteins, and specific antibodies can be generated for specific proteins. For small target molecules, antibodies are more difficult to obtain, but some success has been demonstrated. Some ultrasensitive applications have been reported (8) for the conducting polymer poly(3-hexylthiophene) when it was used to detect hydrazine

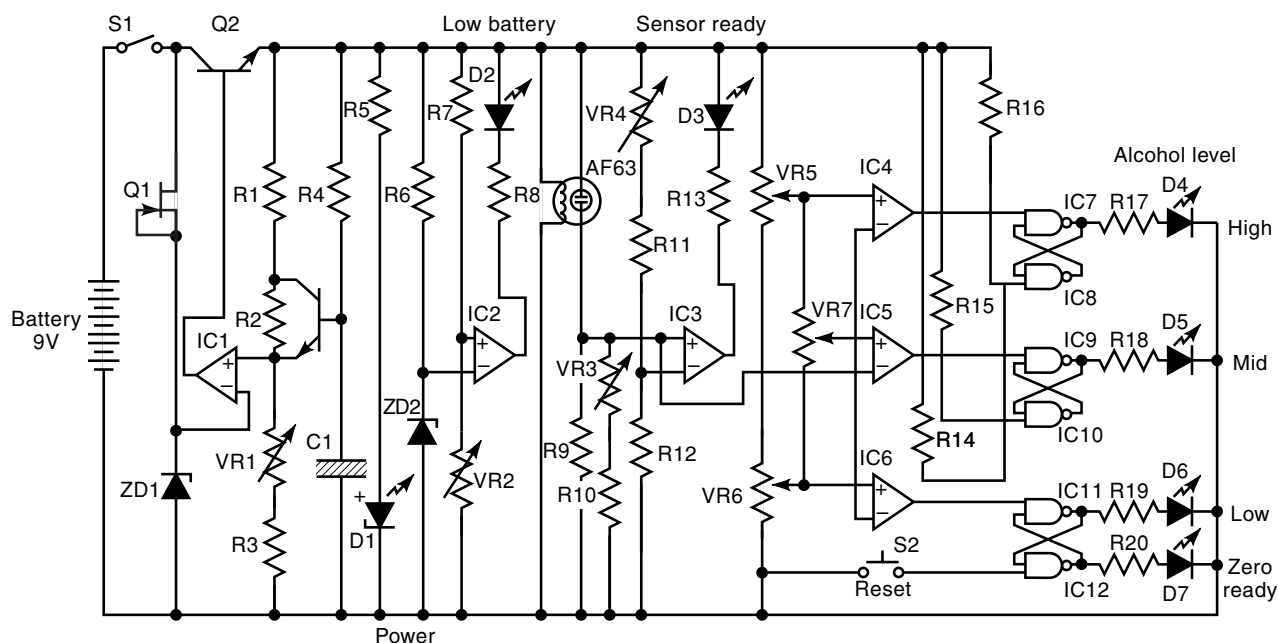


Figure 7. A circuit for utilizing an MOS sensor in the detection of gas-phase ethanol (alcohol).

vapor. This sensor, however, operated as an irreversible device and depended on the cumulative changes in resistance to measure hydrazine vapor concentrations. When operated in such a manner, these devices are often referred to as dosimeters.

Optical Sensors

Optical chemical sensors use interactions with light as the transducer component of the sensor. The current market for optical chemical sensors is smaller than the electrochemical sensor market, but it is a rapidly growing segment. Many of the optical measurements employed in the design of sensors have their basis in analytical spectroscopy and employ optical waveguides to transmit light to and from a suitable detection region. The availability of high-quality optical fiber (cylindrical) waveguides, high-power light sources at a variety of wavelengths, and sophisticated detectors, due primarily to advances in communications technology, has been a primary impetus for, and has greatly accelerated the use of, optical chemical sensors. It is also possible to use planar waveguides as components of optical sensors and these devices have, in fact, been among the first of the optical sensors to be commercialized. Applications that rely strictly on spectroscopic measurements without fiber or planar waveguides are also widespread, but they are not discussed in this overview. The general field of analytical spectroscopy for composition analysis of gases, liquids, and solids is covered in many other textbooks, handbooks, and encyclopedias that are devoted exclusively to the topic.

An optical waveguide is simply a medium through which light is transmitted in such a way that there are distinct light input and output regions. The transmission of light from the input to the output region is done with minimal loss, or the loss is designed to correspond to the sensing capabilities of the waveguide. In the case of a fiber configuration, the waveguide consists of a solid cylindrical “wire” (core) typically made of “clear” glass that is surrounded by a coating (clad) of another material with a different refractive index. Light is transmitted through the fiber by internal reflections as demonstrated in Fig. 8, where the core–clad interface is shown along with the position of a light source and detector. According to Snell’s law, the geometric relationship between refractive index and refraction angle (corresponding to a bend in the light path) is

$$n_1 \sin \Theta_1 = n_2 \sin \Theta_2 \quad (17)$$

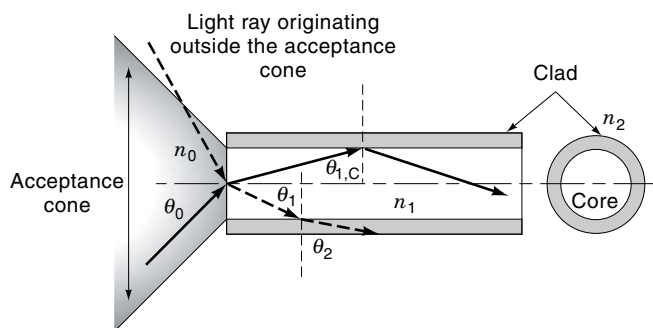


Figure 8. Refraction of light at an optical waveguide interface and at the entrance of a fiber waveguide. Light entering (or exiting) the fiber within the acceptance cone will be totally internally reflected. Additional sensor components often include specific choices for a light source and detector, various lenses, gratings, and/or filters.

If $n_1 > n_2$, the refracted angle (measured for the light ray passing into the medium of lower refractive index) can go to 90° , and essentially all of the light will be reflected. Therefore, a critical angle exists that depends on the ratio of refractive indices above which the light will be totally internally reflected. For a borosilicate glass with a refractive index of 1.52 and a quartz clad of refractive index 1.46 [so-called hard-clad silica (HCS) fiber], the critical angle is

$$\Theta_{1,c} = \sin^{-1} \frac{n_2 \sin 90^\circ}{n_1} = 74^\circ \quad (18)$$

Fiber waveguides are made of different types of glasses or plastics, and the variety of possible core–clad combinations lead to different critical angles. Snell’s law can also be applied to the light entrance region of a fiber, and the maximum angle of light that can be directed into the fiber and still undergo total internal reflection can be found. This “acceptance” angle is summarized in terms of the numerical aperture of the fiber, NA:

$$NA = \sin \Theta_0 = \frac{\sqrt{n_1^2 - n_2^2}}{n_0} \quad (19)$$

The numerical aperture is a convenient number to summarize the light collection efficiency of optical fibers and varies between 0 and 1. An NA of 0.5, for instance, corresponds to a 30° angle, which in three dimensions defines an acceptance cone. At the input to a fiber, therefore, NA gives a measure of the light-collecting ability of the fiber, and high NA is preferred. At the output of a fiber, NA describes the cone of light exiting the fiber. Fibers with high NA, however, direct light less well and present problems for coupling to detectors or other fibers where collimated beams are usually desired. Therefore, a compromise may often be needed.

Absorption Sensors. The light passing through an optical waveguide can interact with chemical species present within the acceptance cone of the fiber through absorption phenomena. At different wavelengths, the light is attenuated due to interactions with the electronic, vibrational, and/or rotational energy levels of the molecules. In the visible region (400 nm to 800 nm), electronic interactions dominate while at higher wavelengths (lower energy), vibrational and rotational energy transitions cause light absorption. As light of intensity P passes through an absorbing medium, the rate at which the intensity changes with distance penetrated, l , is directly proportional to both the local light intensity (Lambert’s law) and to the concentration of the absorber, c (Beer’s law). This can be expressed as

$$\frac{dP}{dl} = -kcP \quad (20)$$

This expression is an approximation (despite the use of the word *law*) that becomes more exact for dilute solutions and in the absence of secondary reflection, refraction, or re-emission effects. For a uniform concentration distribution, the integrated form of this equation yields the expression

$$\frac{P}{P_0} = e^{-kcl} = 10^{-kcl/2.303} \quad (21)$$

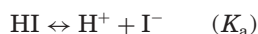
where the initial (input) light intensity is denoted by P_0 . The ratio of light intensities is commonly called the transmittance (or percent transmittance when multiplied by 100) and the logarithm (base 10) of the transmittance is known as the absorbance, Abs. Thereby, an absorbance value of 2, for example, corresponds to absorption of 99% of the light entering the sample. From this, the Beer-Lambert relation can be found, which conveniently summarizes the absorption of light as

$$\text{Abs} = \epsilon cl \quad (22)$$

The parameter $\epsilon = k/2.303$ is known as the extinction coefficient or molar absorptivity. Different compounds will have characteristic absorption spectra due to the wavelength dependence of the molar absorptivity and this is used for identification as well as quantification. The linear dependence between the absorbance and the concentration can lead to simple quantitative assays for measuring the composition of unknown samples.

Sensor selectivity can be achieved, in principle, by using the inherent spectral features of the target molecule. In most cases, however, additional steps are needed since the spectral features are highly overlapping and it is impossible to account for unknown interferents. As with electrochemical sensors, target specific reactions and membranes have been developed.

pH Optodes. In particular, pH optical sensors (referred to also as pH optodes or pH optrodes) utilize an otherwise transparent membrane film that contains an entrapped pH indicator. In the simplest case, the absorbance properties of the film will then vary as a result of the acid-dissociation equilibrium taking place with the indicator molecules distributed throughout the film. The reaction scheme is



The absorption spectra of the protonated indicator, HI, is different from the absorption spectra of the unprotonated indicator, I^- . Since the equilibrium shifts with increasing or decreasing hydrogen ions, by monitoring the change in the spectra, a calibration can be developed between the absorbance at a particular wavelength and the local pH in the vicinity of the film. The range over which the pH is measured depends on the $\text{p}K_a (= -\log K_a)$ value of the indicator and is usually restricted to one pH unit above and below the $\text{p}K_a$ value. Therefore, the pH optode has a much more limited dynamic range than the glass electrode. Nevertheless, it is of interest in applications where small size and electromagnetic immunity are important.

As an example, consider the (aqueous solution phase) absorption spectra of the pH-sensitive dye 8-hydroxypyrene-1,3,6-trisulfonic acid (HPTS) which is shown in Fig. 9(a) as a function of different solution pH values. By tracking the absorbance at 450 nm, a calibration curve relating pH to absorbance can be easily developed, and this is demonstrated in Fig. 9(b). The absorbance values are conveniently normalized according to the relationship

$$\gamma = \frac{\text{Abs} - \text{Abs}_{\text{prot}}}{\text{Abs}_{\text{unprot}} - \text{Abs}_{\text{prot}}} = \frac{I^-}{I_{\text{tot}}} \quad (23)$$

where Abs is the experimentally measured absorbance at the appropriate wavelength. The subscripts refer to protonated and unprotonated species. An analytical expression can be derived from the definition of the acid dissociation equilibrium constant

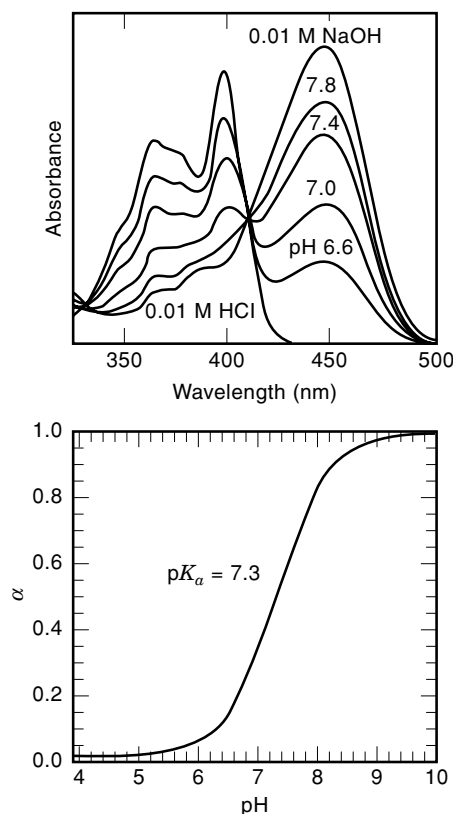


Figure 9. (a) The change in absorption spectra of HPTS (aqueous solution) at different pH values. (From R. P. Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 6th ed., Eugene, OR: Molecular Probes, 1996, p. 557, with permission.) (b) The absorption maxima at 450 nm can be plotted as a function of the pH to generate a calibration curve. The overlap of different curves at 405 nm is seen to be independent of the pH and defines a useful calibration wavelength.

and a mass balance on the total indicator concentration, $I_{\text{tot}} = \text{HI} + \text{I}^-$:

$$(\alpha_{\text{H}^+})K_a^{-1} = \frac{1 - \gamma}{\gamma} \quad (24)$$

This equation can also be written as

$$-\text{pH} + \text{p}K_a = \log_{10} \frac{1 - \gamma}{\gamma} \quad (25)$$

and emphasizes that when $\alpha = 0.5$, $\text{pH} = \text{p}K_a$. A membrane containing the HPTS dye at the face of an optical fiber could function in this way as a suitable pH optode. However, its high water solubility and “blue” (low visible wavelength) absorption spectra make it less desirable for such applications than other dyes that better match the light output available from cheap light-emitting diode (LED) sources and matching detectors. Fortunately, many other dyes are known with a wide range of $\text{p}K_a$ values and absorption spectra. This allows for a broad range of optical as well as chemical characteristics to be designed into a pH sensor. The other feature in Fig. 9, which is seen for many indicators, is the so-called isosbestic point or pH-independent absorption wavelength at 405 nm. By ratioing measurements at 450 nm and 405 nm, a self-calibrating pH sensor can be

designed that accounts for fluctuations in light source intensity as well as degradation or loss of the dye from the membrane phase.

Ion-Sensitive Optodes. The pH optode can also be adapted to the measurement of other cations and anions. The mechanism by which this takes place has been summarized for a wide variety of ion-exchange mechanisms (9). Briefly, if a membrane is designed that contains a pH indicator as well as an adsorption site for other ions (mediated by the incorporation of an ionophore into the membrane), the requirement for electroneutrality causes charged species to move in pairs. If a positively charged cation is adsorbed on the surface, for instance, a proton is exchanged in the opposite direction to preserve electroneutrality. For anionic sensors, a negatively charged anion brings with it the positively charged proton to the surface. In either case, the affinity sites for the ions act together with the absorption spectra of the particular pH indicator to generate a color change. Such optical ion sensors will suffer from selectivity problems, however, in much the same way as the corresponding ion-selective electrochemical electrodes. The selectivity of the pH-sensitive dye for hydrogen ions is usually absolute, but the ionophores chosen for other ions are more flexible in their capacity to bind different chemical species. It is also necessary to independently monitor or fix the pH of the solution to ensure that the measured spectral changes are associated with ions other than hydrogen.

The selectivity for the sensor can be evaluated from a consideration of the following mechanism, which is an expansion of the simpler acid-dissociation scheme written earlier (see pH Optodes):



The subscript f denotes film-associated components, and S refers to the ionophore. This equation is written for an anion exchanger with a stoichiometric binding ratio of 1:1. The indicator, I_f , and the anionic binding site, S_f , are both confined to the film while the hydrogen ion and target anion, X_i^- , can freely exchange with the bulk solution. A similar equation exists for each anion, X_i^- . In this equation, the exchange coefficient, K_{X_i} , plays the same role as the previously described acid dissociation constant to summarize the reaction equilibrium. We also consider only an anion with a single charge of valence = -1 for illustrative purposes. Using an electroneutrality balance together with mass balances on the indicator and ionophore, the relation equivalent to Eq. (24) is

$$(a_{X_i^-})(a_{H^+})K_{X_i} = \frac{1 - \gamma}{\gamma} \frac{\omega_i}{\gamma\beta} \quad (26)$$

where the a_i are the activities and β is the (molar) ratio of indicator to ionophore loaded into the membrane. The fraction of ionophore that is complexed with the target anion is ω_i and $\Sigma \omega_i = 1 - \alpha\beta$.

For a solution containing only an individual anion ($i = 1$), a calibration curve can be developed from these equations and, in particular, the exchange coefficient can be found. A single-ion response function is then defined by

$$a_{X_1}^0 = \frac{1}{K_{X_1} a_{H^+}} \frac{1 - \gamma}{\gamma} \frac{1 - \gamma\beta}{\gamma\beta} \equiv f_1(\gamma) \quad (27)$$

where the superscript on the anion activity is used to emphasize that the response function is measured in the absence of interferents. If the equations are now expanded to consider a solution of competing (interfering) anions with activities a_{X_j} , the result can be compactly expressed as

$$a_{X_1} = f_1(\gamma) - \sum K_{1j} a_{X_j} \quad (28)$$

The selectivity coefficient $K_{1j} = K_{X_1}/K_{X_j}$ is found from experiments with pure component data. This relation is the optical analogue to the Nikolskij-Eisenman equation, Eq. (16), introduced previously for ISEs. It summarizes how the presence of interferents will lead to an overprediction of the target ion concentration if the single-ion response function is blindly used. Similar equations can be developed for solutions of ions with mixed valences as well as for the analogous cation sensor (9).

Gas-Sensing Optodes. It is also apparent that pH optodes can be used for the measurement of gases such as CO_2 and NH_3 in a manner exactly analogous to that described for gas-sensing electrodes. Many patents have been issued, and several commercial designs of optical sensors have been put forward that detail exactly how the light from a fiber waveguide will pass through the measuring sample.

Evanescent Sensors. Light passing through an optical fiber can also interact with the surrounding medium at the core-cladding interface through the so-called evanescent field. This electromagnetic field extends from the interface into the medium surrounding the fiber (or any reflection surface) only a short distance—typically 0.1 μm . Nevertheless, this makes it possible to carry out optical measurements using a fiber in a “closed-loop” configuration. It also allows for the use of membranes that impart selective advantages toward different target analytes while at the same time function as an integral part of the optical waveguide. The light intensity penetrating into the evanescent field is much lower than the intensity found in the acceptance cone region. This is compensated for by using long pathlengths, which are introduced by exposing correspondingly long sections of the fiber to the analyte fluid.

In environmental sensing applications that deal with the measurement of volatile organic compounds in water, hydrophobic membranes can effectively separate and concentrate target analytes from the dilute aqueous background. In fact, the plastic-clad (silicon) silica fibers that are readily available from many commercial sources have been demonstrated to function quite well as “evanescent fiber-optic chemical sensors” (EFOCS) for the analysis of chlorinated hydrocarbons in water (10). Organic species absorbed into the silicon clad interact with the evanescent light field leading to wavelength-dependent light absorption that is characteristic of the composition of the surrounding solution. Figure 10 shows a comparison of the absorbance spectra measured with an EFOCS and with a conventional spectrometer for toluene, trichloroethylene, and trichloroethane. It is seen that long lengths of fiber are needed to generate even small absorbances (note also the $10\times$ scale change) compared with conventional transmission spectroscopy. The spectra are plotted as a function of the wavenumber, ν , which is related to the wavelength, λ :

$$\nu (\text{cm}^{-1}) = \frac{10^7}{\lambda (\text{nm})} \quad (29)$$

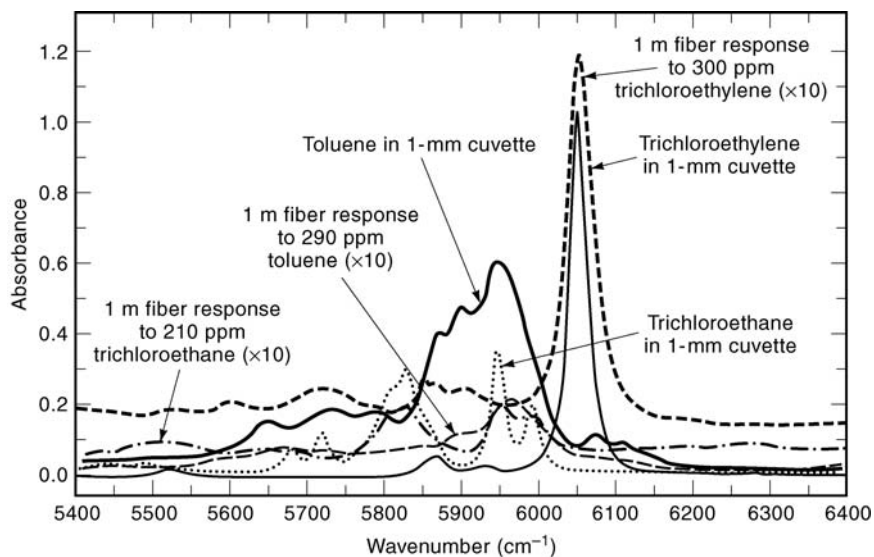


Figure 10. Equilibrium EFOCS responses at specified analyte concentration (magnified $\times 10$) and transmission spectroscopy spectra of the pure components measured in a conventional instrument. (From Ref. 10, with permission. Copyright 1997 American Chemical Society.)

Fluorescence Sensors. Absorption of light can also lead to re-emission at lower energies (longer wavelengths). This can take place from an excited, singlet state to the ground level and leads to what is referred to as fluorescence. Applications of fluorescence to chemical sensing using optical waveguides are very similar in terms of the target compounds and mode of operation to those seen for absorbance sensors. One important additional feature, however, distinguishes this type of spectroscopy: The detection of emitted light often takes place against a zero (dark) background. As a result, fluorescence measurements are often done at a high signal-to-noise ratio. In combination with pulsed light sources and phase-resolved signal analysis, very accurate measurements can be taken of very small numbers of molecules. In this way, small changes can be monitored, and fluorescence assays are usually many orders of magnitude more sensitive than competing absorbance-based assays. The previously cited HPTS molecule, for example, emits light via fluorescence when excited at the absorbance wavelengths shown in Fig. 9. Such molecules are referred to as fluorophores and display a separate emission spectrum for each absorbance wavelength. The membrane-based design described earlier can now be used in a fluorescence-sensing mode by separating the incident excitation light from the emitted light using optical filters, gratings, and/or tunable light sources. In fact, by alternately exciting at 405 nm and at 450 nm, a ratioing scheme can also be devised for this type of fluorescence sensor that allows for internal calibration. Fluorescent optical sensors have found widespread application in many biosensing schemes where coupling of enzymatic reactions with fluorescence detection leads to extraordinary selectivity and sensitivity combinations.

Fluorescence quenching is another phenomenon that leads to reduction of the emitted light due to the presence of so-called quenching molecules. Optical sensors have also been designed around quenching effects, and oxygen, in particular, is a well-known quencher molecule. The relation between fluorescence intensity and oxygen concentration, for a constant excitation intensity, is often summarized by the Stern–Volmer equation:

$$\frac{F_0}{F} = 1 + K_{SV}[\text{O}_2] \quad (30)$$

This equation predicts a linear relation between the inverse of the fluorescence intensity, F , and the oxygen concentration, $[\text{O}_2]$. Such quantitative behavior is often seen in practice, but sensors are also empirically adapted to a modified form of the Stern–Volmer equation in order to generate useful calibration functions. The temperature dependence of the Stern–Volmer constant, K_{SV} , is significant, and simultaneous temperature measurement or temperature control is often needed in interpreting results. This is true of any fluorescence-based chemical sensor, and the fluorescence is typically decreasing with increasing temperature. In contrast, the absorbance-based detection schemes are more immune to temperature fluctuations.

As with absorbance-based sensors, evanescent interactions can also be exploited and many of the designs proposed for fluorescence-based assays utilize fluorophores that are immobilized on a short segment of optical fiber at the core–clad interface. The fibers are often tapered to allow for better interaction of the fluorophore with the excitation light and better capture of the fluorescence emission. The thin interfaces that are used in evanescent detection can also lead to faster response times for the sensor compared with situations observed for thick clads.

Refractive-Index Sensors. Optical chemical sensors have been developed that are based on measurements related directly or indirectly to changes in the refractive index. The cladding can be used as a selective adsorbent (typically a polymer) which changes its bulk refractive index due to absorption of target molecules (11). Such a sensor functions as a refractometer, although absolute values of the refractive index are rarely obtained. Measurements of changes in the light intensity are related to the changes in the critical angle for total internal reflection described previously. Light can essentially “leak” from the fiber if the refractive index of the clad increases from its nominal level. The refractive index of the clad also affects the NA value of the fiber that can be observed in the exit light intensity measured for a fixed solid spherical angle. Such refractive index sensors, for example, have been introduced for monitoring aromatic hydrocarbons in groundwater.

Absolute refractive index changes can also be monitored through so-called surface plasmon resonance (SPR) effects.

Surface plasmon resonance occurs when polarized light is reflected off a thin metal film covering a high refractive index material, such as glass, at particular incidence angle and wavelength combinations. Gold is often used to produce the thin film. The resonance effect is observed by rapid attenuation of the light intensity when either the angle or the wavelength of light is scanned across critical values. The critical values will change with the local refractive index at the interface; thus, molecules that adsorb to the gold surface and modulate the refractive index can be detected. A refractive index SPR sensor is most often fabricated on planar (flat) waveguides, but it has also been demonstrated on an optical fiber. Selectivity can be achieved by modification of the surface with specific affinity ligands for the target compound of interest. For example, in biosensing applications, antibodies have often been used for this purpose. One company, Biacore International AB (Uppsala, Sweden), markets instruments based on SPR technology for a variety of biosensor applications. Other designs have also been introduced (12) that utilize compact, integrated electrooptic features.

Sensor Arrays

Selectivity is one of the most important parameters in developing useful sensors (see Fig. 1). This is true for electrochemical as well as optical designs, and much effort has been devoted to fabrication and discovery of specialized surfaces. An alternative approach is to forgo absolute specificity and rely, instead, on simultaneous measurements over many incrementally selective surfaces. These sensor arrays can provide for versatile operation without highly specific interactions taking place. In the detection of volatile compounds and in odor-sensing applications, they are often referred to as "electronic noses." They have been used in analysis of food processing operations, perfumes and cosmetic production, environmental sensing and a variety of other application areas where odor is a recognizable quality indicator. Some companies that produce electronic noses commercially include Alpha M.O.S. America (Belle Mead, NJ), AromaScan (Hollis, NH), and Neotronics Scientific (Essex, UK).

Sensor arrays have been most often developed with conductimetric sensors such as MOS and CP devices. They can also include chemical sensors based on surface acoustic wave (SAW) transduction schemes [see Surface acoustic wave devices], and often a mixture of sensing elements is used to impart the greatest degree of flexibility. The interpretation of signals from such arrays requires extensive calibrations with pure components and/or known mixtures. The signal analysis often uses statistical methods, including principal component and partial least-squares analysis, or artificial neural networks to recover compositional information on unknown samples. Many times, it is sufficient to categorize the mixture in terms of "good" or "bad" or other loosely defined quality descriptors.

For example, a 14-element CP array has been developed that was shown to be capable of discriminating 8 different solvents (13). The CP array was fabricated from the conducting polymer poly(pyrrole), which was blended with 12 different plasticizers and two different combinations of oxidant. The polymers were dip-coated onto interdigitated electrodes, and resistance measurements were simultaneously performed on all elements of the array after exposing the array to an air-vapor gas phase

of fixed composition. The data were analyzed for each array element by a simple binary criteria related to whether an individual sensor element was responding within a given range or not. The (saturated) pure components in air could be distinguished from each other and, with arrays of 19 elements, individual saturation levels as well as binary mixtures in air could also be quantified. The presentation of the data is often done with respect to the principal components, and in this case it was found that three components were sufficient in this particular experiment.

Arrayed sensors represent a rapidly emerging area of chemical sensor development. New coatings and structures as well as new methods for signal analysis greatly expand the role for chemical sensors into applications that were once deemed too complex. The ability to multiplex diverse sensors will continually expand their application to new areas of use. Together with miniaturization and efficient fabrication techniques, it is expected that chemical sensors will, in the near future, become pervasive in industry, medicine, the environment, and indeed everyday use.

BIBLIOGRAPHY

1. R. F. Taylor, Chemical and Biological Sensors: Markets and Commercialization, in R. F. Taylor and J. S. Schultz (eds.), *Handbook of Chemical and Biological Sensors*, Bristol, UK: IOP Publishing, 1996, pp. 553–579.
2. N. Yamazoe, Chemical sensors R&D in Japan, *Sens. Actuators B*, **6**: 9–15, 1992.
3. J. Y. Lucisano, J. C. Armour, and D. A. Gough, In vitro stability of an oxygen sensor. *Anal. Chem.*, **59**: 736–739, 1987.
4. L. R. Jordon, P. C. Hauser, and G. A. Dawson, Amperometric sensor for monitoring ethylene, *Anal. Chem.*, **69**: 558–562, 1997.
5. R. G. Compton and G. H. W. Sanders, *Electrode Potentials*, New York: Oxford Univ. Press, 1996.
6. A. J. Bard, R. Parsons, and J. Jordon (eds.), *Standard Potentials in Aqueous Solution*, New York: Marcel Dekker, 1985.
7. K. Denbigh, *The Principles of Chemical Equilibrium*, 3rd ed., London: Cambridge Univ. Press, 1971.
8. D. L. Ellis et al., Conductive polymer films as ultrasensitive chemical sensors for hydrazine and monomethylhydrazine vapor, *Anal. Chem.*, **68**: 817–822, 1996.
9. E. Bakker and W. Simon, Selectivity of ion-sensitive bulk optodes, *Anal. Chem.*, **64**: 1805–1812, 1992.
10. D. S. Blair, L. W. Burgess, and A. M. Brodsky, Evanescent fiber-optic chemical sensor for monitoring organic compounds in water, *Anal. Chem.*, **69**: 2238–2246, 1997.
11. S. M. Klainer, Fiber optic which is an inherent chemical sensor, US patent 4,846,548, 1989.
12. J. Melendez et al., A commercial solution for surface plasmon sensing. *Sens. Actuators B*, **35** (1–3): 212–216, 1996.
13. M. S. Freund and N. S. Lewis, A chemically diverse conducting polymer-based "electronic nose," *Proc. Natl. Acad. Sci.*, **92** (7): 2652–2656, 1995.

HENRIK PEDERSEN
Université du Québec à Hull

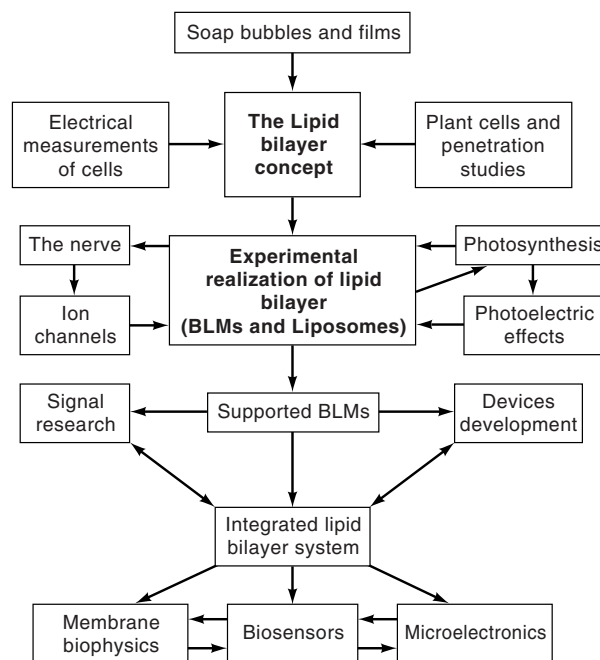
HITOSHI MUGURUMA
Naval Research Laboratory

ERNESTO DELA CRUZ
Rutgers, The State University of
New Jersey

INTELLIGENT BIOSENSORS

BACKGROUND

In the past decade, there have been a number of reports on self-assemblies of molecules as “advanced materials” or “smart sensors”. Without question, the inspiration for this exciting development comes from the biological world, where, for example, the lipid bilayer of cell membranes is among the foremost self-assembling system. Although the first reports on self-assembled bilayer lipid membranes (BLMs and liposomes) appeared in the 1960s (1), experimental scientists have been dealing with BLM-type interfacial adsorption phenomena since Robert Hooke’s time (see Table 1). BLMs (planar lipid bilayers) have been used in a number of investigations ranging from basic membrane biophysics to the conversion of solar energy via water photolysis, and to biosensor development using supported bilayer lipid membranes (s-BLMs). A cogent reason that self-assembled BLMs and liposomes are of sustained scientific and practical interest is due to the fact that most physiological activities involve some kind of lipid bilayer-based ligand-receptor contact interactions. For instance, by embedding a receptor or a host of specific entities in a supported BLM, it is possible to create a sensor that will interact with its environmental counterpart. In the accompanying figure (Fig. 1), an overview is presented. From the top of Fig. 1, soap bubbles were among the first ultrathin films quantitatively characterized, from which the origin of the lipid bilayer concept of cell membranes and its experimental realization, as we know it today, can be traced (1). This is shown in the center of Fig. 1. In this connection, electrical activities of the nerve and photoelectric effects in photosynthesis have been elucidated with the aid of experimental planar BLMs and liposomes. In nerves, translocation of ions across protein channels embedded in the lipid bilayer play the pivotal role. In photosynthesis, light absorption by pigments, confined in the lipid bilayer, initiates electronic charge generation and separation, leading eventually to redox reactions on opposite sides of the membrane. Presently, electron and charge transfer processes through the lipid bilayer are being actively investigated, from both theoretical and biotechnological viewpoints. The crucial role played by the lipid bilayer may be summarized as follows. Living organisms are made of cells bound by their membranes. As we know them, living beings are self-assembling entities, each is organizing a particular combination of phospholipids in the form of a bilayer with other constituents (e.g., proteins) embedded in it. This lipid bilayer, existing in all biomembranes, is most unique



BLM = Bilayer Lipid Membrane

Figure 1. Ultrathin films, biomembranes, and applications (an overview).

in that it serves not merely as a physical barrier but functions as a two-dimensional matrix for surface reactions, as well as for transmembrane processes. The lipid bilayer acts as a conduit for ion transport, as a framework for antigen-antibody binding; as a bipolar electrode for redox reactions; as a reactor for energy conversion (e.g., light to electrical and/or to chemical). Further, a modified lipid bilayer performs as a transducer for signal transduction (e.g., molecular sensing), as well as numerous other functions (1–6). All these myriad activities require the ultrathin (~5 nm) lipid bilayer which is the focus of our research. Many attempts have been made to develop biosensors and molecular devices based on planar BLMs; the task has been a daunting one until a few years ago (7,8), since a 5 nm thick BLM is an extremely labile structure with limited lifetime. Planar BLMs can now be formed on various substrates with long-term stability, thereby opening the way for basic research and development work. Before explaining the lower portion of Fig. 1, mention should be made concerning the self-assembly. Nature uses self-assembly as a strategy to

Table 1. Self-Assembling Systems Containing Amphiphilic Molecules

System	Literature Source
1. Soap films	R. Hooke, in <i>The History of the Royal Society of London</i> , 3 : 29, 1672.
2. Monolayers	I. Langmuir, <i>J. Am. Chem. Soc.</i> , 39 : 1848, 1917.
3. Langmuir–Blodgett multilayers	K. B. Blodgett and I. Langmuir, <i>Phys. Rev.</i> , 51 : 964, 1937.
4. Planar bilayer lipid membranes (BLMs)	P. Mueller et al., <i>Nature</i> , 194 : 979, 1962.
Liposomes (lipid microvesicles)	A. D. Bangham, <i>BioEssays</i> , 17 (12): 1081, 1995.
5. Nucleopore supported BLMs	J. D. Mountz and H. T. Tien, <i>Photochem. Photobiol.</i> , 28 : 395–400, 1978.
6. Gold supported monolayers	(a) L. Taniguchi et al., <i>J. Electroanal. Chem.</i> , 140 : 187, 1982.
	(b) R. G. Nuzzo and D. L. Allara, <i>J. Am. Chem. Soc.</i> , 105 : 4481, 1983.
	(c) L. Netzer and J. Sagiv, <i>J. Am. Chem. Soc.</i> , 105 : 674, 1983.
	H. T. Tien and Z. Salamon, <i>Bioelectrochem. Bioenerg.</i> , 22 : 211, 1989.
7. Metal supported BLMs (s-BLMs)	(a) H.-P. Yuan et al., <i>Mater. Sci. Eng. C</i> , 4 : 35–38, 1996.
8. Salt-bridge supported BLMs (sb-BLMs)	(b) X.-D. Lu et al., <i>Bioelectrochem. Bioenergetics</i> , 39 : 285–289, 1996.

create complex, functional structures such as viral protein coatings, and DNA, besides the above-mentioned *lipid bilayer* of cell membranes. Supported BLMs, formed on metallic wires, conducting SnO₂ glass, and hydrogel substrates, as well as on microchips, possess properties resembling biomembranes. These self-assembled, supported BLMs, have opened research opportunities in studying hitherto unapproachable phenomena at interfaces. Some recent findings demonstrate potential for investigating processes at solid-liquid interfaces (1–6,8). Possible applications include catalysis, corrosion, electrochemistry, lithography, and membrane biophysics (9–15). In the lower portion of Fig. 1, the focus is on *Integrated Lipid Bilayer System*, from which research endeavor is divided into three related areas, namely, membrane biophysics, biosensors, and microelectronics. These are delineated below under separate headings.

Membrane Biophysics. In connection to information processing of the brain, i.e., “How does the brain work?” This is based on the fact that living beings possess excellent mechanisms for recognition, learning, memory, and inference. Understanding the mechanisms will bring about a novel architecture for information processing that is quite different from current computing systems. Experimentally, we have studied the olfactory system of an animal (e.g., frog) whose system is oscillatory neural network, with which highly parallel and distributed information such as odors are processed (1). Concurrently, we would like to carry out molecular simulations using a modified planar lipid bilayer to gain insight at the molecular level (9,14–17). Here, perhaps the concept of force measurement on the nanometer scale, and the method developed to evaluate them would provide a quantitative logic to ponder about the new ways in which molecules behave at nanometer distances.

Biosensors. The research is based on the lipid bilayer from which system sensor concepts can now be advanced. The lipid bilayer provides an environment for receptors similar to the biomembrane (such as immune receptors, growth factor receptors, cytokine receptors, etc). For example, currently most microbiological tests are centralized in stationary laboratories because complex instrumentation and highly qualified technical staff are required. Traditional methods for the detection of bacteria involve at least four basic steps (2–4,18,19). Hence, a complex series of tests is often required before any identification can be confirmed. These methods are often slow (up to 72 h), and may vary in time since the development of a colony containing 10⁶ organisms will take 18–24 h. With the availability of s-BLM probes coupled with microelectronics techniques, an entire new approach is ripe for molecular sensing. The plan is to decentralize the above mentioned procedures using s-BLMs, so that they can be performed virtually anywhere and under field conditions. The long-term goal is to develop a novel type of an electrochemical biosensor for label-free detection of bacteria in real-time (<10 minutes) for medical and environmental application with immediate on-the-spot interpretation of the results. The goal is to achieve extremely high selectivity and sensitivity (5–20 bacteria per sample). Concurrently, working closely with our colleagues in microelectronics, a small analytical device for on-site monitoring of different kinds of pathogenic bacteria in air and water will be studied. Previous results have demonstrated the suitability of the electrochemical approach for the label-free measurements of specific bimolecular interactions. The biosensor will exploit the phenomena of the trans-

membrane ion-current modulation of biospecific interactions in the solid-state surface layer (1–4). Detection of channel opening or closure will be realized directly by monitoring a transient change. Nonspecific binding at a biospecific membrane does not interfere with the ion flux through ion channels. Theoretically, the sensitivity of such a device can approach one molecule. This principle could be also applied as a strategy for sensor design using BLM-doped receptors or ligands.

Microelectronics. The use of a strip thin-film microsystem, based on the silicon diaphragm or glass substrate, is progress. Experiences with thin-film heater and planar interdigital array of electrodes have been reported (20–22). The thin-film and silicon micromachining techniques are compatible with microelectronics IC technologies that allow a further development toward smart sensors, based on the supported s-BLMs. Ideally, one would like to develop a metal/BLM/liquid integrated system that functions as a transducer-actuator. A suitably modified BLM-bifacial can receive appropriate information, act upon it, and initiate a response. In physicochemical terms, a ligand interacts, with a lipid bilayer-bound receptor, resulting in a signal. This signal is transduced from the contact interaction into an electrical output, which in turn actuates a chemical or a physical process. The stimulus for interactions might be electrical, mechanical, light, or chemical, with the electrical perturbation most elegant and easy to control and manipulate. The electric field across the BLM/bifacial is very sensitive to bifacial tension and BLM thickness, which are in turn susceptible to the embedded compounds. Externally applied electric fields act on the BLM/bifacial by inducing surface charge and by creating transient pores, although the exact process is still obscure and remains to be studied (10,13). In this connection, according to solid-state physics, when electrons are confined in a narrow space (i.e., a lipid bilayer) comparable to an electron wavelength, the electrons are subject to interrelated quantum effects, size quantization and resonance. The former causes the continuum of energy levels that usually exists in the conduction band of a solid to become articulated into discrete energy quanta or states. A planar lipid bilayer, a self-assembled supramolecular structure is microscopic in thickness (~5 nm). Such a structure is in a liquid-crystalline state, and inherently dynamic. Upon modification with compounds (e.g., fullerene C₆₀), non-linear electronic behavior and photoelectric phenomena may be examined for unusual quantum-size effects. Thus, a supported BLM is a research area of many exciting possibilities, since it is the dynamic lipid bilayer surface that enable our senses to contact with the environment (1). For our senses, lipid bilayer-bound receptors are involved. These are rod and cone receptors for vision, olfactory bulbs for smelling, tectorial cells for hearing, taste buds for tasting, and nerves for touching. From environment to the brain, signal transduction and information processing must be transmitted in a language that the brain can understand. Insofar as can be ascertained, this language is electrical, for most part based on the translocation of ions and electrons. Protein-channels and electron transfer chains, embedded in the lipid bilayer, are the keys to our understanding (1).

It is worth pointing out that, with all other systems, the entity of interest is immobilized in a rigid, solid-like structure (e.g., polymers, L-B films), whereas in the BLM it is embedded. By embedding is meant that the entity of interest (i.e., membrane modifiers such as polypeptides, proteins, electron acceptors, donors, tissues or whole cells) in the lipid bilayer

is relatively free to adapt to its surroundings. The functions of lipid bilayers are mediated via specific modifiers, which assume their active conformations only in a hydrophobic lipid environment (11,13,14,23,24). Further, the presence of the lipid bilayer greatly reduces the background noise (interference) and effectively excludes hydrophilic electroactive compounds from reaching the detecting surface, thereby causing undesired reactions. The development of conventional BLMs (planar lipid bilayers) and later s-BLMs and sb-BLMs, have made it possible for the first time to study, directly, electrical properties and transport phenomena across a 5 nm ultrathin structure separating two phases, in particular as models of biomembranes. As a result of extensive studies over the past four decades, biomembranes have now been recognized as the basic structure of Nature's sensors and molecular devices. To impart relevant functions in BLMs, a variety of compounds such as ionophores, enzymes, receptors, pigments, tissues, and so on have been embedded. Some of these incorporated compounds cause the BLMs to exhibit non-linear phenomena and photoelectric effects (25–29). The self-assembled lipid bilayer, the most crucial component of all biomembranes, is in a liquid-crystalline and dynamic state. A functional cell membrane system, based on self-assembled lipid bilayers, proteins, carbohydrates and their complexes, should be considered in molecular and electronic terms; it is capable of supporting ion and/or electron transport, and is the site of cellular activities in that it functions as a "device" for either energy conversion or signal transduction. Such a system, as we know intuitively, must act as some sort of a transducer capable of gathering information, processing it, and then delivering a response based on this information. Today, planar lipid bilayer research is a matured field of endeavor, as a result of applications of many disciplines and techniques including interfacial chemistry, electrochemistry, patch-clamp techniques, spectroscopy, microelectronics, and others. In membrane reconstitution experiments, for example, the evidence is that intracellular signal transduction begins at membrane receptors. The research area covered in this paper is highly interdisciplinary. Emphasis has been placed on basic research. The past work has been benefited by a cross-fertilization of ideas among various branches of sciences. The biomimetic approach to practical applications is unique and full of exciting possibilities. We can glean the design principles from Nature's successful products and apply them to our research and development, from which molecular medicine and advanced biosensors may ultimately depend. It should be stated that manipulating and controlling conduction processes of an integrated lipid bilayer system at the nanometer dimension remains a challenge for both experimentalists and theoreticians. From specificity, selectivity, and design points of view, a supported planar lipid bilayer is an ideal natural milieu for embedding a host of materials of interest. For example, mention should be made concerning a most unique group of compounds, namely fullerenes that are of interest in materials science. Fullerene C_{60} , for instance, is an excellent electron mediator and behaves as n-type semiconductor. It is strongly hydrophobic. Hence, a BLM is an ideal environment for the compound (12,17,24,30–45). We, along with others, have investigated electron transfer across an s-BLM containing C_{60} , which can act both as an electron carrier and a photosensitizer. In this connection, fullerenes are expected to play an important role in photodynamic therapy and device development in the coming years. Finally, it should be stated that manipulating and controlling conduction

processes of an integrated lipid bilayer system at nanometer dimension, a system of choice remains a challenge for both experimentalists and theoreticians.

OVERVIEW OF A BIOSENSOR

The interaction of biophysics with cell physiology and molecular biology has given rise to an exciting area of research termed membrane biophysics, which integrates up-to-date findings on molecules and processes involved in inter- and intracellular recognition and communication. Knowledge of the ideas and findings resulting from such interdisciplinary research are now being used for practical applications in analytical chemistry, immunology, photobiology, chemical/biological sensors and transducers, and molecular electronics.

What is a Biosensor?

A biosensor is an analytical device incorporating biological and chemical sensing elements either intimately connected to, or integrated with, a suitable transducer, enabling the conversion of concentrations of a specific species into digital electronic signals. A majority of the biosensors developed thus far have incorporated an enzyme as a biological recognition component. For example, all types of enzyme sensors are based on the classic idea which measures glucose by detecting the reduction in oxygen when the oxidation of glucose is catalyzed by the enzyme. To date, there are many substrates that have been studied by the use of oxidoreductases, and the majority of the enzyme biosensors have been designed specifically for the determination of a large number of "cardiac" enzymes in blood.

Tissue materials from plant and mammalian sources have been successfully employed for the construction of biosensors as well (46). This class of biocatalytic materials simply maintains the enzyme of interest in its natural environment (e.g., lipid bilayer), which results in a considerable stabilization of the desired enzymatic activity.

The microbial sensors are composed of immobilized microorganisms and an electrochemical device and are suitable for the on-line control of biochemical processes. These sensors involve the assimilation of organic compounds by the microorganisms, change in respiration activity, or production of electroactive metabolites. These changes have been monitored directly by an electrochemical device.

The sensitivity of electrical measurements developed for the electrochemical biosensors, coupled with the specificity of antigen–antibody reactions, provides a useful tool for immunology. However, more recently, the optical sensors for immunoassays have been receiving considerable attention in research laboratories and also *in vivo* applications. Among different types of optical biosensors, two types appear to be especially promising. One is based on a surface plasmon resonance phenomenon, and the second one is a fluorescence capillary fill device. Surface plasmon resonance in a thin metal film deposited on a wave guide can be induced by an electromagnetic wave generated when the light is reflected within the wave guide, and it is highly sensitive to variations in the refractive index of the immediate surrounding medium. This phenomenon is monitored by a reduction in the intensity of the reflected light (47).

Biosensors: Diversity and Development

Generally, the biological components used in a biosensor construction can be divided into two categories: (a) those where the primary sensing event results from catalysis (such as whole microorganisms, tissue slices, organelles, and enzymes) and (b) those which depend on an essentially long-term binding of the target molecule (i.e., cell receptors, nucleic acids, and antibodies). The essential element, however, in making a successful biosensor is to provide a suitable immobilization procedure for biological compounds. Five main approaches to enzyme immobilization have been employed: (1) physical adsorption at a solid surface, (2) entrapment in polymeric gel or within microcapsules, (3) cross-linking by means of bifunctional reagents, often in combination with (1) and (2), (4) covalent binding to a reactive insoluble support, and (5) embedding in a lipid bilayer. This last-mentioned approach is unique in that embedding means that the compound(s) (membrane modifiers such as electron acceptors, donors, mediators, polypeptides, proteins, etc.) of interest in the lipid bilayer is relatively free to adapt to its surroundings. The functions of biomembranes are mediated by specific modifiers, which assume their active conformations *only* in the lipid bilayer environment. Furthermore, the presence of the lipid bilayer greatly reduces the background noise (interference). It also effectively excludes hydrophilic electroactive compounds from reaching the detecting surface that may cause undesired reactions. From the specificity, selectivity, speed of response, and design points of view, a supported planar bilayer lipid membrane (s-BLM) is an ideal natural environment for embedding a host of materials of interest. Hence, the s-BLM system offers a wider opportunity for the biosensor development (48). In the following we shall be mainly concerned with the electrochemical biosensors based on supported BLMs using electrical detection methods.

In practice, researchers continue to search for ideal combinations of biocatalysts—enzymes, antibodies—antigens, bacteria, plant slices isolated receptors, and even whole cells. In this context, genetic engineering technology will have a role in improving the biological component of enzyme-based and whole cell biosensors. Materials science and chemical engineering are instrumental in finding solutions for such problems as suitable immobilization procedures, transducers, and the effective coupling of the biological component to the transducer and the subsequent amplification system. The main types of transducers used in biosensor construction are listed in Table 2.

Table 2. Types of Transducers and Measurements Used in Biosensor Technology

Transducers	Measurement
Oxygen electrode	Electrochemical
Ion-selective electrode	Potentiometry
Modified metal electrode	Amperometry
Field-effect transistor	Conductimetry
Conductometry	Impedometry
Spectrophotometry	Photometric
Laser light scattering	
Optical fibers combine with absorption and fluorescence	
Surface plasmon resonance	
Thermistors	Thermometric
Piezoelectric crystal	Acoustic
Surface acoustic wave device	

Although in recent years a variety of different biological components have been used (46), it is by no means certain that all possible combinations of sensing elements and transducers have been explored.

The main challenge to the scientist in biosensor technology is to find an effective coupling of the biological component to the transducer. This is particularly important in the development of amperometric biosensors, since conventional “clean” metal electrodes are generally very poor voltammetric electrodes for the direct oxidation or reduction of biological components. A host of approaches to the modification of electrodes have been developed and investigated. All those methods, however, can be divided into two groups: (a) the modification of the electrode surface by depositing a monolayer, which is based upon either the adsorption of species at the electrode surface or a covalent attachment of redox mediators to the electrode, and (b) the modification by a multilayer, which is most frequently achieved by the use of polymeric modification of the electrode or sometimes by a vapor-phase deposition. The use of mediators in conjunction with oxidoreductases is not a recent innovation. Molecules such as quinones, organic and inorganic ions, and redox dyes have all been used. More recently, a novel electrode based on highly conducting metals and polymeric electron transfer mediators has been used. A practical mediator, however, should fulfill certain criteria. One of the most successful classes of electron mediators have been highly conjugated compounds such as ferrocene (bis-cyclopentadienyl iron), tetracyanoquinodimethane (TCNQ), tetrathiafulvalene (TTF), and their derivatives. However, very recently, avidin biotin coupling arose as a promising method for preparing enzyme-based biosensors (46,49).

Recent success in the interdisciplinary research in biology combined with electronics has led to exciting new developments based on enzymology and transducer techniques. They are known as enzyme electrodes, enzyme thermistor, CHEM-FET/ENFET devices, and immunosensor or enzyme transistor. Collectively, they are called “biosensors or biochips” (50). A common feature of all these devices is a close connection between the enzyme and the transducing system, which is used to follow the enzymatic reaction. The essential principle of the devices, broadly speaking, is predicated on the ligand–receptor contact interactions (51). Application of such developments in the fields of medicine, pharmaceuticals, biochemistry, environmental settings, robotics, and the food industry are obvious. For example, enzyme thermistors make use of the heat which is liberated during an enzymatic reaction. Their usual sensitivity is around 0.01°C. A recent modification of enzyme thermistor is the “TELISA” electrode, which achieves a sensitivity of about 10^{-13} M using an immunoabsorbent. It is expected that this measuring technique will find a broad application in continuous measurements of the release of hormones and/or antigens–antibodies in blood circulation. The quick attainment of a new steady state in the reaction occurring at an enzyme electrode after a random perturbation makes the latter ideally suited to monitor an industrial process (e.g., the production of antibodies). Classic calorimetric methods require much more time than an enzyme thermistor assay to perform a quantitative analysis. Two other interesting developments are ellipsometry and piezoelectric crystals. In ellipsometry, a close connection between the enzyme and the transducing device is not required. The method relies on the change in the angle of polarization of incident light which is reflected by a layer of

biomolecules bound to a solid surface. A change in the thickness and conformation of this layer, under the influence of other macromolecules interaction with the layer, can be easily monitored. This principle is now used in the fermentation industry. Piezoelectric crystals can be used in the analysis of the traces of certain compounds, mainly anesthetics. The frequency of the crystal depends strongly on the absence or presence of adsorbed molecules on the surface of the crystal (52). The selectivity of crystals toward a given compound may be increased by a coating process (e.g., with hydrophobic substances such as oils and fats).

Another exciting new research area is the combination of semiconductor technology with enzymes and other biological macromolecules. Here, mostly field-effect transistors (FETs) are used. If sensitivity of a FET toward certain chemicals or ions can be achieved, the prototype of an "ISFET" is born. A common feature of all these devices is the use of metal oxide semiconductor (MOS) structure. In combination with a thin layer of palladium, a high sensitivity toward gaseous hydrogen can be achieved. In this case, the membrane separates the gaseous from the liquid phase. Addition of traces of certain metals (e.g., Ir) to the Pd-MOS device also increased its sensitivity toward ammonia. It has been shown that such a device is capable of monitoring reliably the production of hydrogen by microorganisms (e.g., *Clostridium acetobutylicum*).

Lipid-Bilayer-Based Biosensors

The basic unit of all living organisms is the cell. Each cell is bound by a limiting plasma membrane, the fundamental structure of which is a lipid bilayer modified with proteins and carbohydrates. This is so because of the unique property of the lipid molecule: One end of the molecule is hydrophilic with a strong affinity toward a medium of high dielectric constant, while the other end of the molecule is hydrophobic, which is sequestered away from the aqueous solution. We now know that the remarkable stability of a lipid bilayer is due to the combination of hydrophilic and hydrophobic forces, which makes the lipid bilayer a thermodynamically favored structure (53).

The living cells undertake intercellular communication, which take place across as well as between cell membranes. The intercellular communication is mediated by the signal-transducing system, which is undertaken by various sensitive biomolecules (or sensor molecules) such as enzymes, ion channels, receptors, and carriers. These sensors *in vivo* serve as communicators to send, receive, transfer, and decode signal to adapt the cell to the changed environment. This ability of living cells is so-called biosensitivity. These signals are generally electrical and chemical in nature, predicated upon the presence of membrane. From the viewpoint of membrane biophysics and physiology, biological membranes are essentially the supporting matrix of the nature's sensors and devices (48,49), and the cell membrane plays a crucial role in signal transduction, energy conversion, and information processing. This is due to the fact that most physiological activities involve some kind of lipid bilayer-based ligand-receptor contact interactions. Outstanding examples among these are ion sensing, antigen-antibody binding, light conversion and detection, and gated channels, to name a few. For example, the thylakoid membrane of green plants functions as an energy transducer converting sunlight into electrical/chemical energy, the photoreceptor membrane

of a rod's outer segment detects photons as the initial step in visual perception, and the plasma membrane of cells and organelles possess the ability for ion sensing—for instance, differentiating Na^+ and K^+ with a great specificity. Furthermore, the plasma membrane provides sites for a host of ligand-receptor contact interactions such as antigen-antibody formation (47–49).

These outstanding characteristics have contributed to the discoveries of two artificial bilayer lipid membrane systems—namely, planar BLMs (7) and spherical liposomes (54)—in 1960s. The seminal work on the self-assembly of planar, bilayer or "black" lipid membranes (BLMs) was carried out by Rudin and his associates in 1959 to 1963. The idea started while one of the authors was reading a paperback edition of *Soap Bubbles* by C. V. Boys (55). These early researchers realized that a soap film in air in its final stage of thinning has a structure which may be depicted as two monolayers sandwiching an aqueous surfactant solution. Rudin and co-workers showed that an under water "soap bubble" (i.e., a BLM formed from brain extracts) was self-sealing and resistant to puncture and had many physical and chemical properties similar to those of biomembranes (7,47). Since then, techniques have been developed to incorporate a wide variety of compounds into BLMs to endow them with desired properties. As a result of the efforts of many investigators (56–59), biologically relevant phenomena such as ion selectivity, excitability, antibody-antigen reactions, active ion transport, and photoelectric effects have all been demonstrated.

In recent years, numerous attempts have been made to exploit the BLM system's potential in practical applications in sensor (46). Advances in microelectronics and interest in ultrathin organic films, including the BLM, especially the newly developed self-assembled bilayer lipid membrane (s-BLM) on a nascent metallic surface (60), have resulted in a unique fusion of ideas towards the development of intelligent biosensor and transducer. Furthermore, recent trends in interdisciplinary studies in chemistry, electronics, and biology have led to a new field of scientific-technological endeavor which is a part of a more general approach toward the development of a new, post-semiconductor electronic technology, namely, molecular electronics with a long-term goal of molecular computers (61).

SELF-ASSEMBLED LIPID BILAYER: TECHNOLOGY AND MEASUREMENT

Techniques of Planar BLMs and s-BLMs

The history of the BLM system and its development as a model for biomembranes has been recounted elsewhere (7). It should be mentioned that there is one major difference between Langmuir-Blodgett (L-B) film layers on rigid substrates and the BLMs. Apart from its bimolecular thickness, a BLM is a liquid-like, dynamic structure in a metastable state from a self-assembling point of view; it is envisioned that for long-term stability, a BLM separating two aqueous solutions requires the presence of a Plateau-Gibbs border. For biosensor development, it is our opinion that a fluid bilayer is of crucial importance. The aim of this section is to describe in sufficient detail how to set up a simple BLM system using the self-assembling techniques (see Table 1).

Conventional Planar BLMs. An experimental BLM, composed of either a mixture of 1% phosphatidylcholine and cholesterol in *n*-decane or oxidized cholesterol (ox. chol.) in *n*-octane, is formed in an aperture (diameter from 0.1 mm to 2 mm depending on the experimental requirement) punched in the side of a 10 ml Teflon cup which sits in an outer chamber of 20 ml volume. Before the introduction of the lipid solution, both chambers are filled with 0.1 M KCl. The lipid solution is then injected over the orifice of the Teflon cup in the bathing solution with a micro-syringe. The lipid solution should cover the opening over the entire orifice. One should see the light reflected from the thinning lipid film with iridescent color patterns. These colors gradually disappear as the membrane thins, and the membrane appears "black" when a BLM is finally formed. That is because such an ultrathin planar BLM possesses two interfaces and its thickness is less than 7 nm compared to the wavelength of visible light, and the membrane appears "black" because the light reflected from the front interface undergoes a half wavelength phase shift and interferes destructively with the light reflected from the back interface, which experiences essentially no phase shift. The BLM formation is monitored by the increase in membrane capacitance via a pair of Ag/AgCl electrodes. Electrical parameters of the BLM are measured with a high gain electrometer and a picoammeter. A number of simple setups for the BLM study have been published (7,62), and several comprehensive reviews on the subject are available (56–59).

Supported Bilayer Lipid Membranes (s-BLMs): The New BLM System. Since BLM's inception in 1960, many modifications have been made to the original BLM technique, with one notable exception to be described below, but the essential principle has remained the same. A BLM formed in the conventional manner (as described in the last section) is an extremely fragile structure with a limited lifetime. For long-term basic studies as well as for technological applications, the common concern has been the mechanical stability of the BLM. Although a number of improvements have been made to prolong the lifetime of the BLM, they rarely last longer than a few hours. For protracted studies and practical applications such as in biosensors and molecular electronic devices development, a long-lived BLM is a prerequisite. For this reason, it is both desirable and imperative that a method be found so that long-lasting BLMs can be generated. In 1989, a simple and novel method was reported for the formation of self-assembled BLMs on solid substrates, which possessed the desired dynamic properties and the requisite mechanical stability (51,60).

The principle and potentialities of the original method of the s-BLM probe construction are as follows. The two-step procedure consists of the adsorption of lipids from a BLM-forming solution on a nascent metal surface followed by the immersion of the lipid-coated tip into an aqueous solution for a BLM to self-assemble (60). This method is very simple, and it rivals the much more complicated and widely used Langmuir–Blodgett (L-B) film technique for certain biosensor fabrication (Table 1). Experimentally, s-BLMs on solid support have been formed by a number of methods, and the two-consecutive steps technique is carried out in the following manner: (1) Place a Teflon-coated metal wire [e.g., platinum (Pt) or stainless steel (ss)] and cut the tip of the wire with a sharp blade, while it is in contact with a BLM-forming solution. (2) Immerse the lipid layer that has adsorbed onto the metal wire tip into an aqueous solution for

5 min to 10 min for a BLM to self-assemble [see Fig. 2(a)]. For best cutting of the metal wire a miniature guillotine has been used, in which the sharp knife is moved vertically onto the wire placed on a flat surface and immersed in a lipid solution (60). Typically we used either an 1% glycerol dioleate in squalene or 5% phospholipid (PC or lecithin) in *n*-decane. For the probe construction we used Teflon-coated ss wire (0.05 cm diameter). Assuming the molecular area of PC to be $50 \text{ \AA}^2$, about 2×10^{14} molecules are needed to cover 1 cm^2 of metal surface with a BLM.

Alternatively, an s-BLM can also be formed at the end of a Teflon tubing filled with a hydrogel in KCl as a salt bridge used in electrochemistry. The formation procedure consists of three steps. In the first step, a chlorided Ag wire (Ag/AgCl) is inserted into the Teflon tubing which has been previously filled with a mixture of agar (or agarose) and KCl solution saturated with AgCl (0.3 g agar in 15 ml of 3 M KCl). The AgCl electrode and the filled Teflon tubing are glued together with wax at the point of insertion. In this way an Ag/AgCl–Teflon salt bridge is constructed. In the second step, the tip of the other end of the Teflon salt bridge is cut *in situ* while immersed in a BLM-forming solution with a scalpel. In the third and the last step, the Ag/AgCl–Teflon salt bridge with the freshly lipid solution coated tip is immersed in 0.1 M KCl solution of the cell chamber. The second step described above may be carried out in air and then the freshly cut end of the salt bridge is immediately immersed in the lipid solution for a few minutes. In either case, the cell chamber fills with an appropriate aqueous solution (e.g., 0.1 M KCl) containing an Ag/AgCl reference electrode and an Ag/AgCl–Teflon salt bridge with a self-assembled BLM at its end. The lead wires of the two electrodes are connected to the measuring instrumentation (see Fig. 3).

The precise arrangement and degree of ordering of the lipid molecules in the final structure is not known for certain. However, it seems highly probable that the bilayer nature of the assembly is a consequence of the thermodynamics of free-energy changes at the metal–lipid and lipid–aqueous solution interfaces.

Instrumentation. A number of methods that study the properties of BLMs such as optical, electrical, mechanical, transport, and permeability have been developed over the years (47,56,57,63). Of these methods, we shall describe only the electrical methods. In the last decade, many new electrochemical methods have been developed and applied to membrane research (64). Among these, the cyclic voltammetry (CV) turns out to be a very powerful method. The basics of CV consist of cycling the potential of a working electrode in an unstirred solution and measuring the resulting current. The potential of the working electrode is controlled relative to a reference electrode which is provided by a triangular potential waveform generator. The instrumentation used with BLMs can be much simpler than that used in the conventional CV, owing to the fact that the high resistance of BLMs can be studied with a two-electrode setup. Thus, a picoammeter together with a voltage waveform generator is all that is required. If a computer or an X–Y displaying device is available, the current–voltage (*I*–*V*) curves may be obtained, which are known as voltammograms. From such voltammograms, information about thermodynamic and kinetic parameters of the BLM system may be

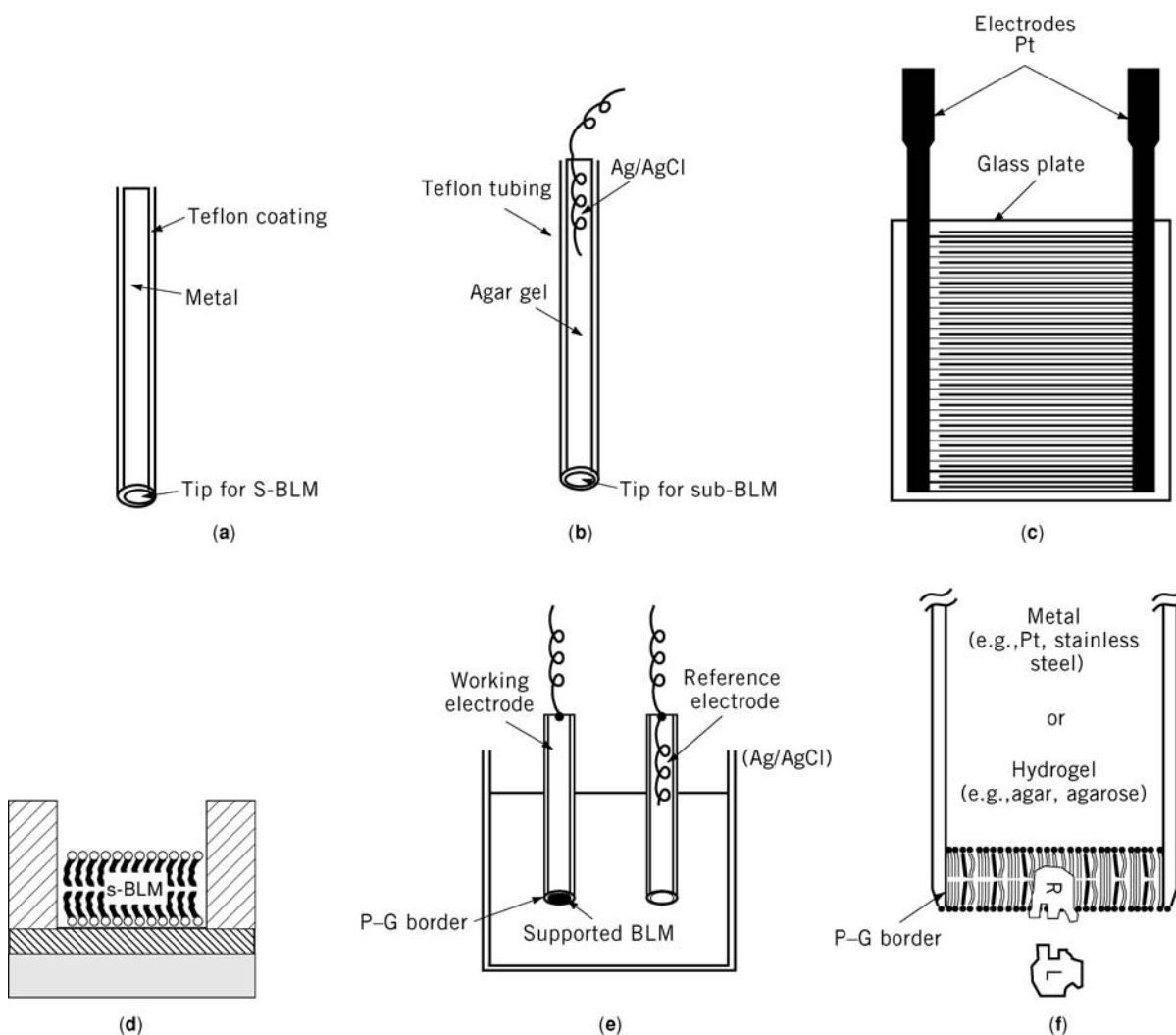


Figure 2. Schematic representations of (a) a metal-supported s-BLM probe, (b) a hydrogel-supported sb-BLM probe (78,93), (c) an interdigitated structure used for supporting an s-BLM (20), (d) a microsystem on a glass substrate for supporting s-BLMs (20), (e) a cell for electrochemical measurements (75), and (f) an enlarged view of the supported BLM illustrating the receptor (R)–ligand (L) interaction. P–G denotes Plateau–Gibbs border (49).

obtained, thereby providing insights into the mechanism of the membrane process under investigation.

A block diagram of the setup for obtaining cyclic voltammograms of the s-BLM system is shown in Fig. 3. The supporting metallic wire serves as a working electrode in a one-cell chamber [Fig. 2(e)]. The reference electrode, usually a chlorided Ag wire (Ag/AgCl) electrode, is dipped in the 0.1 M KCl solution placed in another chamber, and a salt bridge spans over the two chambers. For a two-electrode system as is usually used in the measurement, the newly cut tip of the metallic wire, coated with adsorbed lipids, acts as the working electrode. The current through the s-BLM is measured in the auxiliary electrode (made of a coil of Pt wire) during the CV scanning. The setup is housed in a Faraday cage to minimize interference by the external noise and electrical transients. Despite shielding, the external noise may still be picked up by the switch box; therefore, for the critical measurements the switch box should be incorporated within the same Faraday shield

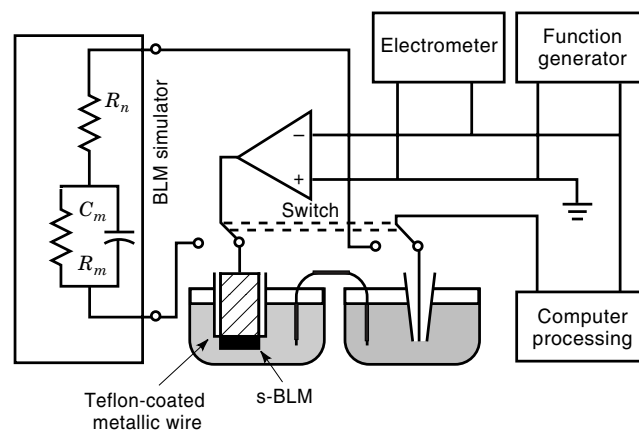


Figure 3. Block diagram of the setup for cyclic voltammetry on the s-BLM system.

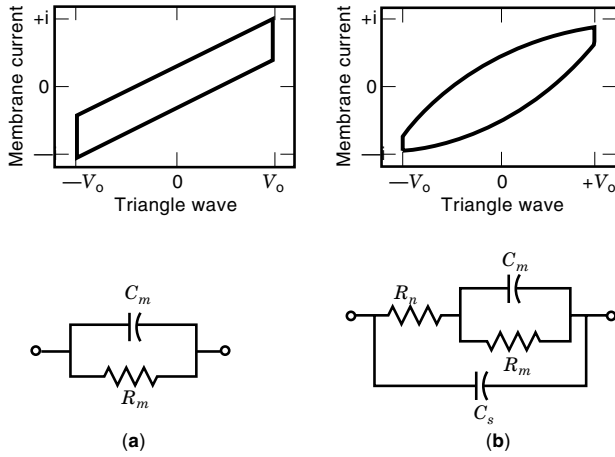


Figure 4. Cyclic voltammograms of BLMs and relevant equivalent circuits: (a) Voltammograms of the planar BLM system (top) and equivalent circuit (bottom). (b) voltammograms of the s-BLM (top) and the improved equivalent circuit (bottom).

as the cell. All cables used are shielded and the shields are grounded.

On-Line Measurement of BLMs from Voltammograms

Parameters Determination of Planar BLMs. The I - V response of an unmodified BLM has a form of parallelogram under the triangle sweeping wave. Figure 4(a) (top) is a typical voltammogram of conventional planar BLM system. A typical equivalent circuit of the planar BLM system is represented as illustrated in Fig. 4(a) (bottom) by a membrane resistance R_m in parallel with a membrane capacitance C_m . The parameters such as the capacitance, resistance, and membrane potential can be determined easily by analyzing these voltammograms. The triangular sweeping wave in the range of $\pm V_0$ with the scan rate A ($\text{mV}\cdot\text{s}^{-1}$) is the input from the circuit. The current in nanoamperes or picoamperes is measured. There are two components in the current through the membrane, namely the charging current i_c and resistance current i_r . The former is determined by the capacitance as follows:

$$i_c = C_m \frac{dV}{dt} = C_m A \quad (1)$$

It can be shown that the capacitance current i_c through the membrane capacitance is a constant. From Ohm's law, the latter component i_r of the membrane current is caused by the membrane resistance, that is,

$$i_r = \frac{V}{R_m} \quad (2)$$

So the net current passing through the BLM can be expressed as

$$i = i_r + i_c = \frac{V}{R_m} + C_m A \quad (3)$$

Equation 3 shows that the current through the resistor increases with increasing scan voltage. In the case of the constant

scan rate A , with fixed values of C_m and R_m , the current i has a linear relationship with the sweeping potential V . Thus the slope reflects the value of R_m , whereas C_m can be determined by measuring i_c according to the graph of the I - V response. i_c will jump to its negative values suddenly ($-i_c$) only at such points where the sweeping wave reaches its maximum and begins to reverse. The jump distance $2h$ equals $2i_c$, and thus C_m will be calculated by merely measuring h .

Parameters Determination of Supported BLMs. When the CV technique is applied to the solid-supported BLM, it has been found that the shape of I - V curve is quite different from that of a planar BLM (parallelogram). The typical voltammogram of an unmodified s-BLM shown in Fig. 4(b) (top) obviously has a different shape when compared with parallelograms obtained for the planar BLMs in Fig. 4(a) (top). The difference between them indicates that the equivalent circuit as well as the corresponding methods for the measurement of conventional BLM properties are no longer rigorously applicable to the s-BLM system. Measurement errors will have a great impact on the accuracy of the parameter determination unless the circuit is improved. Therefore it is necessary to improve the original equivalent circuit and to establish proper parameters relations using a set of differential equations so that the accurate determination of electrical properties of the s-BLM is assured (64,65).

A typical s-BLM system consists of a supported BLM as the working electrode, an electrolyte, and a reference electrode, which may be represented by a number of suitably connected resistors and capacitors such as R_m and C_m . According to the I - V response recorded on the real s-BLM in Fig. 4(b), the improved equivalent circuit for the s-BLM is developed in Fig. 4(b) (bottom), where the nonmembrane resistance R_n is introduced in series with the original circuit, and the parallel C_s is the distributing capacitance for the entire circuit. The membrane is characterized by a high resistance and a high capacitance. Their measured accuracies are far more affected by the proposed nonmembrane resistance R_n , whose composition and effect on the measurement error will be discussed in the following section.

From simulation we have obtained similar I - V responses by CV with the BLM simulator which contains relevant electrical circuit (see Fig. 3). The effect of R_n is achieved by a set of I - V responses generated by a BLM simulator with and without R_n as shown in Fig. 5, where the regular development of the CV waveforms are shown with varying values of R_m , C_m , R_n , and C_s . Figure 5(a) is a set of voltammograms in which C_m is varied while holding R_m constant and R_n at 0 Ω . From cycle 1 to 3, the intercept with the i axis rises gradually in correspondence with the increasing C_m . In Fig. 5(b) the slope drops down from curve 1 to curve 3 as R_m increases while keeping C_m constant.

Figure 5(c) is based on the improved equivalent circuit [see Fig. 3(b)], where both R_n and C_s are involved in the consideration. A detailed observation of the effect of R_n on the shape of the I - V cycle is carried out. In our experiment, R_n varies in a wide range from 10^0 to 10^9 Ω , while the other membrane properties are simulated by holding R_m at 10^9 Ω , C_m at 5 nF, both of which are still within the range of biomembrane. Curves 1 to 7 display the similarities of the I - V cycles which have the shape of parallelograms even if R_n reaches the value of 10^7 Ω . However,

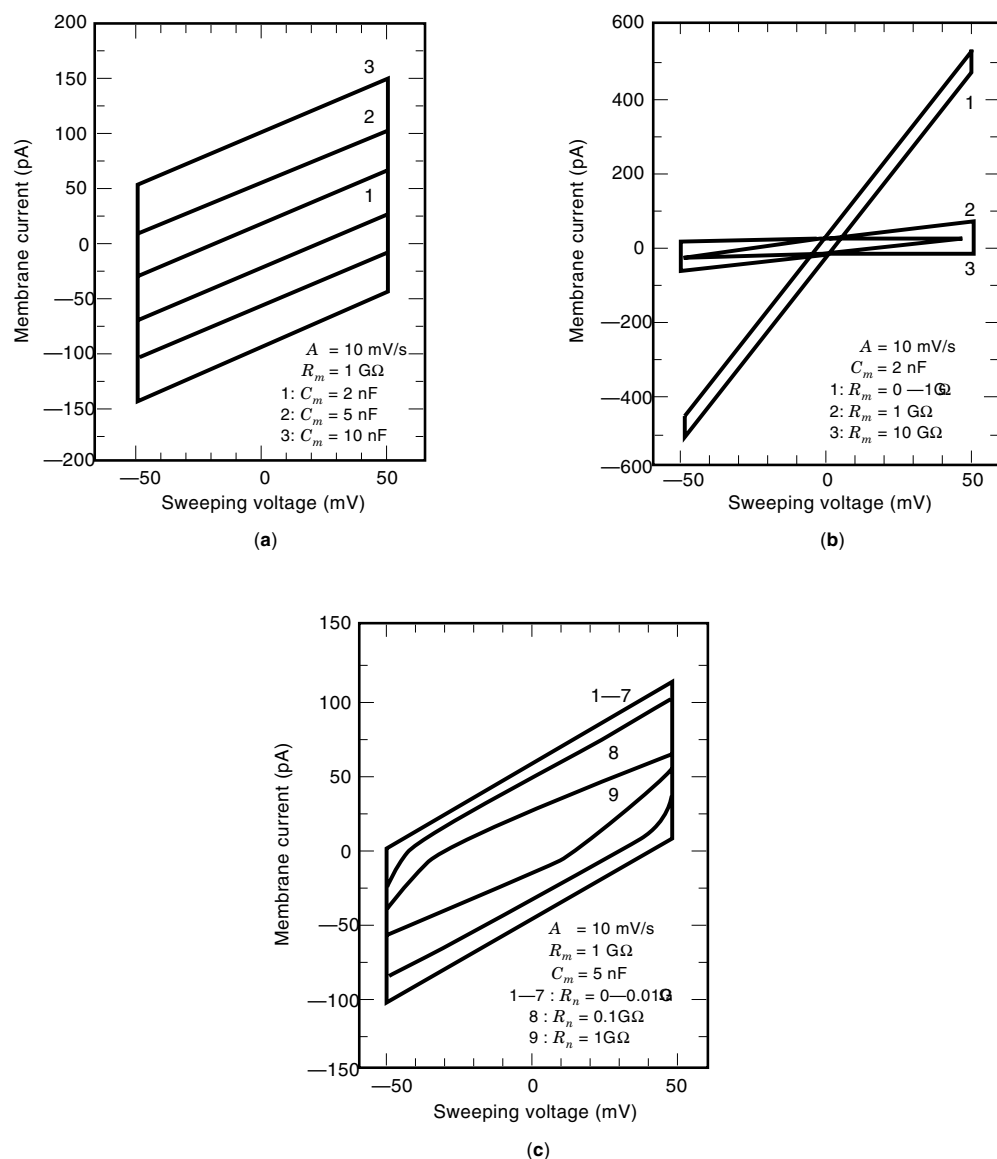


Figure 5. Voltammograms of the BLM simulator: (a) Voltammograms with different values of C_m ($C_m^1 < C_m^2 < C_m^3$) while remaining R_m constant; (b) voltammograms with different R_m ($R_m^1 < R_m^2 < R_m^3$), while remaining C_m constant; (c) affection of R_n on voltammograms (cycles 1–7, $R_n \approx 0-0.01 \text{ G}\Omega$; cycle 8, $R_n \approx 0.1 \text{ G}\Omega$; cycle 9, $R_n \approx 1 \text{ G}\Omega$).

when R_n increases to $10^8 \Omega$ ($\approx 0.1 R_m$), it begins to be characterized by charging current (curve 8). The intercept with the current axis is also somewhat decreased. The most important feature is that, when R_n reaches the same order as R_m ($\approx R_m$), the voltammogram (curve 9) does not display the shape of a parallelogram any longer, and it is very similar to that observed for the s-BLM system [see Fig. 4(b)]. The intercept with the i axis drops down sharply. The slope decreases during the potential sweeping, and it is much lower than the slope in the case of the BLM simulator without R_n or with a low value of R_n (cycles 1 to 7). In this last case, the intercept can no longer reflect the membrane capacitance, and it is the same as the slope of membrane resistance.

Therefore, the measurement error would be greatly increased if one still considers the intercept with the i -axis as the

membrane capacitance and the slope as the resistance. What causes the large value of R_n is still being discussed. The previous report using the LAPS technique has studied the system by considering the solution impedance effect (62). However, the value is not so high that it approaches R_m . Based on our previous research the nonmembrane resistance was not found so high in conventional planar BLM, and the technique for forming self-organized s-BLMs is based conceptually on interactions between a nascent metallic surface and amphipathic lipid molecules. It is supposed that an additional factor is owing to the presence of the metal-lipid interface (interfacial resistance) of R_n . Other important details in R_n include the electrode, electrolyte and solution impedance, and any other components which greatly effect the measurement accuracy.

Here, we propose a new method for determining R_m , C_m , C_s , and R_n accurately from the on-line s-BLM cyclic voltammogram (65). To make clear the effect of R_n on the I - V response, the quantitative relationship of the membrane current with the sweep wave potential has been derived rigorously from the solution of coupled equations based on the equivalent circuit shown in Fig. 4(b). The set of differential equations is as follows:

$$\begin{aligned} i &= i_r + i_c + i_s \\ i_r R_m &= \frac{1}{C_m} \int_0^t i_c dt \\ V &= (i_r + i_c) R_n + \frac{1}{C_m} \int_0^t i_c dt \\ i_s &= C_s A \end{aligned} \quad (4)$$

where i is the total current recorded, i_s is defined as the current through C_s , and all other parameters are defined above.

The final (definite) solution $i = f(V)$ is presented, where R_n and C_s have been included:

$$\begin{aligned} i &= \frac{AR_m C_m}{R_m + R_n} \left[\frac{V}{AR_m C_m} + \frac{R_m}{R_m + R_n} - \frac{2R_m}{R_m + R_n} \right. \\ &\quad \left. \exp \left(-\frac{(R_m + R_n)(V + V_0)}{AR_n R_m C_m} \right) \right] + C_s \cdot A \end{aligned} \quad (5)$$

The initial condition for Eq. (5) is

$$V_{t=0} = -V_0 + A \cdot t_{t=0} = -V_0 \quad (6)$$

During the half-cycle sweep from $-V_0$ to $+V_0$, we can see from Eq. (5) that there are three components included in membrane current: (a) the time-linearly dependent term, which is the resistance current obeying Ohm's law; (b) the time-exponentially dependent term in which R_n , R_m and C_m should be taken into account; and (c) the constant term resulting from the distributing capacitance C_s for the entire circuit.

The traditional way for the parameter determination is to vary each of the parameter values while holding all other parameters constant and to select a fitting curve with a minimal deviation. However, the complexity of the calculation is the major impediment in finding a group of effective suitable values to fit the calculated curve. Despite the aid of a computer, the calculation still requires much more time, which cannot be applied in the on-line parameter determination. So it is necessary to derive a set of accurate expressions from the original solution of Eq. (5), by which one-step calculation for determination of R_m , C_m , R_n , and C_s can be conducted.

Figure 6 presents this method to determine R_m , C_m , R_n , and C_s . First of all, the constant term $C_s A$ in Eq. (5) will be negative at the very moment when the sweep potential reaches its maximum and begins to reverse. So, the height $2h$ in Fig. 6 corresponds to twice the value of the distribution capacitance C_s multiplied with A . So from on-line computer acquire and parameters determination, the height $2h$ can be read out, and C_s

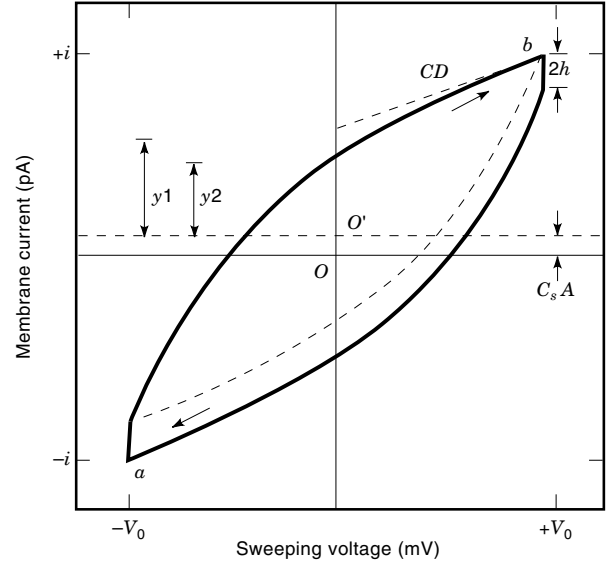


Figure 6. Method for on-line calculation of membrane parameters from s-BLM voltammogram. Four parameters are needed to be acquired from the on-line voltammogram: (1) $1/D$, the slope of the tangent line CD at the half-cycle terminal b ($+V_0$); (2) y_1 , the intercept of tangent line CD with the current i -axis; (3) y_2 , the intercept of the voltammogram with current i -axis; (4) $2h$, the transient current change at the time where the sweeping voltage reaches the maximum, and begins to reverse. Following Eq. (12) in the text, the membrane parameters can be calculated right away after a one-cycle voltammogram is acquired.

is determined by

$$C_s = h/A \quad (7)$$

The differentiation of Eq. (5) is given by

$$\frac{di}{dV} = \frac{1}{R_m + R_n} + \frac{2R_m}{R_n(R_m + R_n)} \exp \left(-\frac{(R_m + R_n)(V + V_0)}{AR_n R_m C_m} \right) \quad (8)$$

The potential of the triangular sweep wave moves from $-V_0$ to $+V_0$ (from a to b in Fig. 6). At the half-cycle terminal b , the exponential term can be rationally omitted and the slope of the tangent line CD at the point b can be expressed as

$$\frac{1}{D} = \frac{1}{R_m + R_n} \quad (9)$$

The intercept of CD with the current i -axis is y_1 , which is expressed as

$$y_1 = \frac{AR_m^2 C_m}{(R_m + R_n)^2} \quad (10)$$

The intercept of the voltammogram with the current i -axis is y_2 ; therefore

$$\Delta y = y_1 - y_2 = 2y_1 \exp \left(-\frac{R_m V_0}{y_1 D R_n} \right) \quad (11)$$

Thus, without any time-wasted fitting procedures, all the membrane parameters can be calculated cycle by cycle from the on-line data acquire as follows:

$$\begin{aligned} R_m &= \frac{m}{m+1} D \\ C_m &= \frac{y_1}{A} \left(\frac{m+1}{m} \right)^2 \\ R_n &= \frac{1}{m+1} D \\ C_s &= \frac{h}{A} \\ m &= \ln \left(\frac{2y_1}{y_1 - y_2} \right) y_1 \frac{D}{V_0} \end{aligned} \quad (12)$$

where D , y_1 , y_2 , and $2h$ are acquired from on-line voltammetry.

In conclusion, the only parameters that need to be acquired directly for a recorded voltammogram of the s-BLM are D , y_1 , y_2 , and $2h$. From the relationship described in Eq. (56), it is now possible to calculate accurate values of the membrane electrical parameters for the s-BLM system. Here a novel method is proposed which can determine the properties accurately without any iteration (65). So the real-time measurement is realized to be suitable for the dynamic analysis.

Solid-Supported BLM Formation. Similar to the time-resolved spectrometry through which the transient molecular reaction can be observed optically, the dynamics of modified or unmodified BLMs, or reconstituted systems using the BLM, is studied electrically by the CV technique where the capacitance, resistance, membrane potential, and current peak are the fundamental parameters in determining the static or the dynamic change of the BLM system. Among these studies, the formation of the s-BLM is a valuable one in analyzing the BLM mechanics and its electrochemical reactions.

Here, the membrane capacitance and resistance are chosen as two principal parameters in monitoring the formation of unmodified s-BLMs. The time-resolved capacitance and resistance are measured following the model described above, based on the recorded s-BLM voltammogram. Table 3 gives the values of s-BLM electrical properties under different BLM-forming solutions. The unmodified s-BLMs were formed from various lipid solutions: 0.5% and 2% lecithin PC (PC = phosphatidylcholine) in *n*-decane. The triangle potential wave sweeps within ± 100 mV, at sweeping rate of 50 mV/s. The setup, materials preparation and techniques for the formation of s-BLM have been described in the above section.

Dynamics of s-BLM Formation. The dynamics of the s-BLM should be studied to establish the membrane parameters of the s-BLM in stable stage (65). Figure 7 presents the time-resolved membrane resistance (a) and capacitance (b) during the s-BLM formation. Several successive stages are featured as shown (c),

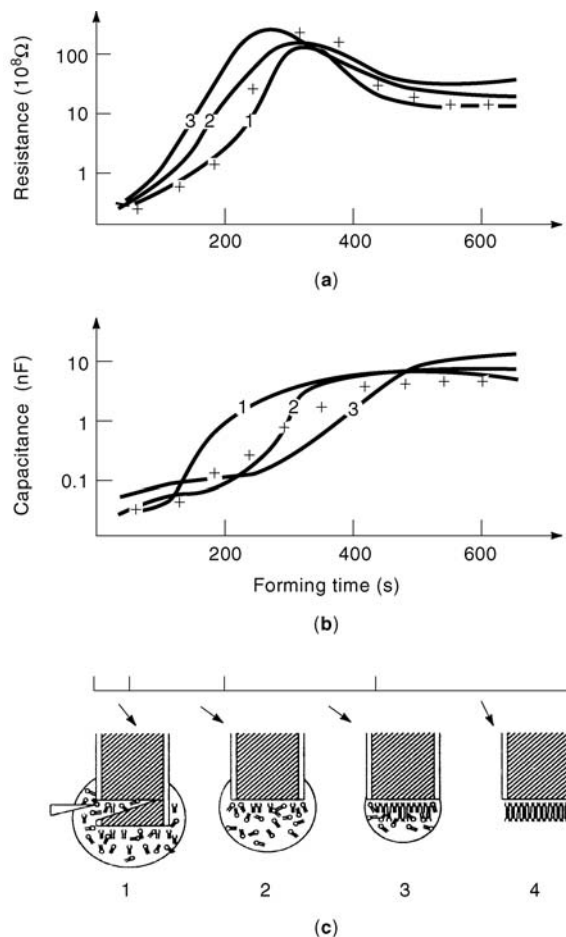


Figure 7. S-BLM formation dynamics: (a) Time-resolved membrane resistance during the formation process (curves 1, 2, and 3, continuous monitoring for 0.5%, 2%, and 10% lecithin solutions, respectively; “+” denotes point monitoring without CV potential effect); (b) Time-resolved membrane capacitance during the formation process (curves 1, 2, and 3, continuous monitoring for 0.5%, 2%, and 10% lecithin solutions, respectively; “+” denotes point monitoring without CV potential effect). (c) Characteristic stages of the s-BLM formation monitored by the membrane capacitance: (1) The tip of Teflon-coated platinum wire is cut off with a sharp blade while immersed in a lipid solution; (2) the newly cut surface of the wire attracts the polar groups of the lipid molecules, and thus a monolayer of lipid molecules is irreversibly adsorbed onto the tip of the wire, while the hydrophobic hydrocarbon chains are in contact with the lipid solution in air; (3) upon immersion of the wire into the aqueous solution, the lipid coating adhering to the metal surface will thin down spontaneously to a lipid bilayer, with the hydrocarbon chains of the two layers facing one another and the polar heads of the second layer of lipid molecules exposed to the water; (4) a self-assembled BLM adsorbed on a metal support has been formed.

according to the mechanics of s-BLM formation described in a previous report (65). In Fig. 7, the continuous curves

Table 3. Electrical Properties of Solid Supported BLM (Diameter: 0.1 mm)

BLM Solution	R_m (MΩ · cm ²)	C_m (μF cm ⁻²)	R_n (MΩ · cm ²)	C_s (pF)
Lecithin (0.5%) in <i>n</i> -decane	0.54	1.56	0.43	27
Lecithin (2%) in <i>n</i> -decane	1.74	0.63	0.40	31
Lecithin (10%) in <i>n</i> -decane	2.25	0.82	0.71	25

correspond to the forming solution of three different concentrations (curves 1, 2, and 3 for 0.5%, 2%, and 10% lecithin, respectively). The formation dynamics indicates that the higher membrane capacitance and resistance of static s-BLM (measured 600 s later) are obtained for the more concentrated solution. This phenomenon is in agreement with the fact that, as the concentration increases, the lipid molecules attached onto the unit area of solid surface (nascent surface of silver wire) become more dense (less solvent), and alter the dielectric constant and resistivity of the lipid bilayer.

Another important phenomenon is that the resistance as shown in Fig. 7(a) has the tendency shifting to the left with the increasing lipid concentration, while the capacitance [Fig. 7(b)] changes in the reverse direction (right shifting). With the denser packing, the lipid molecules quickly cover the freshly cut surface of the metallic wire, and less leak charge can go through the lipid-metal interface, as well as through the Plateau–Gibbs (P–G) border which is provided by the lipid solution between the Teflon coating and the metal wire. This causes the resistance rising in a shorter time for a more concentrated solution. But for the capacitance, though the denser lipid molecules block the current, their orderly parallel arrangement (perpendicular to the cut face of the metallic wire) becomes more difficult to reach the final state. This may be the reason for the capacitance “delay” in more concentrated lipid solutions.

The major principle of the CV technique is to impose an extra potential on the s-BLM in monitoring its dynamics. This influence is detected by the comparison of the “continuous” measurement and “point” measurement (also called nonimpact measurement). The point measurement is conducted through the single cyclic voltammogram recording at several predetermined time points while holding the s-BLM electrically isolated for a different period of time. In contrast, the continuous measurement is the cyclic recording without interruption during the entire formation. The result is reflected in Fig. 7(a, b) (cross-point curve 4) by the point monitoring for the 0.5% lecithin forming solution. The interval between two adjacent measurements is 60 s. The three continuous curves which have been described above are the result of continuous measurement. In Fig. 7(a) (compare curve 1 and the cross-point curve), little change is found regarding the effect of CV on the resistance, but the capacitance right shift in Fig. 7(b) (comparing curve 1 and the cross-point curve) is detected for point monitoring. The polar groups of lipid molecules are supposed to be charged for this behavior. In the initial stage of BLM formation, the lecithin PC polar molecules are in a random state. In the case where no CV potential is imposed on the BLM, the preferred self-organization is being completed only through the attraction between the highly hydrophilic nascent metal surface and the polar groups of the lipid molecules. However, the detecting cyclic potential wave arouses an additional alternative electrical field. The “in field” molecules are forced by the sweeping voltage to accelerate their orientation, which as a result shortens the time for the bilayer formation on the nascent metal surface.

According to the capacitance monitoring of the s-BLM formation (see Fig. 7), in general, there are about four characteristic stages that can be distinguished from the initial “cutting operation” to the final formation of the s-BLM. These stages are as follows:

1. At the beginning, the capacitance fluctuates at random for a few seconds because the tip of the Teflon-coated

metallic wire has just been cut off with a sharp blade while immersed in a lipid solution.

2. During the next few seconds, the capacitance is relatively low due to the droplet of the BLM forming solution on the newly cut surface of the metallic wire. Rapidly a monolayer of lipid molecules is irreversibly adsorbed onto it.
3. The capacitance tends to increase after upon immersion of the wire into the aqueous solution. The lipid droplet on the tip becomes thinner and leads spontaneously to form a lipid bilayer. Moreover, the potential of CV also tends to speed up the s-BLM formation.
4. As the s-BLM adsorbed on a metal support has been formed, the capacitance becomes stable, though sometimes it fluctuates slightly due to the transfer of the solvent and of the excess lipids to the aqueous phase as well as to the P–G border [see Fig. 2(f)]. In fact, all the capacitance values in Table 3 have been measured in this stage.

Self-Assembled Lipid-Bilayer-Based Biosensors. It should be pointed out that, first of all, unmodified lipid bilayers (i.e., BLMs formed from common phospholipids or oxidized cholesterol dissolved in *n*-octane) in 0.1 M KCl will typically have the following electrical properties: R_m greater than $10^8 \Omega \cdot \text{cm}^2$, C_m of about $0.4 \mu\text{F}/\text{cm}^2$, E_m about 0, V_b about $200 \pm 50 \text{ mV}$, and current–voltage (I – V) curves obeying Ohm’s law. However, the electrical properties of BLMs can be drastically altered by incorporating a host of materials such as pigments, dyes, polypeptides, membrane proteins, organic metals, and semiconductor particles (48,49).

Advances of Self-Assembled Lipid-Bilayer-Based Biosensors

There have been a number of reports on self-assembled molecules or structures described as advanced materials or smart materials. Without question, the inspiration for this exciting work comes from the biological world, where the lipid bilayer of cell membranes plays a pivotal role. Insofar as planar lipid bilayers are concerned, these are evidenced by self-assembled lipid bilayers, photoelectric effects in pigmented BLMs, and TCNQ-based BLM rectifiers, and they most recently supported BLMs on interdigitated structures as biosensors by microelectronics techniques (20,66). Our approach to materials science research is a biomimetic one and is centered on experimental BLMs in that the membranes can function in such important processes as electron transfer, signal transduction, and cellular environmental sensing (48,49). Specifically, the following experiments, some of which are in progress, are delineated below.

Incorporation of Ferrocene in s-BLMs. To test the versatility of s-BLMs as a “smart material,” an amperometric sensor was constructed for ferri-/ferrocyanide ions. The results have shown that (1) ferrocene can be very easily immobilized in the lipid bilayer on the tip of a metallic wire (s-BLM) system and (2) ferrocene in an s-BLM on a Pt support increases the sensitivity by about two orders of magnitude for potassium ferri-/ferrocyanide ions compared with that of a bare platinum electrode. This demonstrates that the s-BLM system offers a novel approach to the electrode modification by incorporation of compounds within a lipid bilayer (67–69).

Hydrogen Peroxide-Sensitive s-BLMs. The insertion of appropriate active molecules (modifiers) into the matrix of the lipid bilayer should be able to impart functional characteristics to s-BLMs. We chose TCNQ and dipyriddy-tetrathiafulvalene (DP-TTF) because of their properties as typical electron acceptor and donor molecules, respectively. It was found that DP-TTF not only improved the stability but also increased the range of s-BLM's sensitivity to hydrogen peroxide. In contrast, TCNQ-containing s-BLMs did not show much response to H_2O_2 (68).

Modified-BLMs as pH Sensors. The hydrated hydrogen ion (H_3O^+) is crucial to the functioning of cellular processes because it plays a leading role in enzyme catalysis and membrane transport. To test our concept, we incorporated a number of quinonoid compounds (chloranils) into s-BLMs and found that, indeed, s-BLMs containing either tetrachloro-*o*-benzoquinone (TC_oBQ) or tetrachloro-*p*-benzoquinone (TC_pBQ) responded to pH changes whose potential measurements exhibit a nearly theoretical slope (70,71). In this connection, mention should be made that s-BLMs formed from a glycerol dioleate (GDO) lipid solution containing polypyrrole have been found to be sensitive to pH with a Nernstian slope of 57 mV/decade hydrogen ion concentration as compared with unmodified s-BLMs (slope $\approx$ 20 mV). These new pH-sensitive s-BLMs offer prospects for a ligand-selective probe development using microelectronics technologies.

Modified s-BLMs as Ion Sensors. s-BLMs containing six different kinds of crown ethers were investigated using cyclic voltammetry. In particular, s-BLMs formed from a liquid crystalline aza-18-crown-6 ether and cholesterol-saturated *n*-heptane solution were found to be sensitive to K^+ in the concentration range 10^{-4} to 10^{-1} M with the theoretical Nernstian slope (72,73). The specificity for three alkali metal cations and NH_4^+ of these doped BLMs were also investigated. The order of specificity for most of these bis-crown ethers was found to follow hydrated radii of the cations, that is, $\text{NH}_4^+ > \text{K}^+ > \text{Na}^+ > \text{Li}^+$ (72). The results obtained with these s-BLMs compare favorably with conventional BLMs containing similar compounds, such as valinomycin and gramicidin (7). Concerning the gramicidin, it is well known that it forms ion channels. Cornell et al. (74) reported recently that a lipid-bilayer-based biosensor contains "sliding ion gates" made of gramicidin, and they claimed that a device can be designed for almost any analyte for which a receptor can be synthesized. The idea is again based on ligand-receptor interactions (48,49,69,75).

Recently we have investigated electron transfer across a BLM containing the so-called buckminsterfullerene C_{60} , which can act both as a mediator and a photosensitizer (see a later section on photoactive s-BLMs). Using the s-BLM, it is possible to construct a sensor probe sensitive to iodide ion and to investigate redox reactions across the lipid bilayer (76). The presence of C_{60} greatly strengthened the stability of the s-BLM and dramatically alters its electrical properties. Using the cyclic voltammetry technique, the results show that C_{60} embedded in the BLM acts as an excellent electron carrier/mediator and should be useful for electrochemical biosensor and molecular electronics device development. The cyclic voltammograms contain distinct redox peaks, which are not symmetrical. Further, the C_{60} in the lipid bilayer had the opposite effects on membrane resistance (R_m) and capacitance (C_m), respectively; it caused the R_m to decrease and C_m to increase. The presence of iodine

in conjunction with C_{60} in the s-BLMs further accentuated the effects. It is apparent that both C_{60} and iodine, when embedded in BLMs, facilitate electrical conduction, thereby lowering the R_m . Since the dielectric constant, ϵ , of a typical, unmodified BLM is about 2, the presence of C_{60} and iodine should exert a great influence on ϵ . Since C_m depends on a number of factors such as the surface charge of the BLM, the nature of hydrocarbon chains, and embedded modifiers, an increase in ϵ , and consequently in C_m , is therefore expected. The modified probe increased the detection limit for iodide ion by 100 times, with a linear response in the range from 10^{-8} M to 10^{-2} M. Thus, C_{60} facilitates the discharge of I_3^- at the metal surface, which demonstrates clearly that the embedded C_{60} is indeed an excellent electron mediator (76).

Modified s-BLM as Molecular Sensors. Many authors have reported sensors for the detection of glucose using glucose oxidase (77). Interestingly, using s-BLMs containing redox compounds and electron mediators but without the enzyme, glucose was detected in the buffered solution. The results are preliminary and further experiments are in progress. If a highly conjugated compound such as TCNQ is incorporated in the s-BLM forming solution, the resulting s-BLM was able to detect the presence of ascorbic acid, which is consistent with the findings obtained with conventional BLMs (78).

Molecular Recognition in an s-BLM. S-BLMs can be employed for embedding a number of compounds such as enzymes, antibodies, protein complexes (channels, receptors, membrane fragments or whole cells), ionophores, and redox species for the detection of their counterparts, respectively, such as substrates, antigens, hormones (or other ligands), ions, and electron donors or acceptors (7). The antigen-antibody reaction can be undertaken by using s-BLMs as the probe with electrical detection. The antigen, hepatitis B surface antigen (HBs-Ag) was incorporated into an s-BLM, which then interacted with its corresponding antibody (HBs-Ab or monoclonal antibody) in the bathing solution. This Ag-Ab interaction resulted in some dramatic changes in the electrical parameters (conductance, potential, and capacitance) of s-BLMs. The magnitude of these changes was directly related to the concentration of the antibody in the bathing solution (see Table 4). The linear response was very good from 1 to 50 ng ml^{-1} of antibody, demonstrating the potential use of such an Ag-Ab interaction via the s-BLM as a transducing device (79).

Electron Transfer Experiments in s-BLMs. Research in the field of electron transfer processes in BLMs was first conducted in

Table 4. Effect of Concentrations of HBs-Ab on the Resistance of s-BLMs

[HBs-Ab] (ng m^{-1})	$R_1(\times 10^8)$ ($\Omega \cdot \text{cm}^2$)	$R_2(\times 10^8)$ ($\Omega \cdot \text{cm}^2$)	$\log R_1/R_2$
100	6.05	0.429	3.15
80	6.14	0.504	3.09
60	6.04	1.63	2.59
50	5.07	2.43	2.35
40	5.15	5.50	2.01
30	5.55	8.80	1.80
20	6.30	28.1	1.35
10	6.30	51.2	1.09
1	6.50	162	0.604

the late 1960s to understand the primary step in natural photosynthesis (47,80). It was discovered that a light-driven electron transfer process between donor and acceptor species can occur across the thickness of a pigmented BLM. This finding has subsequently led to the view that the reaction center of natural photosynthesis functions similarly to that of a photovoltaic device of molecular dimensions (56). In the mid-1980s, electron transfer in the dark was seen in BLMs doped with either organic "metals" or semiconducting nanoparticles formed in situ (62). These phenomena were explained in terms of light-induced charge separation, field-driven charge transport, and subsequent redox reactions on opposite sides of the BLM. In the absence of light, the theory of electron tunneling was invoked (see above on TCNQ or TTF containing BLMs). When an s-BLM doped with Zn-phthalocyanine was excited by light, a voltage and a current were recorded, with the action spectra closely parallel to the absorption spectrum of the photon absorber (60). Thus we have shown that a pigmented s-BLM can function as a light transducer or photon-activated switch or detector (76,81).

S-BLMs Deposited on Piezoelectric Quartz Crystals. Smell and taste (olfaction and gustation) are among living organisms' two most vital sensing systems, the biophysics of which have been increasingly elucidated at the molecular level (52,66). Here again the crucial receptors are BLMs. In the preliminary experiments, several kinds of BLMs were successfully deposited on AT-cut quartz resonators (52,82). These were verified by observing frequency (f_m), potential (E_m), capacitance (C_m) and I - V curves. Frequency change (versus that in air) ranged from 9 kHz to 16 kHz, and in cyclic voltammetry experiments no redox peaks could be observed or the peaks were largely damped in the presence of $\text{Fe}(\text{CN})_6^{3-}$. E_m and C_m also showed characteristic values. But the exact values of these parameters were found to be related to the lipid solution, the pH of the bathing solution, and the scan time of voltammograms. If the BLM failed to form or was broken, obvious changes in these parameters were observed. In this case, f_m increased several kilohertz (to about 6 kHz, which corresponds to that induced only by pure viscous loading); C_m and E_m also increased and characteristic redox peaks were observed. Our findings show that BLMs can be formed on piezoelectric quartz crystals and that the piezoelectric techniques can be applied as a powerful tool to characterize the s-BLM system.

S-BLMs on Interdigitated Structures by Microelectronics Techniques. The fact that a lipid bilayer structure can be self-assembled on a solid substrate is intriguing. This novel manner of lipid bilayer formation overcomes two basic obstacles in the way of practical utilization of the BLM structure, namely: (a) its stability and (b) its compatibility with a standard microelectronics technology. As has been repeatedly demonstrated by us and others (20,83), the solid s-BLM system not only possesses the advantages of a conventional BLM structure, but gains new important properties, besides its long-term stability, such as (1) an anisotropic, highly ordered, yet very dynamic liquid-like structure and (2) two asymmetric interfaces, one of which is metallic. With this metallic connection, this type of probe solves the interfacing problem and is applicable to microelectronics technology. On this last-mentioned property, we have extended the experiment described above to the interdigitated structures (IDS). IDS are finger-like electrodes made by

microelectronics technologies and used in microchip applications (20). By forming s-BLMs on IDS made of platinum with a window of $0.5 \text{ mm} \times 0.5 \text{ mm}$, we obtained the following interesting results. First, when an IDS coated with a BLM was formed from asolectin, it responded to pH changes with only a $(15 \pm 2) \text{ mV/decade}$ slope. The conductance of s-BLMs on the IDS was about 50 times higher than the usual s-BLMs. Second, when an IDS coated with a BLM formed from asolectin was doped with either TCoOBQ or TCpBQ, the pH response was linear with a slope close to the theoretical value (68,70). This type of structure (i.e., s-BLM on interdigitated electrodes) can be used to investigate ligand-receptor contact interactions.

S-BLMs on an IDS can be manufactured using microelectronics technologies which already exist without the explicit need for special modification. This finding is viewed as a major "breakthrough" in the biosensor development. In this connection it should be mentioned that experiments on IDS chips modified with a BLM are based on a common basic aspiration—that is, to self-assemble a lipid bilayer containing membrane receptors, natural or synthetic, so that a host of physiological activities, such as ion-molecular recognition, can be investigated (20,52). At the molecular level, most of these activities may be termed collectively as the receptor-ligand contact interactions. The structures which are being reconstituted are inherently dynamic. Receptors and ligands in such close contact will normally vary as a function of time, frequently resulting in non-linear behavior. With an IDS chip modified with a BLM, we now have at last a most unique system for extensive experimentation which will be limited only by our imagination. Thus, insight gained from these studies will guide the preparation of functional BLMs on IDS support. Our aim is to take advantage of microelectronics techniques and apply them to the biochemical and neuroscience research.

Photoelectric Effects in Planar s-BLMs and sb-BLMs. Fullerenes (e.g., C_{60}) have been of great interest in materials science in the last decade. Our interest in fullerenes as a BLM modifier is owing to their most unique properties. Unmodified C_{60} , for example, is a water-insoluble and highly conjugated hydrophobic compound; it behaves as a good electron mediator and as an n -type semiconductor (bandgap = 1.6 eV). Hence, the lipid bilayer is an ideal environment for the compound to reside. The C_{60} -containing s-BLM, considered basically as a "molecular device," is a light-sensitive diode capable of photoinduced charge separation which undergoes redox reactions across the substrate-hydrophobic lipid bilayer-aqueous solution junctions.

Concerning photoactive compounds in BLMs, mention should be made about electron transfer processes in the membranes of photosynthesis and mitochondrial respiration, both of which have been reported in a number of BLM studies (47,53,56,59,80,84). In the 1980s, electron transfer was demonstrated in BLMs. The phenomena were explained in terms of the light-induced charge generation and separation, field-driven charge transport, and subsequent redox reactions on opposite sides of the BLM. In the absence of light, the theory of electron tunneling was invoked (56,64). When an s-BLM doped with Zn-phthalocyanine was excited by light, a voltage and a current were recorded, with the action spectrum paralleled closely to that of the absorption spectrum of the pigment (7,60). This provides direct evidence that the light-induced currents are indeed due to the photoactive compound embedded in the

BLM. Thus, a pigmented s-BLM can function as a light transducer or photon-activated switch or detector. In this connection, the experiments reported by Rolandi et al. (85), Yonezawa et al. (86), Lamrabte et al. (87), Yamaguchi and Nakanishi (88), Bi et al. (89), Ikematsu et al. (90), and more recently by Gruszecki et al. (91) should be of great interest.

Molecular Electronics and Lipid-Bilayer-Based Biosensors

Molecular electronics makes uses of materials at the molecular level in which the species retain their separate identities. As a result, the properties of such materials depend on the molecular arrangement, properties, and interactions. Theory seeks to guide the design and synthesis of effective molecular materials (57). It does so by analysis, interpretation, and prediction, leading to the development and evaluation of concepts, models, and techniques (66). The role of theory in treating molecular properties (mainly by molecular orbital methods) and arrangement (by electromagnetic or quantum-mechanical approaches) is of importance. When these factors are combined, the material properties can be treated more successfully in cases where the interactions are not essential for the existence—for example, in nonlinear optics as opposed to electronic transport properties (92).

The major advantage of molecular electronics with a lower limit on the order of micrometers is the further development of lithographic techniques. The changed physical properties in the submicroscopic region are the major obstacles to further miniaturization in the semiconductor technologies. The physical border for the silicon technologies is about 100 nm, because one cannot overcome the characteristic lengths such as diffusion, Debye, and tunnel lengths. With still smaller dimensions, we enter the realm of biological and molecular systems. Although biotransducers function much slower than silicon-based devices and are not very reliable, they are extremely efficient. Also, despite their disadvantages, nature's molecular devices function more generally and are superior to technical computers or sensors. In contrast to macromolecular biological systems, the main advantage of molecular devices, purportedly, is their relatively simple construction. In this sense, molecular devices may be readily constructed, and they are always easily accessible experimentally from a quantitative point of view (61).

The main elements of molecular electronics are molecular wire, conducting material, molecular-specific transducers of signals similar to the particles, and molecular switches, memories, emitters, detectors, and so on. The flux of information between the molecules can be released in many ways. One of the most important is the transfer of individual charges in terms of electrons, holes, or hydrogen ions, or of other shapes similar to the elements, like solitons, soliton waves, or excitons. Molecular switches may be optical, electrical, magnetic, or thermally reversible systems, such as photochroma-salicylidenanilin. Storage of information in a molecular system can be realized through a change in the electronic as well as geometric structures of the molecules in reversible thermal reactions—for example, conformational or configurational changes upon replacement of hydrogen or protons (66,92).

FUTURE PROSPECTS

In recent years, the development of biosensor configurations has been concentrated largely around the design of the trans-

ducer used. Further researcher's attention, however, should be focused on the mechanism of molecular recognition and catalysis. The fundamental properties of the device must be better understood in order to optimize critical factors such as response time, selectivity, and stability. Immobilization technologies and new membrane materials may basically change the present performance of biosensors.

The key advantage of molecular and biomolecular computing is specificity (92). The large number of variations that are possible with organic polymers such as poly (*N*-isopropylacrylamide as a "linking arm" in the protein ligand recognitions, and artificial photochrome and photorefractive material for future photon information storage and optical computer) allows for fine tuning of electronic motions to a much greater extent than is possible with organic materials. In biological molecules, certain configurational motions are comparable in significance to electronic motions. This is certainly the case in all conformation-based recognition processes. For example, the energy-loan model is applicable to many conformational switching processes based on the "lock-key"-type interactions of macromolecules. This could include self-assembly processes, protein folding, and various motions of biological macromolecules, which are generally attributed to fluctuation phenomena. The electronic instability in the energy-loan model may be thought of as a mechanism for amplifying the effects of fluctuations.

The switching process based on the energy-loan mechanism can mediate important forms of signal processing within biological cells. The enzymatic recognition is itself a basic form of information processing. When proteins and other macromolecules are combined into highly integrated complexes, they become possible for conformational switching processes to propagate over significant distances. The cytoskeleton is a good candidate for such a signal-processing network (92). One hypothesis is that conformational changes propagate through the cytoskeletal network by the energy-loan mechanism, with the conformation change of component macromolecules inducing the required lattice distortion in neighboring macromolecules. Such propagating conformation changes would play a role in transport processes in the cell, and it is likely that they would play a role in information processing as well.

One of the most delicate "molecular wires" is the so-called conjugated hydrocarbon chain, which is the best represented through the chain of carbon atoms in polymers. Most organic polymers are well-known insulators. However, polyacetylene (CH)_x, polydiacetylene, and polysulfinitride (SN)_x, with their conjugated double bonds, are semiconductors or superconductors. Such conjugated systems form the group of organic conductors and semiconductors. The most important organic electroactive polymer is polyacetylene. The foundation of the electron-hole pairs and the positive and negative charges are quite well known. In the outer electric field, the electron and the hole are accelerated in opposite directions. This properties can be used in optical switches for switching on and off the flux of information. Combining molecular "wire" and switchable molecules could lead to the construction of electronic systems based on molecules. Present research is oriented toward discovering peptides/proteins that can transduce electrical current or exist in two-electrical states. These could lead to future "biochips." Research on biochips could lead not only to a better understanding of higher nerve function, but also to the foundation of qualitative computer systems that could provide many of the activities currently performed only by the human brain.

Biochips can be considered, therefore, as highly sophisticated biosensors. The unique properties of biochips are their analog and digital computing potentials, self-perpetuating and potentially self-repairing. Biochips hold promise in a variety of applications such as bionic implants, memory-intensive systems, image processing and storage, artificial intelligence, language processing, and molecular computers. For instance, the analog capability of biochips could enable the creation of "artificial intelligence." As such, biochips are at a very early stage of research and development (66,92). As with biosensors, the current problem is our inability to produce uniform high-activity, stable biomolecular layers and their associated transduction systems. These problems notwithstanding, it seems likely that the initial application of biochips as advanced biosensors, based on the ligand-receptor contact interaction, may be in the clinical setting, where they could serve as automated control devices for drug delivery (61). It also appears probable that in order to extend the capabilities of present silicon-only systems, hybrid biochips and silicon-chip devices would be first produced for computing and memory-intensive systems. The key to the successful application of biochips will be to fill places that are not well-served by current silicon chip technology. Thus, the future development of biochips requires the successful technologies of stable biomolecule embedding and immobilization, biotransduction, and molecular lithography. Several urgent problems to be solved are biologically based amplification, molecular switching, electron transport, and memory function. In the coming decade, the answers to some of these problems will undoubtedly be found. In this connection, the development of lipid-bilayer-based sensors and biological electronic devices seems to be a logical first step. With the BLM systems, especially s-BLMs, we now have an experimental approach for testing some of the new ideas in the development of sensors for practical applications (8,20,46,74).

BIBLIOGRAPHY

- H. T. Tien and A. Ottova-Leitmannova, Membrane biophysics: As viewed from experimental bilayer lipid membranes (planar lipid bilayers and spherical liposomes), Amsterdam: Elsevier, 2000, pp. 648.
- J. M. Kauffmann (ed.), Special issue on electrochemical biosensors, *Bioelectrochem. Bioenerg.*, **41**: 1–108, 1997.
- Y. Umezawa et al. (eds.), Molecular recognition at liquid-liquid interfaces: fundamentals and analytical applications: *Special Issue of Analytical Sciences*, **14**: 1–245, 1998.
- M. A. Brett, and A. Oliveira-Brett (eds.), *Electroanalytical Chemistry, Special Issue of Electrochimica Acta*, **43**: 3587–3698, 1998.
- P. Krysiński, H. T. Tien, and A. Ottova, Charge-transfer processes and redox reactions in planar lipid monolayers and bilayers, *Biotechnology Progress*, **15**, (6): 974–990, 1999.
- A. D. Bangham, Surrogate cells or Trojan horses, *BioEssays*, **17**: 1081, 1995.
- A. L. Ottova and H. T. Tien, Bilayer lipid membranes: An experimental system for biomolecular electronic devices development, *Prog. Surf. Sci.*, **41** (4): 337, 1992.
- H. T. Tien and A. L. Ottova, From self-assembled bilayer lipid membranes BLMs to supported BLMs on metal and gel substrates to practical applications. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **149**, 1999, pp. 217–233.
- S. E. Feller and R. W. Pastor, Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities, *J. Chem. Phys.*, **111**: 1281–1287, 1999.
- T. Y. Tsong, Electroporation of cell-membranes, *Biophys. J.*, **60**: 297–306, 1991; *Bioelectrochem. Bioenerg.*, **24**, 1990.
- C. Karolis et al., Differential effects of cholesterol and oxidised-cholesterol in egg lecithin bilayers, *Biochimica et Biophysica Acta-Biomembranes*, **1368**: 247–255, 1998.
- R. F. Costello et al., Improved gel-protected bilayers, *Biosensors & Bioelectronics*, **14** (3): 265–271, 1999.
- L. Tung et al., Changes in electroporation thresholds of lipid membranes by surfactants and peptides, *Occupational Electrical Injury: An International Symposium*, **888**: 249–265, 1999.
- M. B. Partenskii, V. L. Dorman, and P. C. Jordan, Membrane stability under electrical stress: a non-local electroelastic treatment, *J. Chem. Phys.*, **109**: 10361–10366, 1999.
- M. Tarek et al., Molecular dynamics simulations of supported phospholipid/alkanethiol bilayers on a gold (111) surface, *Biophysical Journal*, **77** (2): 964–972, 1999.
- M. Tsionsky et al., Scanning electrochemical microscopy, Application of SECM to the study of charge transfer through bilayer lipid membranes, *Analytical Chem.*, **71**: 4300–4305, 1999.
- A. N. Parikh et al., Infrared spectroscopic characterization of lipid-alkylsiloxane hybrid bilayer membranes at oxide substrates, *Langmuir*, **15** (16): 5369–5381, 1999.
- D. Ivnitiski et al., Biosensors for detection of pathogenic bacteria, *Biosensors & Bioelectronics*, **14**: 599–624, 1999.
- D. Ivnitiski et al., Electrochemical biosensor based on supported planar lipid bilayers for fast detection of pathogenic bacteria. *Electrochem. Commun.*, **2** (7): 457–460, 2000.
- V. Tvarozek et al., Thin-film microsystem applicable in (bio)-chemical sensors, *Sens. Actuators: B. Chemical*, **19**: 597, 1994.
- V. Rehacek et al., Supported BLM microprobe with thin-film Ag/AgCl reference electrode for pH measurements, *Sensors & Materials*, **10** (4): 229–239, 1998.
- V. Tvarozek, Thin-film microprobe with bilayer lipid membrane for advanced pH-meter, in 21st Inter. Conf. on Microelectronics, organized by Yugoslavia IEEE Section, Electron Devices Chapter in cooperation with Faculty of Electronic Engineering, University of Nis and Ei Holding Co., Nis, under the technical co-sponsorship of IEEE Electron Devices Society, Proc. Conf., Nis-Yugoslavia 1997, 4 pages within the whole volume.
- E. Gallucci et al., Pore formation in lipid bilayer membranes made of phosphatidylinositol and oxidized cholesterol followed by means of alternating current, *Biophys. J.*, **71**: 824–831, 1996.
- B. Schuster et al., Probing the stability of S-layer-supported planar lipid membranes, *Euro. Biophys. J. Biophys. Lett.*, **28** (7): 583–590, 1999.
- Z. C. Bi et al., Photosensitization systems of covalently-linked phthalocyanine complexes in both bilayer-lipid membranes and tin oxide photovoltaic cell, *Photochemistry and Photobiology*, **59** (1): 111–115, 1994.
- C. Neumann-Spallart, F. Pittner, and T. Schalkhammer, Immobilization of active facilitated glucose transporters (GLUT-1) in supported biological membranes, *Applied Biochemistry and Biotechnology*, **68** (3): 153–169, 1997.
- L. Ding and E. K. Wang, Cyclic voltammetry of dye-modified supported bilayer lipid membranes (s-BLMs), *Bioelectrochemistry and Bioenergetics*, **43** (1): 173–176, 1997.
- C. Steinem et al., Impedance analysis of ion transport through gramicidin channels incorporated in solid supported lipid bilayers, *Bioelectrochemistry and Bioenergetics*, **42** (2): 213–220, 1997.
- W. M. Albers et al., Structural aspects of self-assembly of thienoviolon molecular conductors on gold substrates, *Thin Solid Films*, **330** (2) 114–119, 1998.
- J. Sabo et al., A combined AC-DC method for investigating supported BLMs, *Thin Solid Films*, **306**: 112–118, 1997.

31. K. Asaka, A. Ottova, and H. T. Tien, Mediated electron transfer across supported bilayer lipid membrane (s-BLM), *Thin Solid Films*, **354**: 201–207, 1999.
32. K. Asaka, H. T. Tien, and A. Ottova, Voltammetric study of charge transfer across supported bilayer lipid membrane (s-BLMs), *J. Biochem. Biophys. Methods*, **40**: 27–37, 1999.
33. J. Feng et al., Photoelectric measurements of self-assembled and supported planar lipid bilayers: A new technique for studying apoptosis, *Electrochemistry Communications*, **1**: 145–147, 1999.
34. Y.-F. Zhang et al., Study on selectivity of permeating chiral complexes in bilayer lipid membranes (BLMs), *Electrochemistry Communications*, **1**: 238–241, 1999.
35. T. Nakanishi et al., Aqueous electrochemistry of a C-60-bearing artificial lipid bilayer membrane film immobilized on an electrode surface: Thermodynamics for the binding of tetraalkylammonium ion to the fullerene anion, *J. Phys. Chem. B*, **103**: 304–308, 1999.
36. G. Lefevre, J. Bessiere, and A. Walcarius, Cuprite-modified electrode for the detection of iodide species, *Sensors and Actuators B-Chemical*, **59** (2-3): 113–117, 1999.
37. S. X. Xia et al., Copper(II) ion-sensitive electrode using chemical vapor deposited Nb-Ge superconducting films, *Analytica Chimica Acta*, **382** (1-2): 151–156, 1999.
38. L. Wang et al., Receptor-ligand interactions in a reconstituted bilayer lipid membrane, *Electrochem. Commun.*, **2**: 287–289, 2000.
39. P. Kryszinski, M. R. Moncelli, and F. Tadini-Buoninsegni, A voltammetric study of monolayers and bilayers self-assembled on metal electrodes, *Electrochim. Acta*, **45** (12): 1885–1892, 2000.
40. D. P. Nikolelis and S. Pantoulas, A minisensor for the rapid screening of acesulfame-K, cyclamate, and saccharin based on surface-stabilized bilayer lipid membranes, *Electroanalysis*, **12**: 786–790, 2000.
41. Y. L. Zhang et al., Salt-bridge supported bilayer lipid membrane modified with Calix[n]arenes as alkali cation sensors, *Analytical Lett.*, **33** (5): 831–845, 2000.
42. Z. Y. Wu et al., Amperometric glucose biosensor based on lipid film, *Biosensors & Bioelectronics*, **15** (3-4): pp. 143–147, 2000.
43. S. Kalinowski, Z. Lotowski, and J. W. Morzycki, The influence of bolaamphiphilic steroid dimer on the formation and structure of bilayer lipid membranes, *Cellular & Mol. Biol. Lett.* **5** (1): 107–118, 2000.
44. R. M. Nyquist et al., Characterization of self-assembled monolayers for biosensor applications, *Langmuir*, **16**: 1793–1800, 2000.
45. Y. L. Zhang, C. X. Zhang, and H. X. Shen, Studies on electrocatalytically kinetic behavior of horseradish peroxidase at salt bridge supported bilayer lipid membrane, *Analytical Letters*, **33** (12): 2425–2439, 2000.
46. R. F. Taylor and J. S. Schultz (eds.), *Handbook of Chemical and Biological Sensors*, Philadelphia: IOP Publishing, 1996.
47. H. T. Tien, *Bilayer Lipid Membranes BLM: Theory and Practice*, New York: Marcel Dekker, 1974.
48. H. T. Tien, Self-assembled lipid bilayers for biosensors and molecular electronic devices, *Adv. Mater.*, **2**: 316–318, 1990.
49. H. T. Tien, S. H. Wurster, and A. L. Ottova, Electrochemistry of supported bilayer lipid membranes: Background and techniques for biosensor development, *Bioelectrochem. Bioenerg.*, **42**: 77–94, 1997.
50. F. T. Hong (ed.), *Molecular Electronics: Biosensors and Biocomputers*, New York: Plenum, 1989.
51. K. L. Mittal and D. O. Shah (eds.), *Surfactants in Solution*, New York: Plenum, 1989, vol. 8, pp. 133–178; 1992, vol. 11, pp. 61–80.
52. A. L. Ottova et al., Self-assembled and supported BLMs as an adaptive material for biotechnology, *SPIE's 1996 Symp. Smart Structures Mater.*, 1997, vol. 2716, pp. 144–151.
53. G. Dryhurst and K. Niki (eds.), *Redox Chemistry and Interfacial Behavior of Biological Molecules*, New York: Plenum, 1988, pp. 529–556.
54. M. Rosoff (ed.), *Vesicles*, New York: Marcel Dekker, 1996, pp. 49–77.
55. C. V. Boys, *Soap Bubbles: Their Colours and the Forces Which Mould Them*, New York: Dover, 1959.
56. S. G. Davison (ed.), *Prog. Surf. Sci.*, **23**: 317, 1986; **30**: 1, 1989.
57. I. B. Ivanov (ed.), *Thin Liquid Films: Fundamentals and Applications*, New York: Dekker, 1988.
58. J. R. Harris and A.-H. Etemadi (eds.), *Artificial and Reconstituted Membrane Systems, Subcellular Biochemistry*, Vol. 14, New York: Plenum, 1989, p. 97.
59. M. Blank (ed.), *Biomembrane Electrochemistry, Advances in Chemistry Series*, No. 235, Washington, DC: ACS, 1994.
60. A. L. Ottova et al., Self-assembling bilayer lipid membranes on solid support: Building blocks of future biosensors and molecular devices, in R. R. Birge (ed.), *Molecular Electronics and Bioelectronics, Advances in Chemistry Series*, No. 240, Washington, DC: ACS, 1994.
61. J. R. Bourne (ed.), Lipid bilayer-based sensors and biomolecular electronics, *Biomed. Eng.*, **18** (5): 323, 1991.
62. J. Kutnik and H. T. Tien, Application of voltammetric techniques to membrane studies, *J. Biochem. Biophys. Met.*, **11**: 317–326, 1985; J. Kutnik and H. T. Tien, Apparatus and techniques for measurement of electrical properties of bilayer lipid membranes BLM, *J. Electrophysiol. Tech.*, **14**: 211–221, 1987.
63. S. Kalinowski and Z. Figaszewski, A four-electrode potentiostat-galvanostat for studies of bilayer lipid membranes, *Meas. Sci. Technol.*, **6**: 1043–1055, 1995.
64. H. T. Tien, Cyclic voltammetry of electron-conducting bilayer lipid membranes, *Bioelectrochem. Bioenerg.*, **13**: 299–316, 1984; **15**: 19–38, 1986.
65. L. Q. Gu et al., A new method for the determination of electrical properties of supported bilayer lipid membranes by cyclic voltammetry, *Bioelectrochem. Bioenerg.*, **39**: 275–283, 1996.
66. H. T. Tien et al., Bilayer lipid membrane (BLM): An experimental system for biomolecular electronic device development, *J. Mol. Electron.*, **4**: s1, 1988.
67. H. T. Tien et al., *Anal. Lett.*, **26**: 819, 1993.
68. W. Liu et al., in R. L. Opila et al. (eds.), *Polymer/Inorganic Interfaces*, 304, Pittsburgh, PA: Materials Research Society, 1993.
69. A. Leitmannova-Ottova and H. T. Tien, Receptor–Ligand Contact Interactions Using Self-Assembled Bilayer Lipid Membranes, in E. P. George et al. (eds.), *Materials for Smart Systems, The MRS Symp. Proc. Series, reviewed*, vol. 360, Pittsburgh, PA: Material Research Society, 1995.
70. W. Ziegler et al., in V. Tvarozek and S. Nemeth (eds.), *Proc. 7th C-S Conf. Thin Films*, pp. 304–311.
71. A. L. Ottova, D.-L. Guo, and H. T. Tien, *Proc. Conf. Smart Structures Mater.*, Orlando, FL, 1995.
72. (a) Z. H. Tai, L. Cun, and C. Z. Yang, Pt supported BLMs modified by ferrocene and its derivatives, *Mol. Eng.*, **2**: 215–220, 1994. (b) Z.-H. Tai et al., Molecular assemblies of disazafluorenone Schiff-base amphiphiles: Bilayer membrane and its electrochemical oscillations, *Mol. Eng.*, **3**: 285–291, 1994.
73. Y.-E. He et al., Crown ether-modified bilayer lipid membranes on solid support as ion sensors, *Analytical Lett.*, **28** (3): 443, 1995.
74. B. A. Cornell et al., A biosensor that uses ion-channel switches, *Nature*, **387**: 580–583, 1997; B. Raguse et al. Tethered bilayer lipid membranes: Formation and ionic reservoir characterization, *LANGMUIR* **14**: 648–659, 1998.

75. A. L. Ottova et al., Supported BLMs: Biomembrane models and biosensor applications, *Supramol. Sci.*, **4**: 101, 1997.
76. H. T. Tien et al., Electronic processes in supported bilayer lipid membranes (s-BLMs) containing a geodesic form of carbon (fullerene C₆₀), *Bioelectrochem. Bioenerg.*, **42**: 161, 1997.
77. J. Kotowski, T. Janas, and H. T. Tien, Immobilization of glucose oxidase on a polypyrrole–lecithin bilayer lipid membrane, *J. Electroanal. Chem.*, **253**: 277, 1988.
78. X.-D. Lu, A. L. Ottova, and H. T. Tien, Biophysical aspects of agarose supported bilayer lipid membranes BLMs: A new method for forming and studying planar BLMs, *Bioelectrochem. Bioenerg.*, **39**: 285–289, 1996.
79. L. G. Wang, Y.-H. Li, and H. T. Tien, Electrochemical transduction of an immunological reaction via s-BLMs, *Bioelectrochem. Bioenerg.*, **36**: 145, 1995.
80. P. Bianco and J. Haladjian, Control of the electron transfer reactions between c-type cytochromes and lipid-modified electrodes, *Electrochem. Acta*, **39**: 911–916, 1994.
81. H.-M. Wang et al., *Bull. Electrochem.*, **12**: 496, 1996.
82. A. L. Ottova et al., in M. Alper, et al., (eds.), Biomolecular materials by design, *Proc. Mater. Res. Soc.*, Fall Meeting, 1993, p. 330.
83. A.-N. Sun et al., Research on electrical properties of amphiphilic lipid membranes by means of interdigital electrodes, *Mater. Sci. Eng. C*, **2**: 159–163, 1995.
84. L. Song, Y.-J. Xiao, and X.-X. Gao, A simple method to form an asymmetric BLM and its enhancement of photoelectric responses, *Langmuir*, **12**: 4303–4304, 1996.
85. R. Rolandi, D. Ricci, and O. Brandt, *J. Phys. Chem.*, **96**: 6783, 1992.
86. Y. Yonezawa, H. Fujiwara, and T. Sato, *Thin Solid Films*, **210/211**: 736, 1992.
87. A. Lamrabte et al., *Bioelectrochem. Bioenerg.*, **27**: 449, 1992.
88. H. Yamaguchi and H. Nakanishi, Photoresponsive planar bilayer lipid membranes containing azobenzene amphiphilic derivatives, *Sens. Actuators B*, **13/14**: 376, 1993.
89. Z.-C. Bi et al., *J. Photochem. Photobiol. A: Chem.*, **77**: 37, 1994.
90. M. Ikematsu et al., *J. Electrochem.*, **403**: 61, 1996.
91. W. I. Gruszecki, A. Wardak, and W. Maksymiec, *Photochem. Photobiol. B Biol.*, **39**: 265, 1997.
92. F. L. Carter, R. E. Siatkowski, and H. Wohltjen (eds.), *Molecular Electronic Devices*, Amsterdam: North-Holland, 1988.
93. H.-P. Yuan, A. Ottova-Leitmannova, and H. T. Tien, An agarose stabilized BLM: A new method for forming bilayer lipid membranes, *Mater. Sci. Eng.*, **C4**: 35–38, 1996.
94. H. T. Tien, BLM-based electrochemical biosensors, *J. Clin. Lab. Anal.*, **2**: 256–264, 1988.
95. H. Yamada et al., Electron-transfer at a planar bilayer-lipid membrane incorporated with 7,7,8,8-tetracyanoquinodimethane studied by ac-impedance spectroscopy, *J. of Physical Chem.*, **97**: 9547–9549, 1993.
96. A. Winter et al., Gangliosides in phospholipid bilayer membranes: interaction with tetanus toxin, *Chem. Phys. Lipids*, **81**: 21–34, 1996.
97. N. Nakashima et al., A C-60-embedded artificial bilayer membrane film electrode device: Phase-transition-dependent electrochemistry, *J. Phys. Chem. B*, **102** (38): 7328–7330, 1998.
98. K. Winkler, D. A. Costa, and A. L. Balch, Electrochemical properties of fullerene based solid thin layers, *Polish J. of Chem.*, **74** (1): 1–37, 2000.
99. B. Wetzter et al., *Langmuir*, **14** (24): 6899–6906, 1998.
100. H. Sato, et al., Ionophore incorporated bilayer lipid membranes that selectively respond to metal ions and induce membrane permeability changes, *Biosensors & Bioelectronics*, **13**: 1035–1046, 1998.
101. C. G. Siontorou, et al., Rapid methods for detection of Aflatoxin M-1 based on electrochemical transduction by self-assembled metal-supported bilayer lipid membranes (s-BLMs) and on interferences with transduction of DNA hybridization, *Electrochimica Acta*, **43**: 3611–3617, 1998.
102. P. Diao, et al., Cyclic voltammetry and ac impedance studies of Ca²⁺-induced ion channels on Pt-BLM, *Bioelectrochemistry and Bioenergetics*, **45** (2): 173–179, 1998.
103. F. T. Hong, Interfacial photochemistry of retinal proteins, *Prog. Surf. Sci.*, **62**: 1–237, 1999.
104. D. L. Jiang et al., Photoinduced electron transfer across a gold supported octadecanethiol/phosphatidylcholine hybrid bilayer membrane mediated by C-60 in different redox species solution, *J. Photochem. Photobiol. A-Chem.*, **132**: 219–224, 2000.
105. T. I. Rokitskaya et al., Photosensitizer binding to lipid bilayers as a precondition for the photoinactivation of membrane channels, *Biophys. J.*, **78**: 2572–2580, 2000.
106. M. Seitz et al., Formation of tethered supported bilayers by vesicle fusion onto lipopolymer monolayers promoted by osmotic stress, *Langmuir*, **16** (14): 6067–6070, 2000. For more information, see our updated web pages.

Addendum (Some Recent Experiments Using Conventional BLMs and Supported BLMs)

A number of pertinent publications using conventional BLMs and supported BLMs have appeared. The incorporation of these findings into the text is not feasible at the present time. Therefore, a bibliography (annotated Refs. (1,93–105)) is added in this survey volume, so that the literature is up to date. The interested reader may wish to visit our WebPages at:

<http://www.msu.edu/user/ottova/membrane.biophysics.html>
http://www.msu.edu/user/ottova/soup_bubble.html

LI-QUN GU
 Texas A & M University
 ANGELICA LEITMANNOVA-OTTOVA
 Slovak Technical University
 H. TI TIEN
 Michigan State University

SIGNAL CONDITIONING

INSTRUMENTATION AMPLIFIERS

BACKGROUND

The monolithic operational amplifier (op-amp) has, in recent years, surpassed the basic discrete transistor as an analog building block in almost all instrumentation applications. While there are still many application areas (particularly very high speed, high voltage, high current, and ultra low noise) in which the discrete transistor (or even the vacuum tube) predominates, the ease of application of the op-amp has vastly simplified analog design at the system level, at least at frequencies below 100 MHz or so. The origins of the op-amp are buried deep in early negative feedback techniques, but suffice it to say here that the term “operational amplifier” appears to have been coined in a paper by Ragazzini and his colleagues (1), although the term “computing amplifier” survived for a while. The op-amp consisted of a high-gain inverting amplifier that remained stable with 100% negative feedback. This enabled highly controlled closed-loop functions (particularly integrators) to be realized, which, in turn, formed the basis for many analog computing functions. Later op-amps included a noninverting input, modifying the transfer function to one of high gain appearing *differentially* between the two inputs.

Of course, differential amplifiers were not new; simple long-tailed-pair amplifiers with controlled gain date far back to the early days of electronics. But the op-amp was different in that the gain control came from purely external components (assuming high enough open-loop gain), potentially enabling precision gain control from purely passive components.

However, herein lies the dilemma: When feedback is applied around an op-amp, the inverting input becomes a low (ideally zero) impedance. The noninverting input, though, remains at high impedance. Many instrumentation (and other) systems require a precise differential amplifier with high (relative to signal source) impedance for *both* inputs, and this function is now known as an *instrumentation amplifier*. Obviously, simple op-amps require considerable modification to fulfill this function.

DEFINITION

An *instrumentation amplifier* is a precision amplifier with single-ended output and differential inputs and with precisely controlled gain for voltages appearing between its inputs. Ideally, voltages common to both inputs should not affect the output (this will lead to a discussion concerning common-mode rejection in due course). Additionally, both inputs are expected to have high impedance (relative to the source impedance), and this is normally expected to be symmetric, at least to a first order.

Unlike the op-amp, the origins of the term “instrumentation amplifier” are somewhat nebulous. Some early monolithic precision op-amps, notably the μ A725 from Fairchild Semiconductor (1969) and the OP-07 from Precision Monolithics (1975), were referred to as instrumentation operational amplifiers.

This was undoubtedly a marketing label intended to emphasize their precision input characteristics, necessary in many instrumentation applications; but they were still op-amps and not instrumentation amplifiers in the sense now generally accepted. The term “differential amplifier” has its origins lost to obscurity, but such terms as “high-accuracy differential amplifier” or “direct-coupled differential amplifier” appear regularly in texts prior to 1960, and these, of course, are terms that basically describe the quintessential instrumentation amplifier. In the mid-1960s the term *data amplifier* became commonplace to describe such amplifiers (2,3), and several companies were producing self-contained modular amplifiers at this time. The now commonly used term *instrumentation amplifier* was certainly in use by 1967 (4), and the two terms were concurrent for a while.

THE BASIC INSTRUMENTATION AMPLIFIER

Possibly the simplest way to achieve the instrumentation amplifier function is with a single op-amp configured for equal inverting and noninverting gains.

Referring to Fig. 1, the amplifier produces a gain between the inverting input (–IN) and the output, which can be derived from classical op-amp theory as follows:

$$G_{-} = -\frac{R_4}{R_3}$$

The gain from the noninverting input (+IN) to the output can be described by

$$G_{+} = \left(\frac{R_2}{R_1 + R_2} \right) \left(\frac{R_3 + R_4}{R_3} \right)$$

So if the ratio R_1/R_2 is made identical to the ratio R_3/R_4 , the resulting output gain is identical for both inputs but of opposite sign (assuming, for the moment, that A1 is ideal and neglecting any source impedance). This is tantamount to saying that the *common-mode* gain is zero. The absolute value of gain from either input to the output with respect to

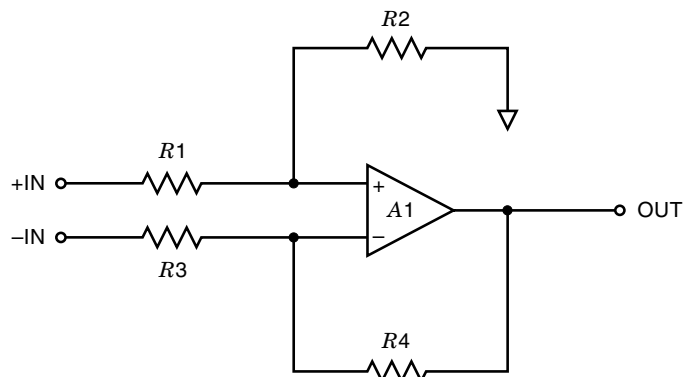


Figure 1. Basic instrumentation amplifier.

the other input is still given by

$$G = \frac{R4}{R3}$$

and this is the *differential-mode* gain. For simplicity, the term “gain” when used without qualification, will subsequently be assumed to be the differential-mode value. This circuit thus performs the basic functions of an instrumentation amplifier.

Although this circuit is commonly used where economy is a major concern, as will be shown, it lacks many of the features demanded in modern instrumentation systems. It is, however, a good starting point to illustrate some of the problems encountered in the design of instrumentation amplifiers in general.

GAIN ERROR

Errors in gain for the simple instrumentation amplifier of Fig. 1 are essentially those of the resistors themselves, unless the gain is unusually high or the op-amp has poor open-loop gain or common-mode rejection ratio (CMRR). For such a simple circuit, the effects of resistor mismatch cause much greater problems by common-mode rejection than by gain error. Small gain errors can often be tolerated or calibrated out, whereas in the case of poor common-mode rejection spurious signals end up mixed with the desired signal and may be impossible to remove. This situation is exacerbated greatly if the common-mode signal is comparable to (or even greater than) the desired one, which unfortunately is often the case (this is the major reason why an instrumentation amplifier is usually needed).

COMMON-MODE REJECTION RATIO

As already mentioned, if the inverting and noninverting gains are perfectly matched, then input voltages common to both of these inputs theoretically never appear at the output. This is, of course, the ideal case. In the case of the simple circuit of Fig. 1, mismatch in the ratios $R2/R1$ and $R3/R4$ results in different gains from each input to the output (neglecting the sign difference, which is obviously intentional). This means that signals common to both inputs will appear at the output to some degree. The accepted standard measurement of the extent to which this will occur is the CMRR. There are various ways to define the CMRR, but the classic definition is simply the differential-mode gain divided by the common-mode gain. This is a rather large number, and for this reason it is almost always expressed in (voltage) decibels:

$$\text{CMRR(dB)} = 20 \log_{10} (\text{CMRR(ratio)})$$

The sign of CMRR can also be confusing. Frequently, CMRR is expressed as something like -70 dB. Strictly speaking, this would imply a common-mode gain greater than the differential-mode gain, but (although it is possible to design an amplifier that would accomplish this) such figures should be treated as common-mode *acceptance* (the inverse of CMRR) when dealing with realistic instrumentation amplifiers. The CMRR of a simple instrumentation amplifier as depicted in Fig. 1 is obviously dependent on the accuracy to which the resistors can be matched. Using the symbol Δ_R for resistor mismatch, and assuming no correlation in the matching, it can be shown that worst-case CMRR owing to this term alone

can be expressed by

$$\text{CMRR} = \frac{1 + G}{4\Delta_R}$$

where G is the selected gain.

For off-the-shelf 0.1% tolerance resistors and a gain equal to 10, the CMRR could be as poor as 2750 or about 69 dB. This would be regarded as extremely meager in most precision applications. Clearly, a fine trim on any of the four resistors can be used to improve this parameter, and that is why integrated circuit instrumentation amplifiers are so commonly used where automated resistor adjustments (by means of a laser or other techniques) can be cost-effectively employed in their manufacture. Using such techniques, the critical resistor *ratios* can be trimmed to a tolerance of 0.005% or so, resulting in a CMRR in the preceding example of over 100 dB.

Even when the resistors are perfectly matched, common-mode errors in the op-amp itself cannot be overlooked. Common-mode rejection in the op-amp shows up directly as a CMRR error in the output multiplied by the gain. Modern precision monolithic op-amps frequently boast CMRR specifications of 110 dB or more. When the instrumentation amplifier is used at low gains, such figures can be considered negligible compared to errors such as resistor mismatch. At high gains, however, this may not be the case. The open-loop gain of the op-amp, ideally infinite, can also upset the behavior of both CMRR and gain error.

An often more serious consideration is the CMRR behavior with frequency. Both CMRR and gain roll-off with increases in frequency are typical of operational amplifiers with dominant-pole (or alternate) methods of compensation designed to be operated with good closed-loop stability and high feedback factors (often 100%). This effect is exacerbated in many precision op-amps that use multiple stages of gain, which, in turn, demand high levels of internal frequency compensation or, to put it another way, tend to have poor ac characteristics.

The most commonly encountered frequency of interest for CMRR is the fundamental frequency of the power-supply grid, which is nominally 60 Hz in the United States and Canada and 50 Hz in much of the remainder of the world. Additional harmonics generated by power transformers, rectifiers, and thyristor control systems produce radiated frequencies that can be far more important than initially expected, given that capacitive coupling tends to increase with frequency whereas op-amp precision tends to decrease with frequency. The subject of RF (radio frequency) susceptibility with frequencies up to several GHz is obviously an extreme (but often important) consideration in the evaluation of CMRR in an instrumentation amplifier.

A full dynamic analysis of factors affecting CMRR (and other effects, such as ac gain error, ac power supply rejection ratio [PSRR], and settling time) is considered beyond the scope of the present discourse. Suffice it to say, however, that even at 50 Hz, many precision op-amps can produce CMRR errors far worse than simple dc analysis would predict.

ADVANTAGES OF THE BASIC INSTRUMENTATION AMPLIFIER

One advantage of the simple amplifier of Fig. 1 is that only one op-amp is used. This is, of course, good not only for economic reasons but also because only one op-amp contributes to the overall error budget. A less obvious advantage is that the input

voltage range is very high. Because of the attenuating effect of R_1 and R_2 , the common-mode input range can extend beyond the input voltage range of the op-amp itself. This is often useful where external signals approach or exceed the supply voltage available to the instrumentation amplifier.

SHORTCOMINGS OF THE BASIC INSTRUMENTATION AMPLIFIER

The most serious shortcoming of the configuration shown in Fig. 1 is that the input impedance is not very high (unless the resistors are made impractically large, with severe noise and bandwidth penalties). To make things worse, the analysis of the effects of the input impedance is highly dependent on the characteristics of the input signals.

The impedance at +IN is equal to $R_1 + R_2$, while at -IN it is equal to R_3 because the feedback action of A1 forces a low impedance at its inverting input. For the usual arrangement, where R_2 is equal to R_4 , and R_1 is equal to R_3 (to preserve symmetry at the op-amp inputs, for input bias current considerations among others), the input impedance is therefore considerably higher at +IN than at -IN. While this can be corrected by reducing the values of R_1 and R_2 (or better, by adding a resistor from +IN to ground), this is often a dangerous practice. For truly differential inputs with matched source impedances, the input impedance is not actually asymmetric (due to the fact that the inverting input of A1 follows part of the signal at +IN). Attempting to balance the absolute input impedances thus could cause severe CMRR errors when the inputs are driven differentially from well-matched source impedances. If the source impedances are not well matched (or at least not well defined), the usefulness of this configuration becomes highly questionable.

For the record, the common-mode input impedance turns out to be $\frac{1}{2}(R_1 + R_2)$ and the differential value is equal to $2(R_1)$. These equations only hold for balanced differential sources.

Another nuisance is that the gain cannot be varied without simultaneously changing two resistors, and again tight matching must always be preserved if CMRR is not to suffer. However, if R_2 and R_4 are each split into two resistors, then an additional resistor connected between the midpoints can be used to increase the gain with minimal effect on CMRR. Now, of course, six matched resistors are required to obtain good CMRR. Such shortcomings are the main reasons to look to alternative instrumentation amplifier topologies.

Despite the limitations, the basic configuration of Fig. 1 is useful enough that at least two companies (Analog Devices and Burr-Brown) are producing this circuit in (laser-trimmed) monolithic form at the time of writing.

THE CLASSIC THREE-OP-AMP INSTRUMENTATION AMPLIFIER

Adding buffer amplifiers to both inputs will clearly remedy the input impedance problem of the aforementioned configuration, but a little more elaboration can solve the gain-setting problem as well. This leads to the classic three-op-amp configuration of Fig. 2. Once more, a definitive reference has proven to be elusive, but at the very least it dates back to a 1966 George Philbrick publication (5).

Referring to Fig. 2, A3, in conjunction with R_1 – R_4 , forms a differential amplifier identical to the circuit of Fig. 1. If R_G is omitted, A1 and A2 act as unity-gain buffers, removing the input impedance problems described previously. In the presence of R_G , the differential gain between the inputs and outputs of A1 and A2 becomes

$$G_{\text{diff}} = 1 + \frac{(R_5 + R_6)}{R_G}$$

The common-mode gain, however, remains at unity. Thus, when R_1 – R_4 are carefully trimmed for optimum CMRR, the differential gain can be increased by reduction of R_G without affecting overall common-mode gain. Because of the way CMRR is defined (as the ratio of differential-mode gain to common-mode gain), the effective CMRR of the amplifier becomes proportional to G_{diff} . The overall gain of the amplifier will be the product of the gain of the second stage and G_{diff} or, (assuming $R_1/R_2 = R_3/R_4$),

$$G = \left(1 + \frac{(R_5 + R_6)}{R_G}\right) \left(\frac{R_4}{R_3}\right)$$

Thus the lower limit on gain is set by R_4/R_3 .

The best distribution of the gain between the first and second stages is the subject of considerable compromise. By using a gain below unity in the second stage ($R_4 < R_3$), the overall CMRR (for a given resistor mismatch) is increased, and the lower gain limit is extended. Unfortunately, this tends to put the gain burden on the op-amps A1 and A2, thus amplifying their input referred errors and reducing their bandwidth.

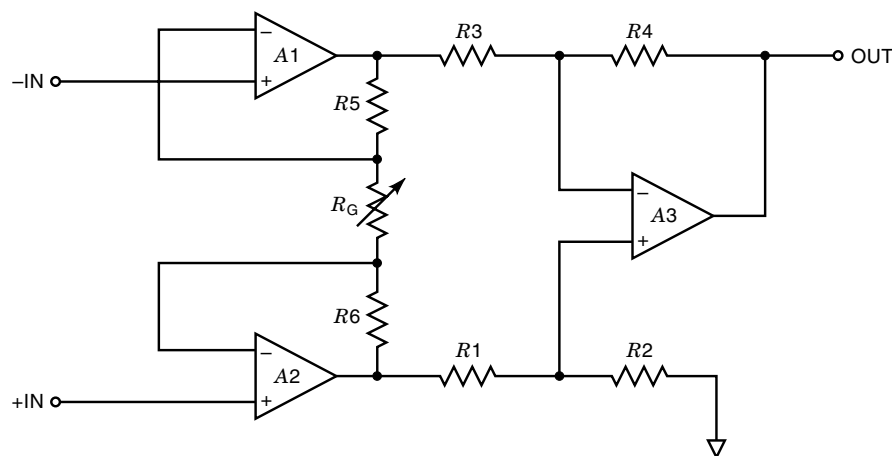


Figure 2. Three-op-amp instrumentation amplifier.

Conversely, taking gain from the second stage ($R4 > R3$) can improve the overall bandwidth at the expense of CMRR (and amplification of the input referred errors of A3) and increases the limit on the lowest achievable gain.

A more subtle (but extremely important) effect, however, is the limit imposed on the common-mode input range. With no differential input, the common-mode input range is limited either by the input voltage range or the output voltage swing (whichever happens first) of A1 and A2. With modern “rail-to-rail” op-amps, this restriction can often be minimal (though most high-precision and high-speed op-amps still require considerable headroom, often a couple of volts or so). Even so, the power supplies dictate a limit on the output swing.

With a differential input applied, the output of either A1 or A2 swings positive about the common-mode input voltage, while the other swings negative. Thus the entire differential output voltage of the first stage directly subtracts from the available common-mode input range. This effect also limits the output swing of the overall amplifier in the case where the second stage is operated at a gain below unity.

As an extreme example, consider the case in which the supply voltages are ± 15 V and all op-amps are limited in range only by these supplies. Also consider the case in which the second stage is operated at a gain of 0.1 ($R3/R4 = 10$). With the outputs of A1 and A2 at opposite supply rails, the output swing will be only 3 V, and even if a 1 V output swing is all that is required, the common-mode input range will be reduced by 10 V, barely making a ± 10 V common-mode input range with symmetric swings for the input amplifiers.

A more practical example might be the case in which the second stage gain is set to unity and ± 10 V swings are required at the output. This output swing can now be achieved, but again the common-mode input range is still barely ± 10 V, even with ideal amplifiers. For realistic op-amps requiring 2 V of headroom, the common-mode input range is at best ± 8 V with a full differential signal applied. If this range is unacceptable, then gain must be provided in the second stage.

The symmetric swing referred to previously actually occurs when $R5$ is equal to $R6$. This is highly desirable because it maintains as much symmetry as possible in the two input op-amps. Making $R5$ and $R6$ dissimilar can cause offset problems (due to the input bias currents of the op-amps flowing through unequal impedances) and ac CMRR problems (because the effective closed-loop gains of the amplifiers are different). Purposely making $R5$ and $R6$ different values can position the common-mode input range closer to one supply or the other, but this is usually a poor solution to the problem.

Suffice it to say that most commercial implementations of this architecture (and there are many) have used symmetric values for $R5$ and $R6$ and a second stage gain of unity (occasionally greater). Even with these limitations, this is one of the most powerful configurations available for instrumentation amplifier realization.

REFERENCE INPUT

Because the instrumentation amplifier has a differential input and single-ended output, it is necessary to refer the output voltage to some reference potential. So far, this has been represented by a symbol commonly referred to as ground, which is a somewhat universal reference point in most analog systems. Often, however, the integrity of a global ground connection is

highly questionable, particularly when high degrees of analog precision are sought. In some cases, it may be necessary to refer the output to some other potential; this is particularly true where only a single supply voltage is available and the so-called ground is actually one of the supplies (usually the negative one). In the circuit of Fig. 2 the ground connection is made via $R2$, but in fact this point can be used as a reference input since it has (close to) unity gain to the output. Most modern, commercially available instrumentation amplifiers therefore label this pin as reference rather than ground. In the interest of clarity, many of the figures in this article show the reference input as a ground symbol. For the purpose of this article, the ground symbol may be assumed to correspond to a reference input, which can be used to refer the output to a different point (or a precision Kelvin ground) inside the system.

SENSE INPUT

The overall feedback loop for an instrumentation amplifier is often completed externally to allow for Kelvin sensing of remote potentials where there exists the possibility of significant voltage drops along the connecting wires (due to finite loading of the amplifier output). Thus, some instrumentation amplifiers feature a sense input, which is almost always connected to the final amplifier output, either locally or remotely via force and sense wires.

Care must be used when taking advantage of this feature, however, since most instrumentation amplifiers can become unstable when presented with large amounts of capacitance at the output, a condition often created by long wires connected to the output (and sense) pins.

INPUT AND OUTPUT REFERRED ERRORS

This is a good point at which to introduce the concept of input and output referred errors. In virtually all variable-gain instrumentation amplifiers, most of the error terms actually have two components. One is called output referred, since it appears at the output independent of gain setting. Examples of this are the CMRR error caused by mismatch of resistors $R1$ – $R4$ in Fig. 2 and all errors caused by op-amp A3. The other component is called input referred, and it appears at the output multiplied by the overall gain of the amplifier. Examples of this are almost all errors attributable to the input op-amps A1 and A2. Most instrumentation amplifier data sheets specify these terms (such as input offset voltage and output offset voltage) separately. Since the sign is generally unpredictable, at any particular gain the terms are usually presumed to add at the output (except for noise, where an RSS [root sum of squares] summation technique is generally applied).

In passing, it should be noted that it is not so much the individual errors of A1 and A2 that appear as the input referred error (again, except for noise) but the difference between them, since any systematic errors appear as a common-mode signal to the second stage. For this reason, a matched monolithic dual op-amp is usually used for A1 and A2 in the configuration of Fig. 2.

THE TWO-OP-AMP INSTRUMENTATION AMPLIFIER

The simple circuit of Fig. 1 attempts to balance the inverting and noninverting gains of an op-amp operated in a closed-loop

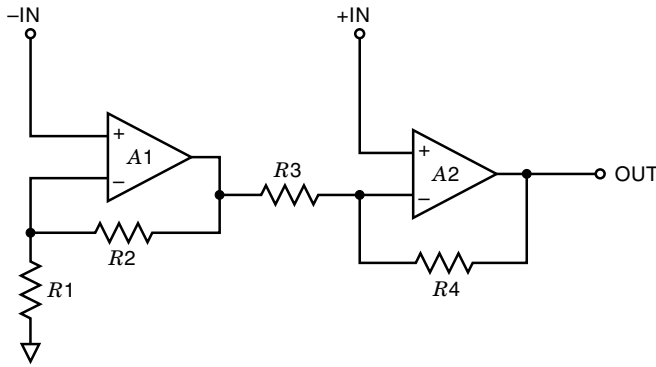


Figure 3. Two-op-amp instrumentation amplifier.

configuration by attenuation of the signal at the noninverting input. Another alternative is to leave the noninverting input alone (thus maintaining its inherently high input impedance) and to use a second op-amp to balance the gains by amplifying the gain at the inverting input. Since the latter op-amp can be operated in a noninverting gain configuration, a high input impedance for both final instrumentation amplifier inputs can be preserved. The basic circuit is shown in Fig. 3. The op-amp A2 provides the differential function, while A1 amplifies the $-IN$ input to equalize the gains between the $+IN$ and $-IN$ inputs. The incremental transfer gain from $+IN$ to the output is given by

$$G_+ = 1 + \frac{R_4}{R_3}$$

whereas the incremental transfer gain from $-IN$ to the output is given by:

$$G_- = -\left(1 + \frac{R_2}{R_1}\right)\left(\frac{R_4}{R_3}\right)$$

To obtain good CMRR, the absolute value of these gains must again be equalized, and this is achieved by making the ratio of R_2/R_1 equal to R_3/R_4 , in which case the common-mode gain is theoretically zero and the differential gain is given by

$$G_{\text{diff}} = 1 + \frac{R_4}{R_3}$$

The resistor matching requirements for CMRR are very similar to the simple configuration of Fig. 1, except that the CMRR depends on $G/4\Delta R$ rather than $(1 + G)/4\Delta R$. Choosing R_2 equal to R_3 and R_4 equal to R_1 balances input bias current errors of the op-amps (but not their dynamic characteristics—more on this later). Of more importance, however, are the gain and common-mode input range limitations. The first thing to be noted is that unity gain (or anything less) is impossible, because under these conditions the value of R_2 must be infinite. Second, even at somewhat higher gains, it should be noted that A1 *always* amplifies the common-mode voltage, producing severe limitation of common-mode input range due to available swing at the output of A1. For this reason, it is rare to find this configuration used in an overall gain of less than five (the closed-loop gain of A1 asymptotically approaches unity as the overall gain is increased, progressively ameliorating this problem).

Another limitation concerns the ac characteristics. Like the circuit of Fig. 2, a matched dual monolithic op-amp used for

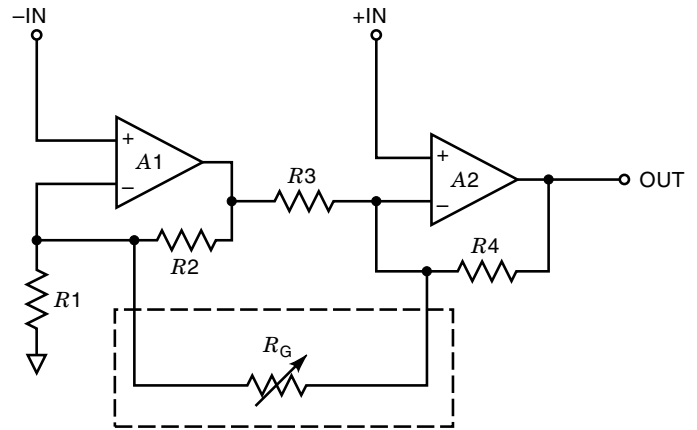


Figure 4. Modified two-op-amp instrumentation amplifier.

A1 and A2 can contribute greatly to the goal of good dc precision. Unlike the circuit of Fig. 2, however, the op-amps operate under very different individual closed-loop conditions. This causes problems when trying to maintain a good CMRR versus frequency, particularly since the entire phase shift of A1 appears in the inverting path but is totally absent in the noninverting one. Some phase compensation techniques can be applied to help this situation, but for the most part if good ac characteristics are required, this is unlikely to be the topology of choice.

A final note concerns the variable-gain characteristics of the circuit of Fig. 3. Like the circuit of Fig. 1, any gain change requires that the ratio matching of R_2/R_1 to R_3/R_4 be left unchanged if CMRR performance is not to be degraded. Bearing in mind that there is a practical minimum gain for this configuration, there is a way to increase the gain without severe CMRR penalty. Figure 4 shows the modification. The addition of R_G between the inverting inputs of A1 and A2 modifies the gain equation to

$$G = \left(1 + \frac{R_4}{R_3}\right) + \left(\frac{2R_4}{R_G}\right)$$

Using this technique, some commercial realizations of this configuration have had their usefulness greatly extended. Companies such as Linear Technology and Burr-Brown feature this configuration in their product portfolio (at the time of writing) in the form of monolithic integrated circuits.

AN ALTERNATIVE APPROACH USING THREE OP-AMPS

Another possibility using three op-amps is shown in Fig. 5 (6). Op-amp A2 does most of the work, with A1 (connected as a unity-gain buffer) restoring a high impedance to the inverting input ($-IN$). In this case, however, the inverting and noninverting gains of A2 are equalized by using a third op-amp (A3) to provide active attenuation of the common-mode signal appearing at $-IN$. This common-mode signal is impressed across R_1 and, by the action of A3, is injected as a nulling current back into the inverting input of A2. It can be shown that optimum CMRR is obtained when the ratio of R_3/R_1 is made equal to the ratio of R_2/R_{FB} . For symmetry, generally R_1 will be made equal to R_{FB} and R_2 will be made equal to R_3 . This topology has some interesting characteristics:

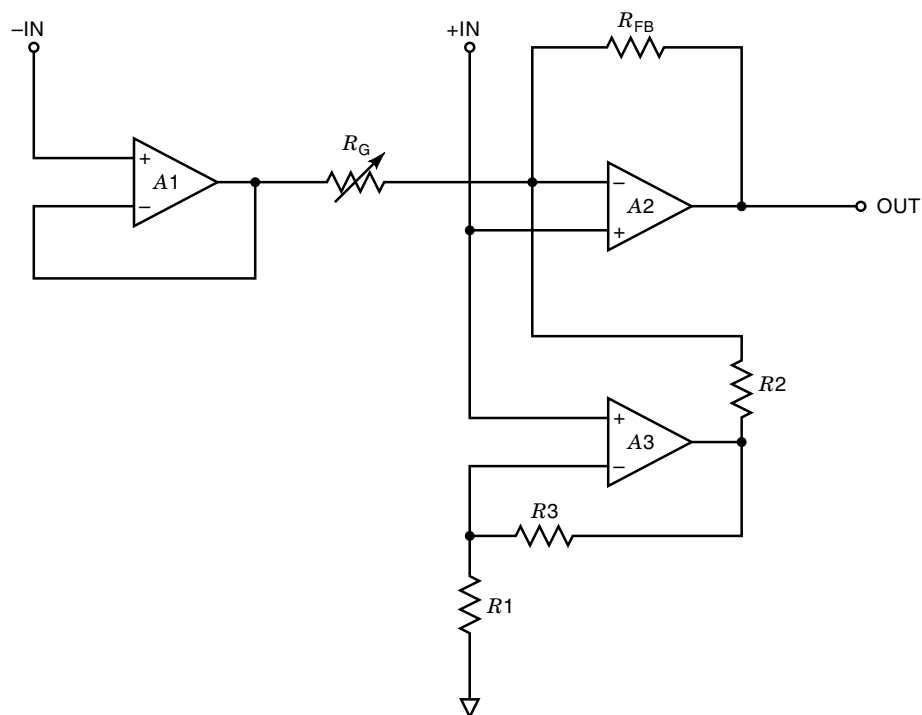


Figure 5. An interesting alternative three-op-amp configuration.

1. CMRR can be trimmed by a fine adjustment on any of R_1 , R_2 , or R_3 without affecting the gain of the amplifier.
2. The overall differential gain is simply R_{FB}/R_G . Thus gains from zero to any practical value are available by adjusting only one component (R_G). The CMRR (referred to output) is not affected by this gain adjustment.
3. The circuit is capable of a very wide common-mode input range. Op-amps A1 and A2 limit this range by their input/output swings in the usual fashion, of course, but the main limitation comes from the output swing of A3 since it amplifies the common-mode voltage by the factor $(1 + R_3/R_2)$. This limitation can be removed, however, by choosing the ratio of R_3/R_2 to be small enough that the other amplifiers become the limiting factor. The common-mode input range of the overall amplifier can thus approach the limits of the op-amps themselves.

This circuit (as might be anticipated) is not without drawbacks. The most significant of these is due to the fact that the errors of A3 are effectively amplified by the ratio R_{FB}/R_2 (this is an output referred error, not affected by gain setting). So attempts to maximize the common-mode input range in this manner tend to produce higher output offset voltages and noise. Also, A1 and A2 operate in very different closed-loop conditions, making it difficult to maintain good CMRR at high frequencies. Finally, the inverting input ($-IN$) is loaded by only one op-amp, whereas the noninverting input ($+IN$) is loaded by two op-amps, making the input characteristics somewhat asymmetric, especially in the case where the op-amps have significant input bias currents. Nevertheless, this configuration has been used (with some modification) to produce a monolithic instrumentation amplifier capable of operating from a single 5 V supply, where the common-mode input voltage range extends all the way to the negative supply rail (7).

VENTURING BEYOND OP-AMP DESIGN TECHNIQUES

The examples presented so far have relied on configurations formed from conventional op-amps and resistors. True, all of the aforementioned circuits have been integrated in monolithic form at some time or another; but they all could also be effectively produced using commercially available off-the-shelf components. In parallel with developments along these conventional lines, the monolithic integrated circuit industry has produced many topologies for the realization of the instrumentation amplifier function, most of which would be difficult, and certainly uneconomic, to produce outside the environment of a monolithic integrated circuit. The remainder of this article focuses on some of the more important techniques used to produce monolithic instrumentation amplifiers.

CURRENT-FEEDBACK TECHNIQUES

The traditional design approach of using conventional operational amplifiers with feedback consisting of resistive components is nowadays often referred to as *voltage feedback*. In the mid-1980s a new term started to appear in op-amp literature: *current feedback*. A current-feedback op-amp differs from a conventional one in that its inverting input is internally held at a low impedance; the displacement current in the compensation capacitor is ultimately derived from the current flowing in the feedback network (8). This enables such op-amps to have very high slew-rate characteristics. Also, to the extent that the inherent input impedance of the inverting input is less than that of the feedback network, such op-amps maintain a more constant bandwidth as the closed-loop gain is increased than their conventional counterparts (which tend to have a fixed gain-bandwidth product). The drawback here is that such op-amps have intrinsically imbalanced

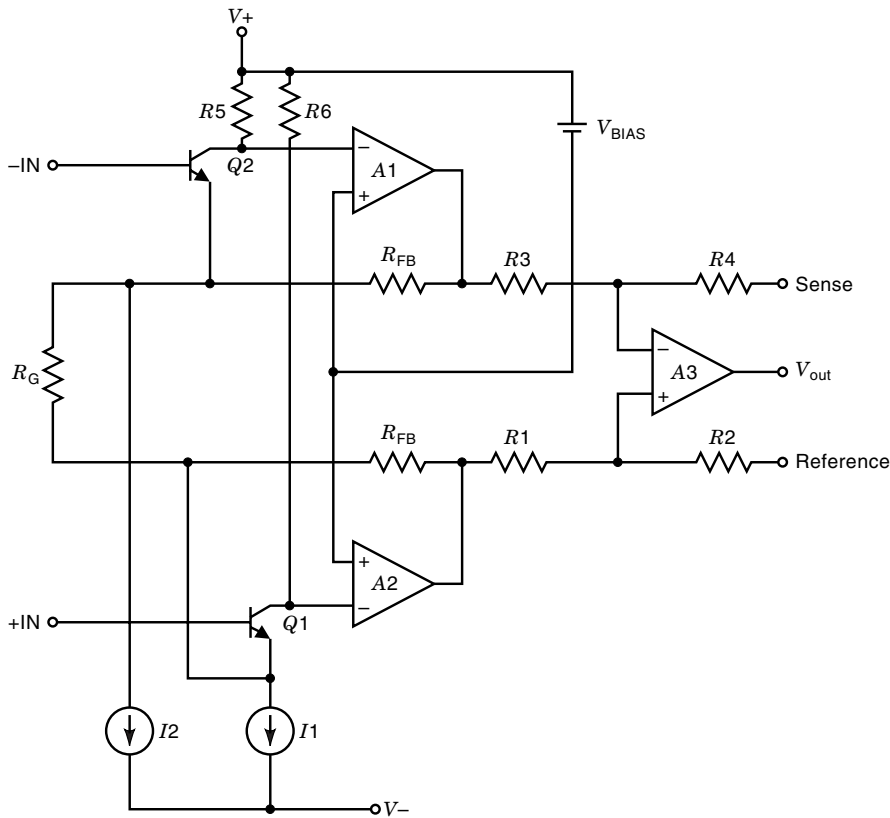


Figure 6. Current-feedback version of Fig. 2.

input stages and cannot approach the precision of more conventional types, despite many ingenious schemes to balance them up. Actually, current-feedback is an offshoot of what used to be called *cathode feedback* in the vacuum-tube days and is not a fundamentally new technique from a circuit theory viewpoint.

Applying current feedback to an instrumentation amplifier is actually much easier than in the case of a general-purpose op-amp (although all configurations described so far can be implemented using current-feedback op-amps). This is because the instrumentation amplifier is an inherently dc balanced structure with well-defined feedback components.

Figure 6 shows a current-feedback approach to the configuration of Fig. 2 (9), now produced in integrated circuit form by several manufacturers. Essentially, this consists of the classic three-op-amp design preceded by a preamplifier consisting of $Q1$ and $Q2$ and their associated load resistors, $R5$ and $R6$. Feedback (via the resistors R_{FB}) is now returned directly to the emitters of the input pair rather than to inherently high-impedance op-amp inputs, as in previous examples. The bias currents for $Q1$ and $Q2$ are not set by current sources $I1$ and $I2$ (as might first be thought) but rather are provided from the outputs of $A1$ and $A2$, in a common-mode feedback loop controlled by V_{bias} . Since these currents must flow through the feedback resistors (R_{FB}), $I1$ and $I2$ are added in order to center the common-mode swing at the outputs of $A1$ and $A2$ (these current sources are sometimes omitted when the input stage currents are small enough to produce negligible voltage drops across the feedback resistors).

The gain equations for this arrangement are identical to that of the example presented in Fig. 2, at least under dc condi-

tions. However, to the extent that the dynamic impedance at the emitters of $Q1$ and $Q2$ is lower than the value for R_G , the latter component does not greatly attenuate the overall ac feedback, resulting in an approximately constant bandwidth (rather than a constant gain-bandwidth product typical of voltage-feedback configurations) as R_G is varied. The configuration does slow down at higher gains (as R_G becomes comparable to or less than the input transistors' dynamic emitter impedance), but it can still offer a considerable practical improvement in overall bandwidth and settling time.

Other advantages stem from the fact that only two transistors comprise the input stage, rather than the four necessary for two conventional op-amps. This leads to reduced input-referred errors (particularly noise). A minor disadvantage comes from the fact that the uncorrelated portion of the error currents in $I1$ and $I2$ (and the input transistors) has to flow through the feedback resistors and appears as an increase in output referred offset and noise. With careful design, however, this effect can be kept small. Similar current feedback configurations can be devised for the other voltage-feedback topologies presented previously.

ACTIVE FEEDBACK

In the sense used here, *active feedback* involves the use of an active voltage-to-current converter as a feedback element, instead of the resistor, which until now has been shown as the basic feedback component. Confusingly, the term *current feedback* has also been employed for this type of feedback, and there is no real accepted standardization in common usage (maybe

this article will encourage such a standard). At least I feel I have defined my own nomenclature (with some historical justification), but for the record I am one of the many whose past publications have (unintentionally) contributed to the present state of confusion.

CONCEPTUAL ACTIVE-FEEDBACK INSTRUMENTATION AMPLIFIERS

In the basic circuit of Fig. 7, two identical differential bipolar transistor pairs (they could be field-effect transistors in theory) are degenerated by R_{G1} and R_{G2} (assumed to be equal for now, as are the current sources $I1$ through $I4$). The degeneration resistors are large enough to ensure that any desired differential input voltage (impressed between $+IN$ and $-IN$) will retain some sensible current in both of the input transistors, $Q1$ and $Q2$. Both differential pairs are summed into differential loads consisting of $R1$ and $R2$, which are further sensed by op-amp $A1$.

A negative feedback loop is provided from the output of $A1$ to the second differential pair (note that the base of $Q3$ is grounded). When the loop is closed (frequency compensated by capacitor C_C), the output voltage will thus be closely equal to the differential input, independent of the common-mode input voltage (10). The nonlinearities of the two transistor pairs nominally cancel under this arrangement; thus the circuit forms a unity-gain amplifier with common-mode rejection limited only by second-order nonidealities in the transistors themselves. The most important nonideality (for CMRR purposes) is the mismatch of output impedance of the two transistor pairs. Without resort to any kind of trimming, this can be a large number indeed, leading to a high CMRR.

The drawback of this arrangement is that it is difficult to adapt it to variable gain. At first glance, making R_{G1} a variable component does the trick (and will result in variable gain), but under these circumstances the nonlinearities of the transistor pairs no longer cancel (except at unity gain), resulting in cubic distortion products that are likely to be unacceptable at higher gains.

Another idea is to place a resistive attenuator between the output of $A1$ and the base of $Q4$. While this technique will preserve linearity, it has the unfortunate effect of amplifying both input and output referred errors. If the gain range is small

and the corresponding input range is well defined, a very good amplifier can result, since the transconductance of both stages can be optimized. For wide gain ranges, however, the result is a noisy instrumentation amplifier with poor output referred errors and loss of bandwidth at high gain settings.

One possible method of producing a more usable instrumentation amplifier from this general idea is shown in Fig. 8. In this configuration, the nonlinearities of the transistor pairs are corrected by enclosing them in localized operational amplifier feedback loops. This enables the gain to be varied (without distortion penalties) by adjustment of the ratio of R_{G1} to R_{G2} . Some merging of the op-amp functions is possible, and a monolithic design based on this idea has been reported in the literature (11).

One interesting aspect of this configuration is that the major feedback loop (completed by op-amp $A1$) is almost totally divorced from the input section. This effectively means that as R_{G1} is reduced to increase gain (assuming R_{G2} is left alone), the loop bandwidth does not change. Thus the overall amplifier bandwidth tends to remain constant even as the gain is varied over a wide range (gains of zero to any practical upper limit are possible with this arrangement and with most other active-feedback configurations).

Unfortunately, this configuration has developed a rather complex input stage, which tends to produce high levels of input referred errors (particularly input offset voltage, input offset drift, and input referred noise). Other schemes that do not need such a complex input stage have been developed in order to alleviate this situation. Before continuing, however, a note concerning output referred errors is in order. Active-feedback amplifiers, in general, tend to have high output referred errors (their major drawback compared to more conventional techniques using op-amps and resistors). This is because the active voltage-to-current converters used to provide the feedback have much larger offset and noise components than simple resistors. The high degree of gain flexibility (usually from zero upward), intrinsically high common-mode rejection (without trimming), and potentially high speed characteristics of the active-feedback instrumentation amplifier come with a penalty in terms of precision at low gains. As a crude generalization, modern active-feedback instrumentation amplifiers offer significant advantages in terms of speed and CMRR at most gains but tend to be poor in most other respects at gains below 50 or so. At lower gains (certainly below 10), more conventional

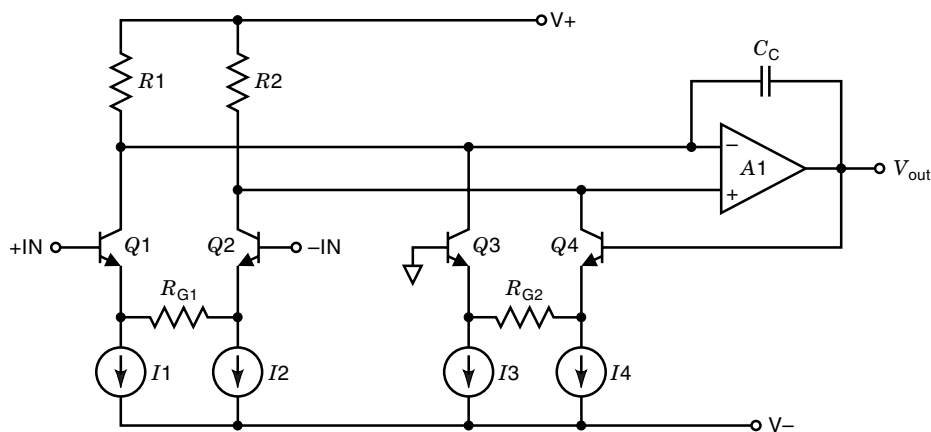


Figure 7. Conceptual active-feedback instrumentation amplifier.

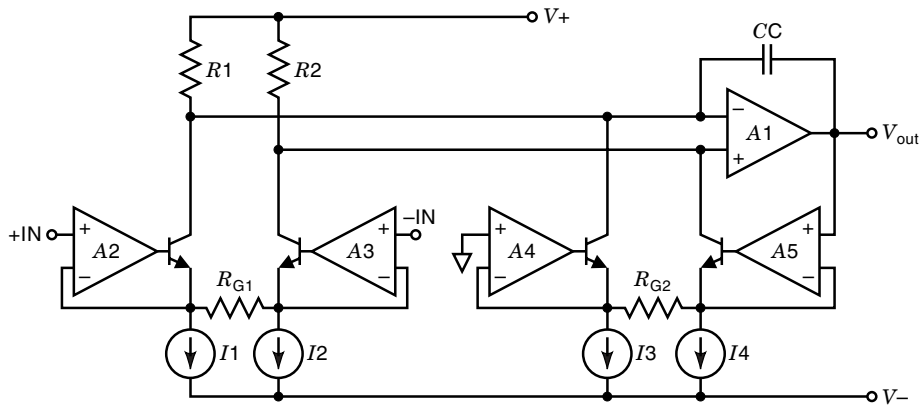


Figure 8. Active-feedback instrumentation amplifier with variable gain.

techniques are likely to provide better overall performance and, in conjunction with the current-feedback approach, are likely to be competitive in terms of speed.

PRECISION ACTIVE-FEEDBACK INSTRUMENTATION AMPLIFIERS

The basic idea for an active-feedback amplifier with a precision input stage is depicted in Fig. 9. Transistors $Q1$ and $Q2$ are biased at a quiescent point determined by the standing currents from a highly linear voltage-to-current (V -to- I) converter. Feedback provided by op-amp $A1$ to the V -to- I converter controls the differential characteristics. If the open-loop gain is high, the feedback loop will force the currents in $Q1$ and $Q2$ to be equal—regardless of differential input—while the output currents of the converter are not equal due to the presence of the gain-setting resistor, R_G . Because the input transistors are operated under identical conditions, the differential input voltage is directly forced across R_G , so the differential output currents of the V -to- I converter are $2(\Delta V_{in}/R_G)$. Since the V -to- I converter is presumed to be linear, the output voltage is now equal to the input voltage multiplied by the product of R_G

and half the differential transconductance of the converter, independent of the common-mode input voltage because of the converter's inherently high output impedance.

The output referred errors (as in previous examples) are those of the active V -to- I converter and can be comparatively high. The input errors, though, can closely approach the intrinsic errors of $Q1$ and $Q2$ alone, yielding a theoretical input stage precision about as good as anything available on a monolithic integrated circuit.

The structure of the V -to- I converter is obviously critical to the performance of this topology, and various methodologies have been used from time to time. With the inherent advantage of high CMRR without trimming, it is not surprising that the first fully integrated monolithic instrumentation amplifier would be an active-feedback design (12) (which became commercially available as the Analog Devices part number AD520). Four years later, a much improved design (13) was introduced (the AD521). Figure 10 shows the basic topology. Amplifier block $A2$ adjusts the current sources $I3$ and $I4$ to maintain constant currents in the input transistors, $Q3$ and $Q4$. Under these circumstances, the differential input voltage is accurately forced across the resistor R_G . The difference in $I3$ and $I4$ is now simply twice the input voltage divided by R_G . The current sources $I1$ and $I2$ are slaved to $I3$ and $I4$, so their difference is exactly the same. The amplifier $A1$ forces half of this difference to appear across the resistor R_s . The gain is now simply R_s/R_G . In the example presented, both R_s and R_G were external components, although there is not much flexibility in the choice of R_s because of limitations of the maximum values of the internal current sources.

The need for two amplifier blocks and parallel-connected controlled current sources tends to exacerbate output referred errors in such designs, and one method of alleviating this is to place the V -to- I converter in series with the input stage. This is not a trivial task, since the resulting V -to- I converter must reside entirely in the space left between the extremes of the common-mode input range and one of the supplies (usually the negative). One method of achieving this is shown in Fig. 11, first implemented by National Semiconductor (14).

The V -to- I converter is composed of $A2$, $A3$, $Q3$, and $Q4$ with $I1$ and $I2$ providing emitter bias current for the input stage, $Q1$ and $Q2$. To prevent negative common-mode excursions at the input from causing $Q3$ and $Q4$ to saturate, the output is attenuated and level shifted by resistors $R3$ and $R4$. Balance is restored to the other side of the V -to- I converter by an

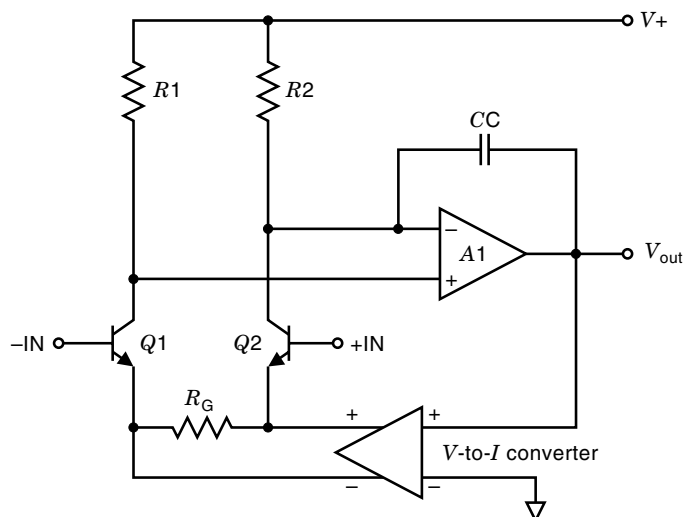


Figure 9. Conceptual topology for a precision active-feedback instrumentation amplifier.

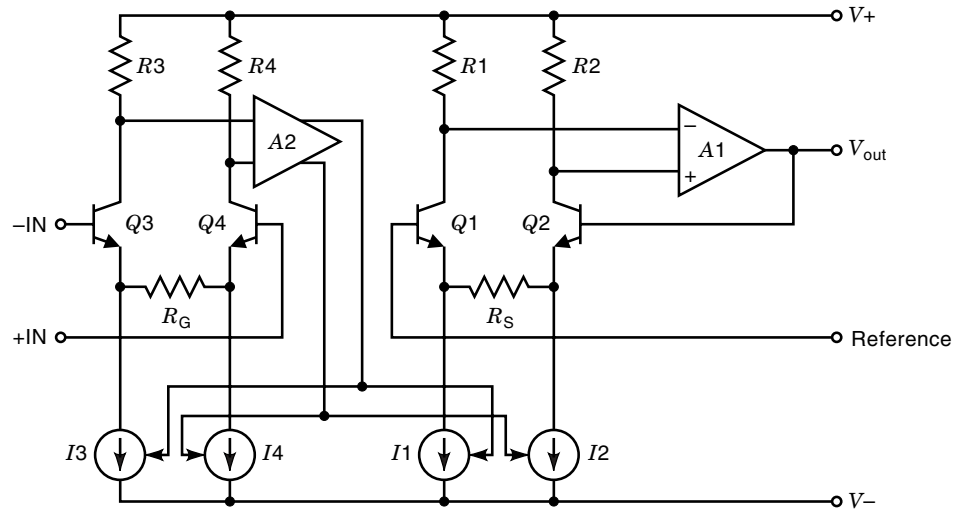


Figure 10. Active-feedback instrumentation amplifier using a parallel V-to-I converter.

identical network consisting of $R5$ and $R6$. $A2$ and $A3$ can now be made extremely simple because no level shifting is required and any systematic offset will cancel between the two op-amps. One drawback of this technique is that mismatches in the ratios of resistors $R3/R4$ and $R5/R6$ can cause severe degradation of the negative power-supply rejection.

The effective differential transconductance of the V-to-I converter is twice the inverse of R_{G2} , and similarly the effective input stage transconductance is twice the inverse of R_{G1} . The overall transfer function is now given by

$$G = \left(\frac{R4}{R3 + R4} \right) \left(\frac{R_{G2}}{R_{G1}} \right)$$

For the example of Ref. 14 (National part number LM363), the gains are selected by pin-strapping internal resistors to provide

gains of 10, 100, or 1000 (fixed-gain versions with gains of 10, 100, or 500 are also provided).

Other implementations of this architecture (15,16) (Analog Devices part numbers AMP-01 and AMP-05) have left R_{G1} and R_{G2} as external components with the ratio of $(R3 + R4/R4)$ internally set at 20. This results in an overall differential gain equal to $20(R_{G2}/R_{G1})$ and, like most configurations of the active-feedback amplifier, allows a user-defined gain setting from zero to any sensible upper limit.

GLOSSARY OF FREQUENTLY ENCOUNTERED TERMS

Gain

The gain specification relates to the transfer function of the instrumentation amplifier. Typically, gain can be fixed, adjustable

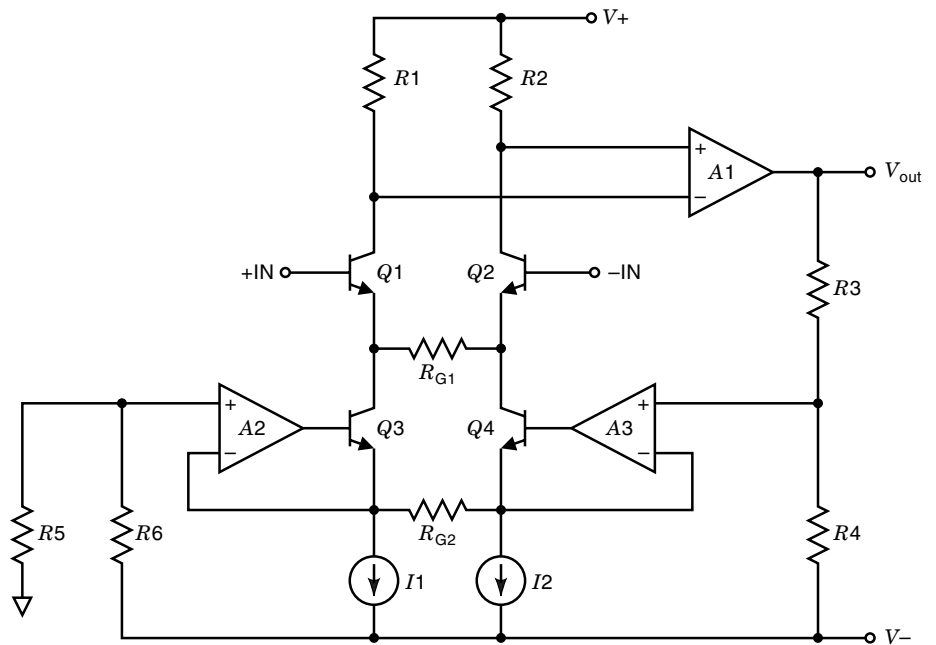


Figure 11. Active-feedback instrumentation amplifier using a serially connected V-to-I converter.

by pin-strapping, digitally selectable, or controlled by an external resistor (R_G). In the latter case a transfer equation will normally be provided.

Gain Range

The gain range is the overall range over which the gain equation is considered valid. At the lower end it is generally limited by the type of configuration used, while at the upper end it is often theoretically unlimited. Practically, at very high gains, all instrumentation amplifiers eventually exhibit errors that render them unusable. When an upper figure appears on a data sheet, it is usually the point above which the device manufacturer is not willing to provide any guaranteed specifications.

Gain Error

The number given by the gain error specification (usually expressed as a percentage) describes the maximum deviation from the gain equation. This is, for convenience, often quoted (and tested) at several fixed gains, with the user left to interpolate between them. Note that when an external resistor appears in the gain equation, the absolute tolerance of this component also appears as part of the gain error.

Nonlinearity

The instrumentation amplifier is assumed in simple theory to have a linear transfer characteristic from the differential input to the output. Obviously, in practice some nonlinearity will exist, and this represents an error that is nearly impossible to trim out or compensate for.

Nonlinearity is generally expressed as the peak deviation from a straight line superimposed on the plot of the output voltage as the input voltage is varied (at a particular gain setting) over a range wide enough to sweep the output over a specified excursion range. It is normally expressed as a percentage of the *maximum output excursion*, not the value of the output excursion at any other point. As such, nonlinearity is thus generally defined as an output referred error, and specifications will usually be quoted at a selection of representative gains.

The manner in which the straight line is defined can also cause some confusion. A line drawn between zero and theoretical full scale is probably the obvious one, but one argument suggests that this would produce a nonlinearity specification that includes other error terms (particularly offset voltage and gain error), which are generally specified separately (and in theory can be calibrated out). If the resulting line is drawn between measured zero and full scale, complications can occur when positive and negative signals are accommodated, since the resulting line drawn between the positive and negative full-scale output may not pass through zero. Generally, a line drawn between two arbitrary points (such as zero and theoretical full scale) gives rise to the term *end point nonlinearity*, whereas a line skewed to pass through two or more measured points with minimum peak error is termed *best-fit nonlinearity*. Careful reading of a data sheet specification is necessary to determine the effect that a given nonlinearity specification will have on an actual system.

Offset Voltage

The offset voltage of a practical instrumentation amplifier consists of two terms: the input offset voltage and the output offset

voltage. The input offset voltage is an error voltage that appears at the output multiplied by the selected (differential) gain of the amplifier. Therefore, it tends to predominate when the amplifier is configured for high gain applications. The output offset is an error term always present at the output, regardless of gain setting. Theoretically, it is defined as the error voltage at the output when the gain is set to zero, though this has to be extrapolated for many configurations where zero gain setting is impossible. The output offset voltage is obviously most troublesome at low overall amplifier gains.

Input (Bias) Current, Offset Current, and Input Impedance

The input current is simply the current drawn by one or both of the inputs when the amplifier is operated in its normal region (sometimes expressed as an average, or as a maximum of the two). This is often called an input bias current because such currents (at least in a bipolar junction transistor amplifier) are the base currents of the input transistors necessary to maintain them at their selected bias point. For amplifiers with field-effect transistor inputs, the input current generally reflects leakage currents associated with details of their fabrication.

The input offset current is the difference between the two input currents, of paramount importance when balanced source impedances are used. This is a measure of how well the two input currents are matched on a particular device.

Related specifications include differential and common-mode input impedance. The former represents the change in input offset current when a differential voltage is applied; it is normally defined as the reciprocal of the change in input offset current times the differential voltage used to induce it. The common-mode input impedance is normally defined as the reciprocal of the change in the sum of the two input currents multiplied by the common-mode voltage used to induce it. For most modern instrumentation amplifiers, the effect of both input impedance terms is usually negligible compared to the overall values of the input currents.

Common-Mode Rejection Ratio

The CMRR is a measure of the change in output voltage when both inputs are changed by equal amounts. These specifications are usually given for both a full-range input voltage change and a specified source imbalance. Because CMRR is usually a large number, it is usually expressed in decibels. For example, a CMRR of 10,000 could be expressed as 80 dB. Because CMRR generally consists of both input and output referred components but is always specified referred to the input, it will normally (apparently) increase with gain. For this reason, CMRR is almost always specified at several representative gain settings.

Common-Mode Input Voltage Range

The common-mode input voltage range represents the maximum excursion common to both inputs over which the CMRR specifications are guaranteed. For some instrumentation amplifiers, this is a function of differential input voltage, and often the input voltage range will be expressed by an equation rather than a fixed value. Another way to express the input voltage range is to specify a maximum excursion for either input, since for some amplifiers (particularly active-feedback types) this is a more realistic definition.

Power-Supply Rejection Ratio

The power-supply rejection ratio (PSRR) is a measure of the change in output voltage either when both power supplies are changed by equal amounts (in opposite directions, to remove any CMRR component) or when each supply is varied independent of the other (of course, there is only one supply to be varied in the case of a single-supply amplifier). Like CMRR, PSRR is often expressed in decibels, generally consists of both input and output referred components, is normally specified referred to the input, and will normally increase with gain.

Settling Time

Settling time is defined as that length of time required for the output voltage to approach and remain within a certain tolerance of its final value. It is usually specified for a fast full-scale input step and includes output slewing time. Since several factors contribute to the overall settling time, fast settling to 0.1% does not necessarily mean proportionately fast settling to 0.01%. In addition, settling time is not necessarily a function of gain. Some of the contributing factors include slew rate limiting, underdamping (ringing), and thermal gradients (long tails).

BIBLIOGRAPHY

1. J. R. Ragazzini, R. H. Randall, and F. A. Russell, Analysis of problems in dynamics by electronic circuits, *Proc. IRE*, May 1947.
2. J. Rose, Straight talk about data amplifiers, *EDN Magazine*, November 23, 1966.
3. R. Y. Moss, Errors in data amplifier systems, *Hewlett-Packard Journal*, July 1967.
4. Instrumentation amplifiers: A survey, Staff Report, *Electromechanical Design*, March 1967.
5. *Applications Manual for Computing Amplifiers for Modelling, Measuring, Manipulating, and Much Else*, George A. Philbrick Research Incorporated, 1966.
6. D. F. Bowers, "Instrumentation Amplifier with Single Supply Capacity and Simplified Gain Equation," US Patent 5,075,633, December 24, 1991.
7. D. F. Bowers, A new configuration for instrumentation amplifiers, *Proc. 10th European Conference on Circuit Theory and Design (ECCTD-91)*, Vol. 3, pp. 1324–1332, September 1991.
8. A new approach to op-amp design, Comlinear Corporation Application Note 300-1, March 1985.
9. S. A. Wurcer and L. Counts, A programmable instrumentation amplifier for 12-bit resolution systems, *ISSCC Digest of Technical Papers*, February 1982.
10. B. Gilbert, A high-performance monolithic multiplier using active feedback, *IEEE J. Solid-State Circuits*, **SC-9**: December 1974.
11. R. J. Van De Plassche, A wide-band monolithic instrumentation amplifier, *IEEE J. Solid-State Circuits*, **SC-10**: December 1975.
12. H. Krabbe, A high performance monolithic instrumentation amplifier, *ISSCC Digest of Technical Papers*, February 1971.
13. A. P. Brokaw and M. P. Timko, An improved monolithic instrumentation amplifier, *IEEE J. Solid-State Circuits*, **SC-10**: December 1975.
14. C. T. Nelson, A 0.01% linear instrumentation amplifier, *ISSCC Digest of Technical Papers*, February 1980.
15. D. F. Bowers, A versatile precision instrumentation amplifier, *ESSCIRC'83 Digest of Technical Papers*, September 1983.
16. D. F. Bowers, A fast settling FET input monolithic instrumentation amplifier, *ESSCIRC'85 Digest of Technical Papers*, September 1985.

DEREK F. BOWERS
Analog Devices Incorporated

CURRENT SHUNTS

Electric current can be measured directly, through the physical effects of the current itself, or indirectly, by measuring the resultant voltage as the current flows through a suitable, well-defined impedance, usually called a current shunt.

Direct techniques are exemplified by the torque or force produced on a current carrying coil in a magnetic field, such as in classical D'Arsonval meters and electrodynamic meters. Other approaches have used the force on an "iron vane" produced by a current carrying coil, the expansion of a "hot wire" carrying a current to move an indicator needle, and the heating of a thermocouple junction to produce a voltage. The latter case is arguably an indirect measurement, illustrating the somewhat arbitrary nature of the distinction.

Even electrochemical methods have been used to measure current, such as the transfer of mercury across an electrolytic gap in a glass capillary tube, or the mass of a deposited metal in an electrolytic cell after a fixed time interval. These might more accurately be considered amp-hour meters, as they measure the time integral of current.

Indirect techniques are typified by the use of a low-value resistive "shunt" to convert current into a corresponding voltage. Although resistances of any value might be used to measure small currents, shunt resistances are typically in the milliohm to microohm range for currents above one amp. The use of current shunts is increasingly common with the advent of modern Digital Volt Meters (DVM), which are essentially voltage-measuring devices.

CURRENT SHUNT DEVELOPMENT

Early moving coil meters were capable of measuring currents from microamps to several amps. For higher currents, a low-value "shunting" resistance was placed in parallel with the meter, bypassing a large and well-defined fraction of the current around the meter and increasing the measurable current. For large currents, the shunt would be mounted outside the meter housing for better cooling and to allow the shunt and meter to be physically separated for convenience.

It became relatively standard to design panel meters and shunts to operate with 50 mV at full meter scale, while precision "laboratory" grade shunts were more often 100 mV full scale. This standardization allowed the same meter movement to be used for nearly any current, requiring only a suitable shunt resistance and meter scale. Modern shunts for DVMs are usually made in values divisible by a power of ten, allowing current to be read directly from voltage simply by shifting the decimal point; for example, the conductance in a 100 $\mu\Omega$ shunt is 10 A/mV, or 10,000 A/V.

Current shunts may be used for both dc and ac currents, although the design of shunts for high-frequency (kHz to MHz) ac currents is more involved than for dc or low-frequency ac current.

DC CURRENT SHUNT DESIGN CONSIDERATIONS

The principal considerations in dc and low-frequency ac shunts are accuracy and stability, which are more difficult to achieve with low resistances. The power dissipation in high-current shunts is also a complication, as shall be seen.

Contact Resistances

Contact resistance is a major factor in defining and accurately measuring resistances below 1 Ω . Current shunts are usually designed with two connections for current, and two additional connections for sensing shunt voltage (or “potential”), as shown in Fig. 1, often termed a four-terminal “Kelvin” connection. Measurement errors due to the indefinite current contact resistances (R_1 and R_2 in Fig. 1) are eliminated, as only the voltage across the shunt resistance R_s is sensed. Voltage-sense contact resistances R_3 and R_4 form a resistive divider with the voltmeter, but these resistances are usually negligible, compared with the input resistance of a voltmeter, and are thus not a significant source of error.

Power Dissipation Effects

Although the voltage drop is relatively low on a current shunt, the power dissipation can be quite significant; for example, a 1000 A, 100 mV shunt will dissipate 100 W at the rated current. The resulting temperature rise can cause two sources of error: a change in shunt resistance, and a thermal electromotive force (emf), or voltage where dissimilar metals join (a thermocouple effect).

A net thermal emf can occur when the two ends of the shunt are not at the same temperature, resulting in a shunt voltage measurement error. A thermal emf may add to or subtract from the shunt resistive voltage, depending on the algebraic sign of the thermal emf relative to the resistive voltage with a dc current.

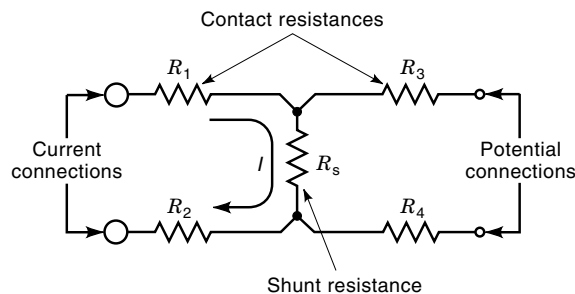


Figure 1. Separate shunt current and potential connections are used to minimize errors due to contact resistances.

These errors are much less when measuring ac currents, as the thermal emf adds to the resistive voltage during one half of the ac cycle, and subtracts during the other half cycle. The result is a complete cancellation of the thermal emf with an average responding ac voltmeter. If the dc component is not blocked to an rms responding voltmeter, the measurement error is still only 1% with a 10% thermal emf, and a 0.01% error with a 1% thermal emf component.

Selecting a Shunt Material

Pure metals have the low resistivity desirable for shunts, but the temperature coefficients of resistance (TC, or TCR) are high, ranging from about 3 m Ω /(Ω ·°C) for platinum to 6 m Ω /(Ω ·°C) for nickel (0.3%/°C and 0.6%/°C, respectively). Alloys of metals typically have higher resistivities but lower TCRs—occasionally much lower. The thermal emf can range from less than one to tens of μ V/°C. The approximate electrical properties of selected metals and alloys are given in Table 1; resistivity and TCR may vary significantly with impurities, alloy composition and degree of work hardening or annealing.

Nichrome is a common resistance alloy with high temperature capability and good corrosion resistance, but the TCR and thermal emf are both relatively high (see Table 1). Nichrome is also physically hard and difficult to work, and cannot be soft (tin/lead) soldered, but must be welded or hard (silver) soldered. Constantan has a much lower TCR than Nichrome, but the thermal emf effect is even stronger.

Table 1. Selected Properties of Electrical Metals and Alloys

Metal	Typical Composition	Resistivity $\mu\Omega \cdot \text{cm}(20^\circ\text{C})$	TCR ppm/°C	Thermal emf ^a eV/°C	Ref.
Copper	Cu100	1.7	3,900	+2.86	1
Silver	Ag100	1.6	3,800	+2.73	1
Aluminum	Al100	2.8	3,900	−0.42	1
Tin	Sn100	11.5	4,200	+0.20	1
Lead	Pb100	22	4,300	0.0	1
Solder	Sn63, Pb37	14.7			2
Brass	Cu60, Zn40	6.4	1,000	+0.8	1
	Cu70, Zn30	8.4	2,000		1
Nichrome	Ni58.5, Fe22.5	100	400	+21.3	1
	Cr16, Mn3				
Constantan	Cu60, Ni40	44.1	+2/+33	−39.9	1
Manganin	Cu84, Mn12, Ni4	45	+6/−42 ^b	+1.38	1
	Cu83, Mn13, Ni4	48.2	+15/−15 ^c		2
	Cu86, Mn10, Ni4	38.3	+15/−15 ^d		2

^aReferenced to lead.

^b12±100 °C.

^c15±35 °C (wire alloy).

^d40±60 °C (shunt alloy).

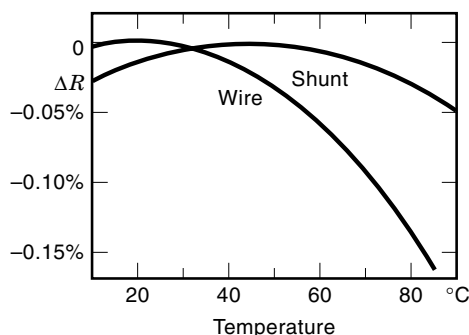


Figure 2. Manganin shunt alloy has a minimum resistance change at elevated temperatures to allow for significant power dissipation, while manganin wire is optimized for room temperature operation of “laboratory standard” reference resistors.

In 1889, Edward Weston discovered that alloys of copper, manganese, and nickel have very low temperature coefficients. Alloys with 10% to 13% manganese and 2% to 4% nickel are trade-named Manganin. Still the principal choice for shunts, Manganin is available from Harrison Alloys, Inc. in Harrison, New Jersey (800) 526-1256 and (201) 483-4800, and Carpenter Technologies in Reading, Pennsylvania (800) 694-6543 and (610) 208-2000. Typical changes in resistance with temperature for Manganin are shown in Fig. 2. A low positive TCR occurs at low temperatures, with a “turnover” to a negative TCR at elevated temperatures. Operation near the TC turnover point is desirable for the highest temperature stability.

Manganin resistance wire is used for precision resistor standards from about one ohm to thousands of ohms. Power dissipation is minimal in this application, so the alloy content is adjusted to have the minimum resistance variation around room temperature. On the other hand, current shunts are usually required to dissipate significant heat (as noted previously), so Manganin, for current shunt applications, has an elevated TC turnover temperature of 40 to 60 °C.

Manganin can be hard (silver) or soft (tin-lead) soldered, and has a thermal emf compatible with all of the metals and alloys commonly used in electronics: copper, brass, bronze, silver, gold, tin, and tin-lead solder. Proprietary alloys, with even lower TC, have been developed for high-precision shunts and resistance standards, such as Zeranin 43 (Cu87, Mn7, balance unspecified) by Isabellenhütte in Germany.

Shunt Resistance

The theoretical resistance of a shunt is calculated from the formula:

$$R_s = \rho l / A \quad (1)$$

where

R_s = shunt resistance;
 ρ = resistivity;
 l = effective shunt length; and
 A = total area of shunt.

This formula is only a close approximation when applied to the shunt material, as there is a small resistance contribu-

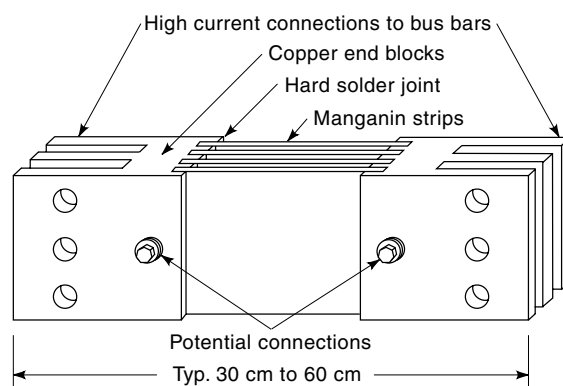


Figure 3. Shunts for currents of hundreds to thousands of amps are designed for mounting on one or more bus bars, with good heat dissipation to the bus bars and ambient air.

tion from the end blocks, leads, or other terminations. In addition, the resistivity of the shunt material will vary somewhat from published values, due to residual impurities and processing variations. Precision shunts are designed to have an initial value slightly below the desired value. A calibration adjustment to the desired value is achieved by carefully removing small amounts of shunt material, which decreases the effective area and increases resistance.

DC CURRENT SHUNT CONSTRUCTION

The construction of typical shunts is shown in Figs. 3, 4, and 5. The high current shunt of Fig. 3 is designed to be bolted directly to bus bars carrying hundreds to thousands of amps. Multiple strips of Manganin are mounted in parallel for low resistance, with spaces between strips for cooling air flow. Copper end pieces are used to minimize their contribution to resistance, and to help conduct heat from the shunt into the bus bars.

A typical 20 to 200 A meter shunt is shown in Fig. 4. The basic construction is similar to that of high-current shunts, but is physically smaller, and wire or heavy cable current connections are often used instead of bus bars. In this case, the shunt is mounted on an insulated base for mounting on a panel. Copper end pieces are preferred for higher currents, while brass may be used at lower currents.

Low-current shunts (below about 10 A) may be designed for printed circuit board (PCB) mounting, such as shown in Fig. 5.

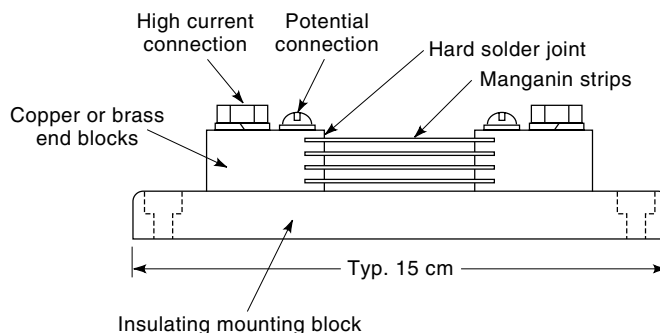


Figure 4. Current shunts in the 20 to 200 A range are typically designed for panel mounting.

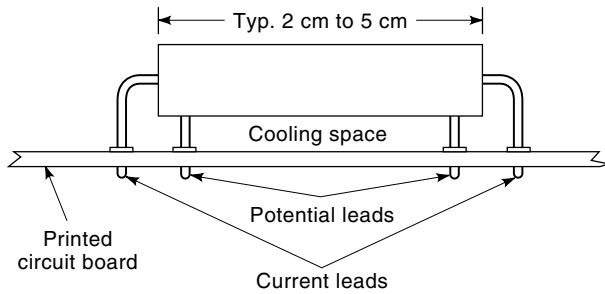


Figure 5. Low-current shunts for 1 to 20 A may be mounted on a printed circuit board. A cooling space between the shunt and board is recommended when power dissipation exceeds one watt.

Current shunts with only two terminals may be used for lower currents when high accuracy is not required; these may have wire leads and resemble a conventional resistor in all but resistance value. Low-value “chip resistor” shunts are also finding increasing use for surface mounting on a PCB or other substrate, and these are also often two terminal devices.

The accuracy of Manganin shunts can range from 1% to 0.001% or less. Obtaining the highest accuracies requires careful design, fabrication, annealing, and aging of the shunt, with operation restricted to a limited temperature range. Precision shunts may be forced-air cooled, in order to improve accuracy or current rating, although an oil bath is more common. The oil bath may be stirred and/or water cooled, in order to keep the temperature change to a few degrees or less.

HIGH-FREQUENCY AC CURRENT SHUNT DESIGN

There are two principal limitations to the accuracy of shunts at higher frequencies; inductive effects, eddy current skin, and proximity effects. Inductive effects are usually the dominant high-frequency error source, with eddy current effects only of concern in essentially “noninductive” current shunts at very high frequencies.

Shunt Inductance at High Frequencies

A typical parasitic inductance of 10 to 50 nH has little impact on the impedance of a 100 Ω or 1,000 Ω resistor below a few hundred megahertz. However, even 10 nH of equivalent series inductance (ESL) severely limits the higher frequency accuracy of a 10 m Ω shunt; the impedance increases by 1% at 22.6 kHz and 10% at 73 kHz. For wattmeter applications, the phase shift between shunt current and voltage is even more important, particularly for low-power-factor measurements, where the system voltage and current are nearly 90° out of phase. For the above shunt, a 1° phase shift occurs at 2.8 kHz, increasing to 10° at 28 kHz.

The effective ESL of a shunt is the sum of intrinsic inductance of the shunt and mutual inductance coupling between the current input and voltage output leads, although it is not always practical, or even meaningful, to try accurately to separate the two components of inductance.

Accurate current measurement with shunts at high frequencies requires that ESL or its effects be minimized; the use of wirewound resistors for low-current shunts should be avoided for this reason. The intrinsic inductance of conventional

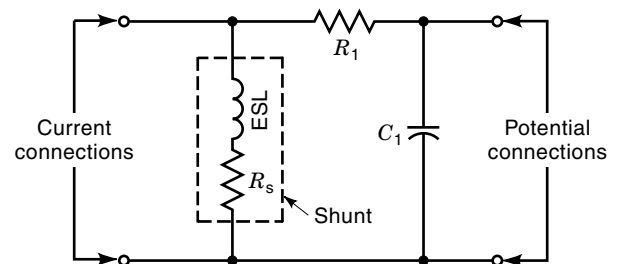


Figure 6. The high-frequency accuracy of a current shunt is typically limited by the equivalent series inductance (ESL). The accurate bandwidth can be extended by compensating the shunt's ESL/ R_s frequency response zero with an $R_1 \times C_1$ resistor-capacitor pole (see text).

low-frequency shunts (Figs. 3 and 4) is not readily reduced, but the larger ESL effects of mutual inductance between current and potential leads can often be reduced by a factor of two to ten, by bringing one or both of the potential sense leads back along the shunt body to a common point, and exiting from the shunt with twisted wires or coaxial cable. Minimizing the inductive area of the current leads would also reduce ESL, in principle, but is usually impractical or ineffective, due to the larger size and relative inflexibility of the conductors.

Shunt ESL Compensation

The effective bandwidth of a shunt can be extended significantly by compensating or canceling the inductive effects with either of two techniques. The first approach is to use a resistor-capacitor (R - C) network, as shown schematically in Fig. 6. The R_1 - C_1 network of Fig. 6 is used to create a response pole at the same frequency as the R_s -ESL zero of the shunt resistance and inductance, which occurs when

$$(R_1)(C_1) = \text{ESL}/R_s \quad (2)$$

This technique can extend the bandwidth of a shunt by orders of magnitude, although some care must be taken in the values of the compensating resistor and capacitor, to minimize other parasitic effects. Compensation resistor values in the range of 30 to 1,000 Ω are recommended; ESL of the resistor may limit bandwidth at lower values, while parasitic capacitance becomes a factor at resistances above this range. Using compensation capacitances below 47 to 100 pF increases sensitivity to loading by other parasitic capacitances, while the ESL of capacitors

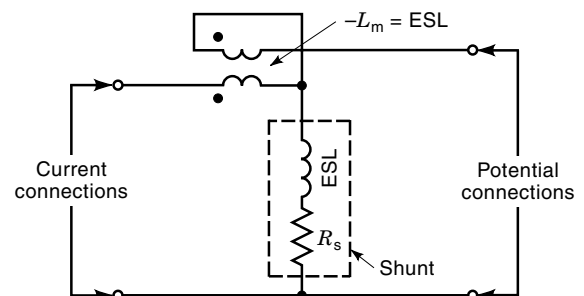


Figure 7. The high-frequency bandwidth of a current shunt can be extended by compensating or canceling the shunt's equivalent series inductance (ESL) with a negative mutual inductance ($-L_m$).

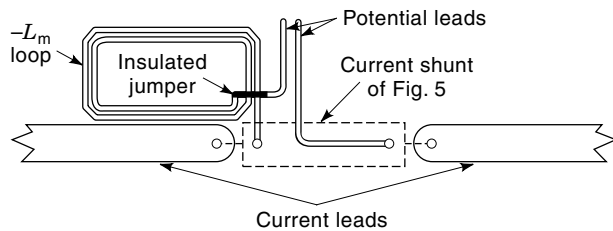


Figure 8. An example of $-L_m$ compensation of shunt ESL, achieved by routing of a potential sense lead back through the magnetic field near a current lead.

above about 1 nF may limit the upper frequency of accurate compensation. The ESL of leaded resistors and capacitors is typically 10 to 20 nH, while that of surface mount “chip” devices is significantly lower, at 1 to 5 nH. The parasitic capacitance of a 1/4 W leaded resistor is about 0.4 pF, while that of a 1206 size chip resistor is less, at about 0.1 pF.

An alternative approach to ESL compensation is to use negative mutual inductance coupling ($-L_m$) between the current and potential leads, as shown schematically in Fig. 7; when $-L_m$ is numerically equal to ESL, the effects of the shunt ESL are canceled. In this approach, the potential sense leads are arranged in the magnetic field of the current leads, in such a way as to create an induced voltage, which cancels the inductive voltage of the shunt. An illustration of this approach is shown in Fig. 8, for the printed circuit board mounted shunt resistor of Fig. 5.

The ESL of a shunt is sensitive to both shunt construction and current and voltage lead placement, so these must both be well controlled and consistent. The ESL compensation must then be calibrated, by comparing the output voltage with the input current at frequencies above and below the R_s -ESL zero frequency.

Although the frequency response of the shunt can be measured with various compensations, the simpler time domain approach shown in the test circuit of Fig. 9 is recommended. A voltage pulse from a generator is applied to a noninductive resistor (typically 50 Ω), which is in series with the (relatively low-impedance) shunt. The current in the 50 Ω resistor will have essentially the same waveform as the voltage, so the shunt compensation is adjusted to match the sensed shunt voltage waveform to the input voltage waveform. In the following discussions, the pulse will be assumed to have a smooth risetime and a flat top or plateau, although this is not essential for ESL compensator calibration.

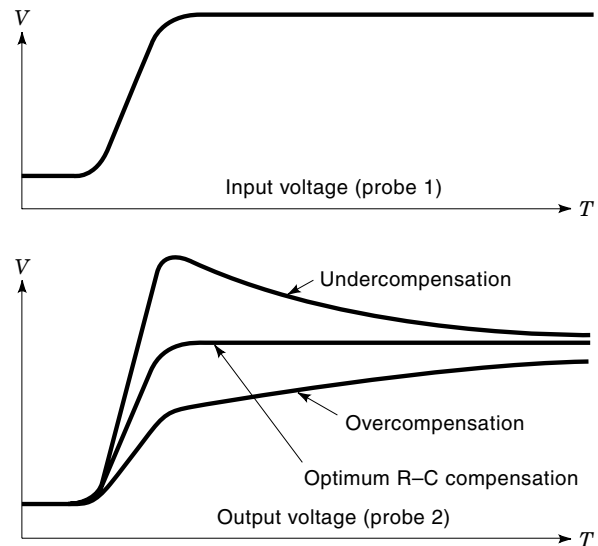


Figure 10. The effects on output voltage waveform of over- and undercompensation of shunt ESL with an R - C network (as in Fig. 6) are shown for a pulse input using the test setup of Fig. 9. The voltage and time scales are arbitrary.

Without compensation, the ESL of a current shunt will cause a large voltage overshoot during the pulse rise time. The overshoot may be tens to thousands of times higher than the steady-state voltage, depending on the shunt resistance, shunt ESL, and the pulse rise time.

When a shunt is undercompensated with an R - C network, the shunt voltage will still overshoot the desired value, as illustrated in Fig. 10, and then decay to the final voltage at the R - C compensation time constant. Overcompensation will cause the initial rise in voltage to be less than the desired value, with a longer rise to the final value, due to the greater R - C time constant. Correct compensation will produce a shunt voltage waveform essentially identical to the input pulse voltage waveform.

Full ESL compensation with negative mutual inductance also produces a shunt waveform equal to the input waveform, but under- and overcompensation produces different waveforms, as illustrated in Fig. 11. Undercompensation produces a similar shunt voltage overshoot as with R - C compensation, but the voltage falls back to the final value immediately after the input rise is completed. Significant overcompensation will cause the sensed shunt voltage to initially

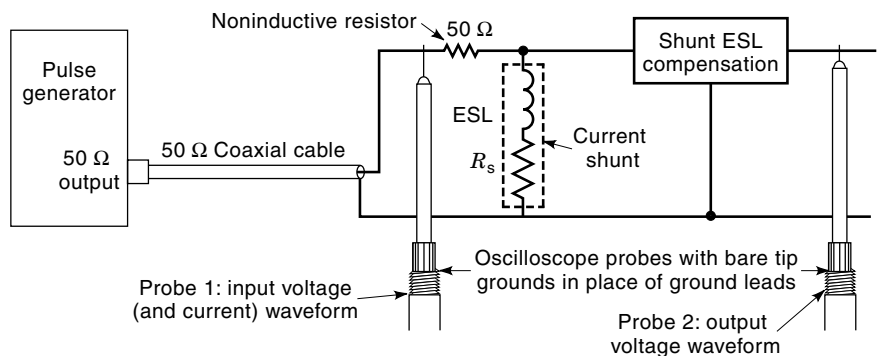


Figure 9. A recommended test setup for calibrating shunt ESL compensation, with either an R - C pole or negative mutual inductance. The compensation is adjusted to achieve the same waveforms on oscilloscope probes 1 and 2.

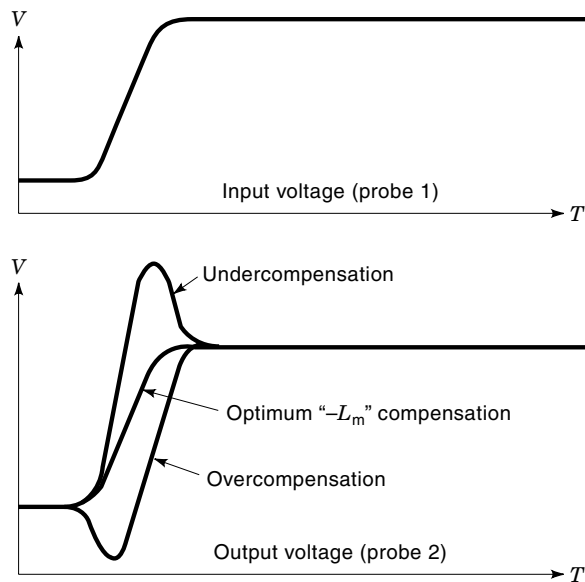


Figure 11. The effects on output voltage waveform of over- and undercompensation of shunt ESL with negative mutual inductance (as in Fig. 7) are shown for a pulse input using the setup of Fig. 9. The voltage and time scales are arbitrary.

undershoot, or swing in the opposite direction, with an immediate rise to the final value at the end of the input current rise.

Thus both resistor–capacitor and negative mutual inductance ESL compensation of a shunt can produce an output voltage waveform that matches the input current. The voltage waveforms due to over- and undercompensation however, are quite different.

Noninductive Shunts

Shunts can be constructed to appear noninductive at the sense voltage, although there is always some finite inductance in series with the current terminals. These shunts effectively utilize negative mutual inductance ($-L_m$) compensation, but in an automatic way, which does not require calibration.

Coaxial Shunts. Coaxial shunts are the more widely known noninductive shunt construction. A cylinder of resistive ma-

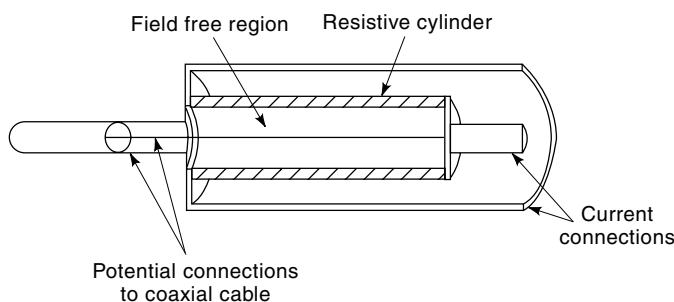


Figure 12. Cross-section of a “noninductive” coaxial shunt, which achieves automatic $-L_m$ compensation of the input ESL by bringing the potential sense leads out through a flux free region inside the resistive cylinder.

terial is incorporated in a coaxial arrangement with current and potential leads, with the potential leads brought out in a “flux free” region. One such construction is illustrated in Fig. 12.

The coaxial shunt has a nearly zero sense voltage ESL, as none of the magnetic field between the coaxial conductor and shunt material is coupled to the sense leads. However, the coaxial shunt can be inconvenient to use in many applications, particularly when coaxial conductors are not already in use. The two-terminal “insertion inductance” may also be significant (perhaps tens of nanohenries), partly due to the magnetic field between the coaxial cylinders, and partly due to the transition joint required to noncoaxial conductors.

Parallel-Plate Current Shunt. An alternative construction for a noninductive shunt is shown in Fig. 13. A relatively thin and annealed Manganin foil may be folded over at the midpoint, with thin insulation (not shown) between the foils. The ends of the foil are soft- or hard-soldered to copper current terminations, which may be left parallel or folded apart at right angles, depending on the application. Thicker or unannealed foils or strips will have to be soldered together, in place of the fold.

Narrower voltage sense lead foils are also soldered to the copper current terminations, and are brought back along the center of the Manganin foils or sheets, as shown in Fig. 13, with thin insulation between the sense leads and the Manganin. Closely spaced foil sense leads may be continued to the shunt voltage sensing point, or a transition made to coaxial cable or a twisted pair of wires.

The parallel plate current shunt can have a sense lead effective ESL nearly as low as the coaxial shunt; Shunts have been constructed with an ESL of less than 50 pH. The insertion inductance may be an order of magnitude lower than that of a coaxial shunt (perhaps several nanohenries or less) when used with noncoaxial conductors.

If the propagation delay along the sense leads of a high-frequency shunt approaches 6 to 8% of a cycle at the highest ac frequency, or a quarter of the 10 to 90% rise time with a pulse current, then transmission line termination techniques should be used to minimize reflections and the resultant waveform distortion and measurement error. If the voltmeter has a high input impedance, a “source termination” at the shunt is advised, using a series resistor equal to the transmission line impedance. This will be 50 or 75 Ω for most coaxial cable, and typically in the range of 100 to 200 Ω for a twisted pair of wires. If the voltmeter input impedance matches the transmission line impedance (typically 50 Ω for RF applications), a source termination is, generally, not necessary.

Eddy Current Effects in Shunts

Skin effect is classically thought of as the nonuniform current distribution of an isolated conductor at high frequencies, due to eddy currents induced in the conductors. In a round conductor at high frequencies, the current is uniform around the surface, while the current falls off exponentially beneath the surface. The “skin depth” is defined as the depth below the surface where the current amplitude is $1/e$ (36.8%) of the surface value. If all of the current were flowing uniformly in the first skin depth of thickness, the effective resistance would be the same. The skin depth varies as the inverse

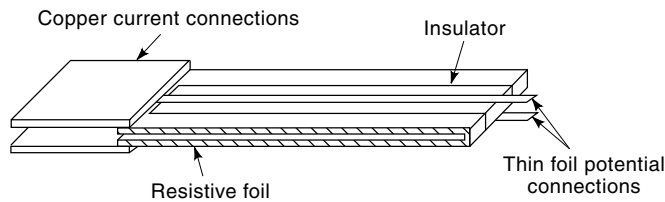


Figure 13. Construction of a “parallel plate” current shunt. The potential leads are brought out through a nearly flux-free region, to produce an essentially noninductive shunt.

square root of frequency, but also increases as the square root of resistivity.

When two or more conductors carrying high-frequency current are brought near to each other, the current distribution on the surface will change, which is termed a “proximity effect.” Currents flowing in opposite directions in two parallel cylinders tend to concentrate on the facing surfaces. The converse effect occurs when the currents flow in the same direction, as the currents begin to avoid the facing surfaces. At the risk of oversimplification, this effect can be envisioned as parallel currents “repelling,” while opposing currents “attract.”

The concept of skin effect as “the current distribution in an isolated conductor,” however, is misleading. For example, the high-frequency current in an isolated bus bar or thin ribbon is not uniform on the surface, but tends to concentrate on the outside edges and corners. The term “skin effect” should be limited to the exponential decay of current normal to the conductor surface; that is, the current distribution perpendicular to the conductor surface.

It is recommended that the current distribution around the conductor be considered a single-conductor proximity effect, caused by the mutual repulsion of parallel current filaments in the conductor. This conceptually explains the uniform surface current distribution in a round conductor and the nonuniform current in a flat conductor, where the current filaments “pile up” at the edges and corners.

The value of this distinction can be appreciated when relatively wide and thin sheets, plates, or bus bars are placed close together and carry high-frequency current in opposite directions, as in the parallel plate current shunt of Fig. 13. The proximity effects now force the high-frequency current to be uniform across the width of the sheets or plates (although only on the facing surfaces if the sheets are thicker than one skin depth).

The skin effect causes the impedance of coaxial or parallel plate current shunts to rise as the thickness approaches a skin depth. The excess impedance is a distributed resistance-inductance effect, asymptotically rising as the square root of frequency with the voltage leading the current by 45°. As such, the skin effect cannot be compensated by a single pole, but would require a distributed R - C network or the equivalent.

Thus, skin effect represents a more fundamental upper limit to a shunt’s useful frequency than ESL, and is best dealt with by using thinner foils. (The relatively high resistivity of Manganin, about 25 times that of copper, is useful in increasing skin depth and raising the frequency of skin effect impedance increases.) For a high-frequency current penetrating a flat or tubular conductor from one side (as in the parallel plate and coaxial shunts, respectively), the impedance increases about

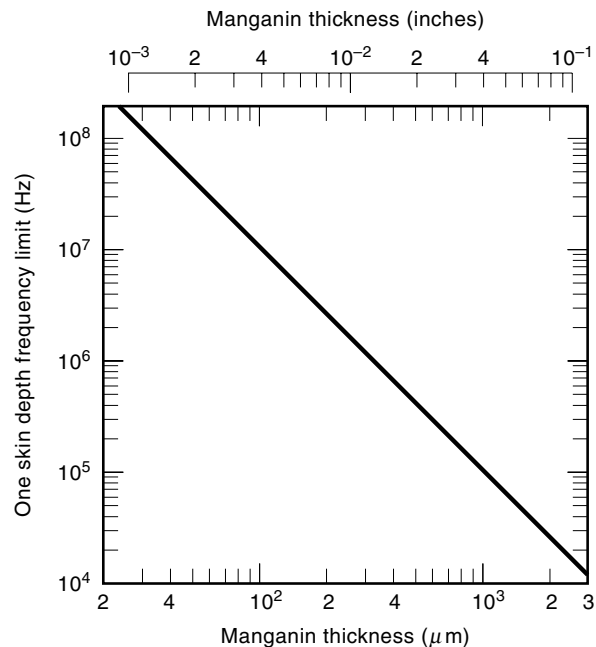


Figure 14. The maximum “accurate” frequency of a coaxial or parallel plate current shunt is limited by skin effects. The “one-skin-depth” frequency limit for Manganin increases rapidly with thinner material.

10% when the conductor is one skin depth thick. For Manganin sheets or tubes with current flow on one surface, the “one skin depth” limit on frequency versus Manganin thickness is given in Fig. 14.

BIBLIOGRAPHY

1. C. D. Hodgman (ed.), *Handbook of Chemistry and Physics*, 41st ed., Cleveland, OH: Chemical Rubber Publishing Co., 1960.

Reading List

- J. A. Ferreira, W. A. Cronje, and W. A. Relihan, Integration of High Frequency Current Shunts in Power Electronic Circuits, *IEEE Trans. Power Electron.*, **10**: 32–37, 1995.
- A. J. Schwab, *Hochspannungsmesstechnik*, New York: Springer Verlag, 1981.
- C. M. Johnsson and P. R. Palmer, Current Measurement using Compensated Coaxial Shunts, *IEE Prec.-Sci. Meas. Technol.*, **141**(6): 471–480, 1994.
- S. Sevénsson, *A Precision Wattmeter for Non-sinusoidal Conditions*, Chap. 5, School Electrical and Computer Engineering, Chalmers University of Technology, Technical Report no. 223L, 1996.
- W. Kohler, *A New Shunt for the Measurement of Low Level ac Currents*, Paper 64.01, Dresden, Germany; 7th Int. Symp. High Voltage Eng., August 26–30, 1991.
- O. B. Oliveira, M. T. Silva, and P. A. Chagas, *Design Construction and Performance Evaluation of Shunts for the Measurements of High Impulse Current*, Paper 64.02, Dresden, Germany; 7th Int. Symp. High Voltage Eng., August 26–30, 1991.

BRUCE W. CARSTEN
Bruce Carsten Associates, Inc.

CURRENT TRANSFORMERS

Current transformers (CT) are magnetic devices, which assist in the measurement of current. Different applications necessitate a wide variety of designs. Current transformers used in the ac power field, along with voltage or “potential” transformers, are generally known as instrument transformers, and are discussed only briefly here. Other types of CT discussed in this article are:

- high-frequency CTs;
- pulse current transformers;
- ac + dc current-sensing transformers;
- precision dc current-sensing transformers;
- Rogowski coils.

Classical ac CTs are used to change an ac current to a more convenient level for measurement, while isolating the ammeter from the current carrying conductor. Current transformers used in electrical power applications typically operate at frequencies of 50 or 60 Hz, but may range from 25 Hz to 400 Hz or so. Measured current levels may be less than an amp to thousands of amperes or more, and transformation accuracies may range from 0.1% or better for critical measurements, to several percent or more for general indication or overload detection.

Higher frequency ac current transformers are used in a number of other applications, including RF transmitters, induction heaters and electronic power converters to control output, monitor operation, or diagnose problems. Frequencies may range from hundreds of hertz to hundreds of megahertz, while current levels are usually lower at milliamperes to hundreds of amperes. Accuracy requirements are often not as stringent, often in the 1 to 10% range, although errors down to 0.1% are obtainable at frequencies up to the low megahertz band. When very wide bandwidths are required, active electronic circuits can be used to extend the low-frequency limit by several orders of magnitude, while improving accuracy at midband frequencies.

Ac transformers can be adapted to measuring pulse currents, finding use in electronic power circuits and pulse energy systems. They can operate with a small net dc current, as long as the current is zero for a sufficient interval between pulses.

Conventional ac current transformers cannot measure dc current, nor can they usually tolerate significant dc current without magnetic saturation of the core. The CT can be adapted to measure dc as well as ac current, with modifications and the addition of electronic circuitry, albeit at some potential sacrifice in available accuracy. These devices are sometimes termed current transducers, to distinguish them from “ac only” CTs. Some oscilloscope current probes also use this approach to provide broadband sensing from tens of megahertz down to dc.

A close relative of the current transformer is the Rogowski coil, which physically resembles a current transformer but (usually) without a magnetic core. The open circuit output voltage is proportional to the time derivative of the measured current, rather than the current itself. Lacking a core, the device cannot be saturated by large ac or dc currents. As such, it may find use where high currents are involved, or at very high frequencies, where magnetic cores become a hindrance to accurate measurement.

HIGH-FREQUENCY CURRENT TRANSFORMERS

High-frequency current transformers (HFCT) are similar in design and construction to an instrument CT, principally differing in the common use of ferrite cores, which perform as well or better than nickel-iron cores above about 100 kHz. Minimization of capacitive loading and parasitic coupling also become much more important at high frequencies.

The maximum initial permeability of ferrite is about 10 k, but this value holds to low flux densities and up to at least 100 kHz. This permeability is superior to that of conventional silicon steel at very low flux densities for any frequency. The permeability of nickel-iron can reach values of 100 k, but begins to fall at a few kilohertz to similar values for ferrite above 100 kHz. However, ferrite often finds use at frequencies well below 100 kHz, due to a significantly lower cost than nickel-iron.

Unlike ac line frequency CTs, HFCTs are often used over a very broad band of frequencies, typically three to six decades of frequency. HFCTs are often used to reduce the current to be measured, as are ac line frequency CTs, but not necessarily for identical reasons. The load or burden on a HFCT is usually a resistance (instead of an ammeter), to convert the current to a corresponding voltage, which is easier to measure at high frequencies. The shunt resistance required to measure currents above a few amperes is typically much less than 1 Ω , in order to limit power dissipation. The parasitic inductance of current shunts becomes problematic at high frequencies, as described in the article titled Current Shunts.

The shunt inductance problem can be greatly alleviated with current transformers, even with a moderate turns ratio. A CT with a ten-turn secondary and a one-turn primary decreases the primary current and increases the voltage by ten times on the secondary. The secondary side current shunt now has 100 times the resistance required on the primary for the same power dissipation. The useable frequency increases by the same factor for a given shunt inductance, and the higher voltage is usually easier to measure accurately. For HFCTs, the secondary side current shunt is usually known as the *CT load resistance*, and this term will be used hereinafter.

Many high-frequency CTs use a single-layer secondary winding on a toroidal core with a single-turn primary, as shown in Fig. 1. Often the primary is composed of the conductor carrying the current to be measured.

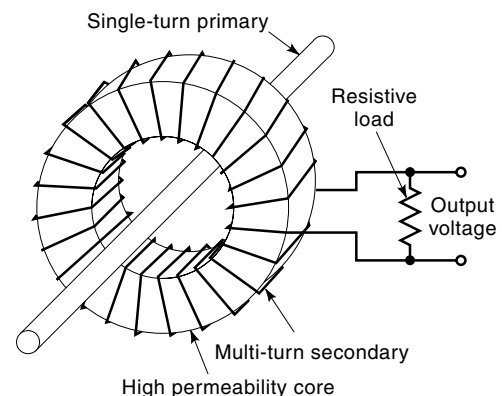


Figure 1. A high-frequency ac current transformer is typically used with a resistive load or “shunt” to measure current.

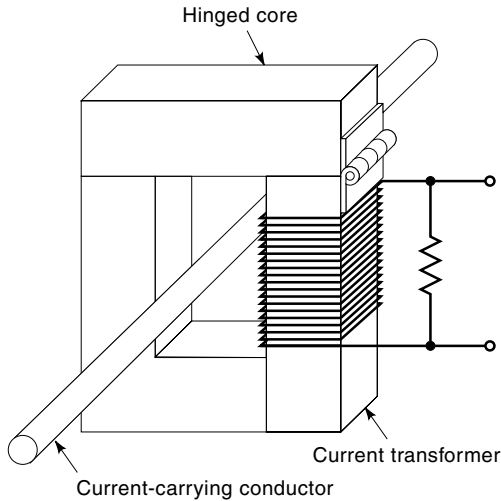


Figure 2. A clamp-on current transformer allows current measurement without breaking the circuit.

In some instances the temporary installation of a current transformer is desired when it is not feasible to break the circuit carrying the current to install the transformer. Clamp-on current transformers are made for these applications, with a split core for placing around conductors without having to open the primary circuit. A hinged version of a clamp-on transformer is shown diagrammatically in Fig. 2. Other configurations may use a split toroidal core with a hinge, or a portion of the core may be made to slide into place to complete the magnetic circuit. The inevitable air gap between core pieces tends to reduce the accuracy and raise the minimum useable frequency of clamp-on current transformers.

Current transformers will algebraically sum the currents in two or more primary windings. This ability can be quite useful and sensitive in comparing two nearly identical currents flowing in opposite directions, as the CT responds only to the current difference. This allows a comparison of two currents to within parts per million in precision laboratory test equipment, or the detection of ground fault currents of a few milliamperes out of 10 or more amperes in the humble household ground fault interrupter (GFI) breaker.

HFCT Error Sources

An equivalent circuit for a typical transformer is shown in Fig. 3, including the usual parasitic “components,” in addition

to the ideal transformer. The relative importance of the various parasitic effects are quite different for current transformers, when compared with other types of transformers.

Primary Resistance and Leakage Inductance. The primary winding resistance R_p and transformer leakage inductance L' have no direct impact on current transformation accuracy, as they are in series with the primary input current I_p . Their only effect is to increase the insertion impedance of the transformer, which may have an influence on the primary current if the insertion impedance is not insignificant, compared with the impedance of the rest of the circuit.

Core Magnetization Currents. The core magnetizing inductance L_m and parallel loss resistance R_m have been shown on the secondary side of the ideal transformer, instead of the more conventional primary side placement. This helps clarify the sources of error, as the secondary current I_s is still an “ideal” transformation of the primary current. The ideal secondary current now divides among four parallel paths:

1. the core magnetizing inductance L_m ;
2. the core's parallel loss resistance R_m ;
3. the parasitic secondary loading capacitance C_s ; and
4. the output load resistance R_l .

The currents in L_m , R_m , and C_s bypass current from R_l , and are direct sources of error in the CT. The magnetizing inductance is principally of concern at lower frequencies, below about 100 kHz for a ferrite core permeability of 10 k, and at proportionally higher frequencies if lower permeabilities are used. As long as core flux densities remain below saturation, the minimum operating frequency f_1 is defined by:

$$f_1 = \frac{1}{[2\pi L_m / (R_l + R_s)]} \quad (1)$$

which is the corner frequency below which the current in L_m exceeds that in the load resistance R_l . The transformation error at this frequency is large, about 30%, and the phase displacement is 45° . If an accurate transformation is desired, the transformer has to be designed for a much lower corner frequency. Lowering the corner frequency requires increasing the core area/magnetic path length ratio, increasing the number of primary and secondary turns, and/or using a higher permeability core if one is available.

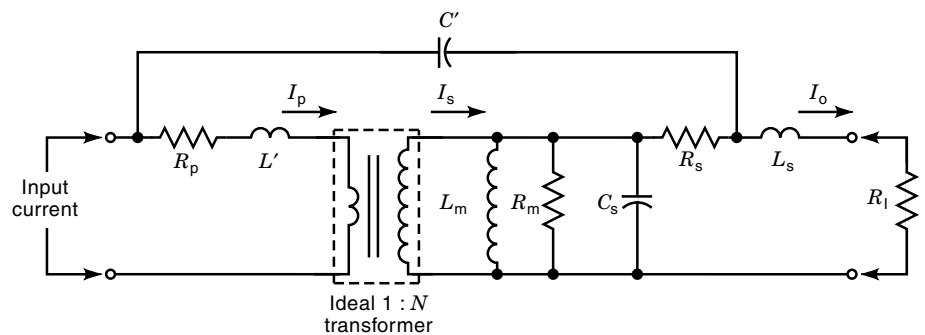


Figure 3. An equivalent circuit of a high-frequency current transformer is useful for identifying sources of error.

A second limitation to the minimum CT operating frequency may be core saturation, which occurs when the maximum secondary voltage experienced is too high for the number of turns used. Full saturation occurs at about 0.3 to 0.4 T in ferrites, but peak flux must be kept below 0.1 to 0.2 T (depending on material and operating temperature) for operation in the high permeability “linear” region of the core B - H loop. The number of turns required with a sinusoidal voltage is given by:

$$N = \frac{U}{4.44BAf} \quad (2)$$

where N is the number of turns; U is the rms voltage across L_m , in V; B is the allowable peak flux density, in T; A is the cross section of the core, in m^2 ; and f is the operating frequency, in Hz. If core flux is too high for the required secondary voltage and operating frequency, the flux can be reduced by increasing the core area and/or the number of turns, while keeping the desired primary/secondary turns ratio. A higher core permeability will not lower the frequency of core saturation for the same saturation flux density.

The parallel loss resistance of ferrites is roughly constant at high frequencies, so R_m represents a fairly consistent “broad-band” error current. This error can also be reduced by increasing the core area/magnetic path length ratio or increasing the number of turns, but a higher or lower permeability core will not significantly affect R_m .

Secondary Capacitance Loading. The upper frequency limit of a CT usually occurs when the impedance of the secondary parasitic capacitance loading C_s becomes less than that of the load R_l . Secondary capacitance is largely composed of winding layer-to-layer capacitance, secondary winding to core capacitance, and turn-to-turn capacitance, in that approximate order. Single-layer windings are often used to eliminate large layer-to-layer capacitances, although some space between the winding ends on a toroidal core must also be provided to reduce end-to-end capacitance. Extra insulation between the winding and a toroidal core is also often useful in reducing secondary capacitance. Turn-to-turn capacitance is usually insignificant in all but the highest frequency CTs with a few secondary turns, when spaced turns or wire with a thicker insulation may be beneficial.

The number of secondary turns tends to have only a moderate effect on secondary capacitance, which may increase or decrease, depending on other winding geometric factors.

Secondary Resistance and Inductance. As long as the secondary winding resistance R_s and the impedance of the secondary output inductance L_s are less than the load resistance R_l , the voltage drop on the output shunt largely determines the voltage on L_m , R_m , C_s , and their respective error currents. Only when the impedance of R_s and L_s become comparable to or larger than R_l do they have a significant or major effect on the currents in L_m and R_m and, hence, on the current transformation error. Increasing the number of turns decreases the magnetizing current errors, but the secondary resistance will increase at least as the square of the number of turns, and as the cube of the turns if the secondary winding is kept to a single layer. A point of diminishing returns is eventually reached when $R_s > R_l$, where more turns begin to degrade accuracy. If higher accuracy is still required, a larger core may be required.

Output inductance is produced by any secondary current flux, which is not coupled to the primary. It is due partly to secondary lead inductance, and partly to “air core” inductance of the secondary winding. The impedance of L_s is usually much less than R_s , and is usually not a significant problem, except at the highest frequencies with a large number of secondary turns.

Primary to Secondary Capacitance. The primary to secondary capacitance coupling C' is usually minimized by having a single-turn primary centrally located in a toroidal CT, with significant insulation thickness on the primary. As such, C' usually has an insignificant contribution to secondary loading capacitance C_s . However, significant error currents can flow through C' , if the CT primary and secondary windings are at a high ac potential with respect to each other, which usually occurs when the CT is placed on a primary conductor at a large high-frequency voltage. The worst case occurs with square or pulse voltage waveforms, as the displacement current is

$$I_c = C'(dV/dt) \quad (3)$$

where I_c is the instantaneous capacitance displacement current, in A; C' is the primary to secondary capacitance, in F; and dV/dt is the time derivative of the voltage, in V/s.

These capacitance currents can be surprisingly high with “square-wave” voltages. For example, a 600 V p-p pulse or square wave with a 30 ns rise or fall time will produce a 200 mA pulse of current through only 10 pF of capacitance.

If large primary ac voltages exist, the best solution may be to try to move the CT to a conductor carrying the same current, but at a lower ac potential. If this is not practical and the problem is not too severe, increasing the primary-secondary spacing and going to an insulation with a lower dielectric constant (ideally “air”) may suffice.

Otherwise, an electrostatic shield is required around the secondary winding, or at least between the primary and secondary winding. This shield is preferably tied to a “ground” point in the primary circuitry, although a connection to the ground side of the secondary may also serve. Superior isolation is achieved when two shields are used, with the one closer to the primary “grounded” to the primary circuit and the secondary shield tied to the “ground” side of the secondary.

It must be noted that the term “ground,” as used here, does not refer to an “earth” connection external to the circuit, but to a low ac potential conductor of the respective circuit. This is often a dc bus for the circuit, and may be known as the *ac quiet* or *RF low* conductor.

Extending HFCT Bandwidth Accuracy with Active Circuits

It has been shown that core magnetizing currents represent the fundamental limit to CT low- and medium-frequency bandwidth and accuracy. There are two ways that active electronic circuitry can be used to reduce these errors: the “virtual ground” load, and flux cancellation through feedback.

CT Virtual Ground. The CT load resistance is usually much larger than the secondary winding resistance, so the secondary current accuracy will improve as the load resistance is decreased to zero, dropping the voltage on the core’s L_m and R_m to a minimum limited by R_s (and L_s , if significant).

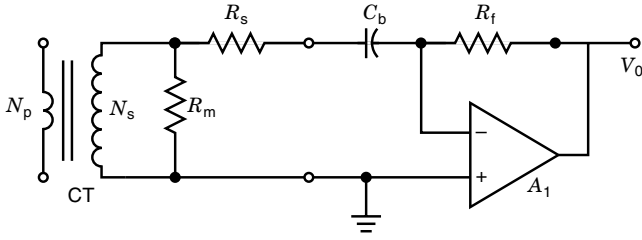


Figure 4. An operational amplifier (op amp) can be used to create a “virtual ground” on the secondary of a CT, improving accuracy and lowering the minimum usable frequency.

Unfortunately, the error in reading the shunt voltage tends to increase with lower voltages, due to system noise and other effects. One way to overcome this tradeoff is to use feedback to force the secondary output voltage to zero (creating a virtual ground on the secondary winding), while measuring the current required to achieve the virtual ground.

The simplest virtual ground load circuit uses an operational amplifier (op amp) A_1 , as shown in Fig. 4. The “ground” side of the CT secondary is connected to the positive (noninverting) input of A_1 , while the other side is connected to the negative (inverting) input. A feedback resistor R_f is connected between the A_1 output and the inverting input. The high gain of A_1 forces the voltage on the negative input to be essentially equal to the “zero voltage” on the positive input. The current required to do this flows through R_f , so the output voltage is, essentially,

$$V_o = I_s \times R_f \quad (4)$$

where V_o is the output voltage, in V; I_s is the secondary current, in A; and R_f represents feedback resistance, in Ω . In effect, the feedback resistor R_f becomes the CT load resistor, but the CT does not “see” this voltage on the secondary.

A blocking capacitor C_b may be required in series with the CT secondary, to avoid gain multiplication of the A_1 input offset voltage V_{os} . Without C_b , the dc output of A_1 would be V_{os} times R_f/R_s , while, with C_b , the dc output is simply equal to V_{os} . C_b must be a relatively large capacitor, to avoid limiting the low-frequency bandwidth; specifically,

$$(C_b R_s) \text{ great } L_m / R_s \quad (5)$$

where C_b is the dc blocking capacity, in F; R_s signifies secondary winding resistance, in Ω ; and L_m is the secondary winding magnetizing inductance, in H.

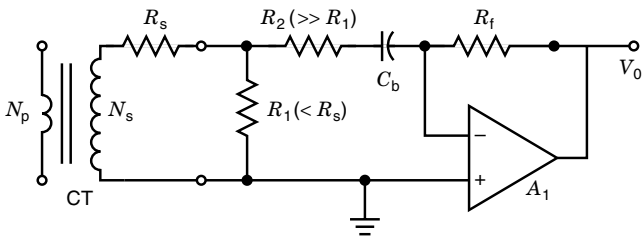


Figure 5. The operational amplifier current required to create a “virtual ground” on the secondary of a CT is reduced by the ratio of $R_1/(R_1 + R_2)$.

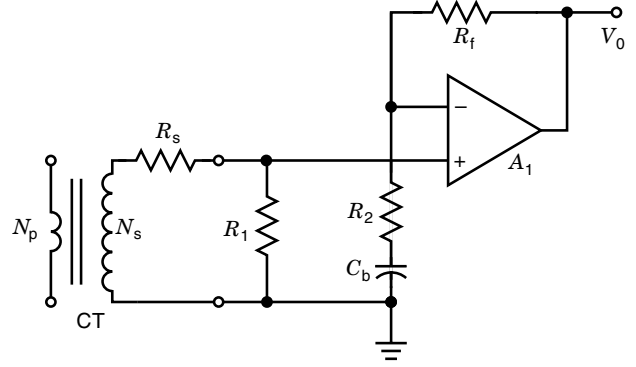


Figure 6. A noninverting buffer amplifier can be used with a CT shunt or load resistance of $R_1 < R_s$, to extend the low-frequency bandwidth.

The equivalent series resistance (ESR) of C_b should be less than R_s , and C_b should have an rms current rating equal to the full secondary current. A low-voltage electrolytic capacitor will usually suffice, which need not be a nonpolar type as the dc voltage will only be the input offset voltage of A_1 , which is usually less than 5 to 20 mV.

A drawback of this simple but effective circuit is that A_1 must be able to source the full CT secondary current through R_f , while the current capability of most op amps is limited to 10 to 100 mA. The feedback resistance R_f will usually be much less than the CT magnetizing resistance R_m , so a unity gain stable op amp is required, and the output HF bandwidth is limited by the gain–bandwidth of A_1 .

This limitation can be partially circumvented by the circuit of Fig. 5, where the op amp current is reduced by the current division ratio of $R_1/(R_1 + R_2)$. R_1 must be less than R_s , to avoid raising the L_F end of the bandwidth. The op amp A_1 need now only be stable at a gain of $R_f/(R_1 + R_2)$, which can increase the HF bandwidth (the actual voltage gain is R_f/R_2). An alternative circuit, with a noninverting gain of $(R_f + R_2)/R_2$, is shown in Fig. 6.

Flux Cancellation with Feedback. The low-frequency bandwidth of a HFCT can be extended even further, with circuits that sense and cancel the CT core flux through feedback techniques (1). A basic circuit that illustrates this principle is shown conceptually in Fig. 7. The HFCT now has a tertiary winding with turns N_t , which drive the inputs of high-gain amplifier

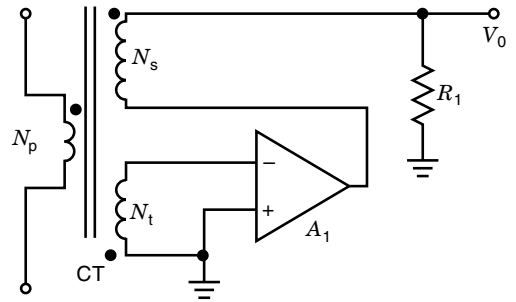


Figure 7. Using feedback to cancel the core flux of a CT extends the low-frequency bandwidth by orders of magnitude and enhances mid-band accuracy.

A_1 , while the output of A_1 drives a current through the secondary winding and load resistor R_1 . The high gain of A_1 forces a secondary current to flow, which almost exactly cancels the primary current, in order to hold the voltage on the tertiary winding to nearly zero. With essentially no current in the tertiary winding, and thus no $I \times R$ voltage drop, the core flux and magnetizing current errors are also forced to be nearly zero.

Whereas the circuits of Figs. 4, 5, or 6 may lower the LF end of the bandwidth by a factor of ten or more, flux cancellation can lower the minimum frequency by a factor of 1,000 or more. Both approaches also improve midband accuracy by similar factors and, in the case of flux cancellation, it becomes difficult to measure the actual midband CT error.

The circuit of Fig. 7 is not a practical circuit as shown. A dc feedback path must be provided around A_1 , and stability of the circuit must be ensured at both the lower and upper unity gain crossover frequencies. With “zero” magnetizing current the LF bandwidth could extend arbitrarily close to dc. In principle, the finite open loop gain of the op amp limits the minimum frequency, but, in practice, the maximum gain must be controlled, in order to achieve a stable unity gain crossover at the low-frequency end.

Ac current transformers tend to be 10 to 100 times larger than power transformers of the same “power” or volt-ampere rating, but active flux canceling allows the size of CTs to be reduced dramatically, particularly at lower frequencies. A practical limitation is that the op amp must again be able to supply the full secondary current.

A major advantage of the circuit of Fig. 7 is that the HF bandwidth is not necessarily limited by the bandwidth of A_1 . If the output of A_1 has a low impedance, and the input can tolerate the tertiary voltage, the circuit reverts to normal CT operation at frequencies greater than the bandwidth of A_1 .

PULSE CURRENT TRANSFORMERS

Pulse current transformers (PCT) may resemble ac current transformers, but the performance parameters of interest are somewhat different. The principle specifications for an ac CT are frequency range, maximum current, maximum burden or load resistance, and accuracy. Critical ratings for a pulse transformer are maximum peak current, rise time, pulse top droop, and the maximum current-time product for a pulse.

Low Duty Cycle PCTs

An ac current transformer with a resistive load can be operated with unipolar pulse currents, if the duty cycle is sufficiently low (often less than 1%) and other conditions are met:

1. The volt-seconds on the CT windings during the pulse must not cause the core flux to reach saturation.
2. The primary current must be zero between current pulses.
3. The interval between pulses must be long enough to allow the core to reset with the resistive load on the secondary.

Residual flux B_r in the core may become a significant portion of the saturation flux with some core materials, which reduces the pulse amp-second product. Occasionally a very small core “air gap” is introduced to reduce B_r . This also allows the PCT to

operate with a slightly greater amount of dc current, but at the expense of a similar drop in magnetizing inductance, which increases pulse droop. For applications with a significant dc current, a “dc compensation” winding must usually be added, to cancel the dc flux in the core.

This type of pulse transformer is usually also rated for high-frequency ac CT use. They are usually supplied with an internal resistive load to produce an output voltage proportional to current, and are typically terminated with a 50 Ω source impedance for driving coaxial cables (2) and/or 50 Ω load impedances. Clamp-on versions are also produced, which can operate over broad frequency ranges (3).

High Duty Cycle PCTs

In some applications, like switched mode power converters, it is desirable to operate a PCT with a relatively high duty cycle. This can be readily accomplished for unipolar pulses with the addition of a rectifying diode D_1 between the CT and the resistive load, as shown in Fig. 8.

During the current pulse, D_1 conducts to connect the CT to the resistive load. After the current pulse, D_1 conduction stops and allows the magnetizing inductance of the CT to “kick back” and generate a reverse voltage, which can reset the core very quickly if the kick back voltage is much higher than the forward voltage during the pulse. Pulse duty cycles of >90% can be accommodated with this simple modification to a CT.

A reset voltage clamp is also shown in Fig. 8, which prevents reverse breakdown of D_1 . The clamp typically consists of a Zener or avalanche diode Z_1 and a second diode D_2 , which blocks forward conduction of Z_1 during the pulse. Schottky diodes may be used for D_1 at high frequencies, partly to minimize additional secondary voltage when the CT resistive load voltage is about a volt or less, and partly for the high switching speed of Schottky diodes. The current rating of the Schottky diode should not greatly exceed that of the actual CT secondary current, as the high diode capacitance will tend to slow down the rise of the reset voltage. $p-n$ junction diodes for D_1 (and D_2) should be of the ultra-fast epitaxial variety, with reverse recovery times of 20 to 40 ns.

The voltage clamp is not always required for short current pulses, particularly for those of less than 1 to 10 μs duration. The maximum possible reset voltage on a PCT is limited by the peak magnetizing current in L_m (in Fig. 3), flowing through the core loss resistance R_m .

This core loss limit to the reset voltage also tends to limit the maximum pulse duty cycle with short pulses. However, an incomplete core reset after one pulse will leave a residual

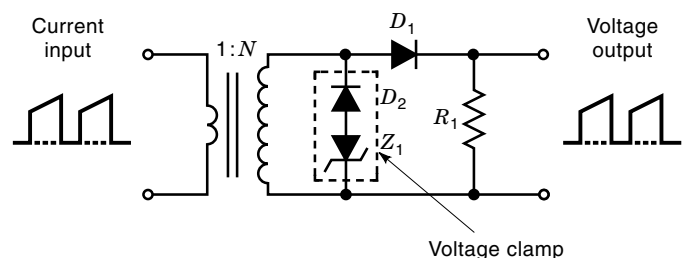


Figure 8. A CT secondary side diode allows unipolar pulse currents to be monitored.

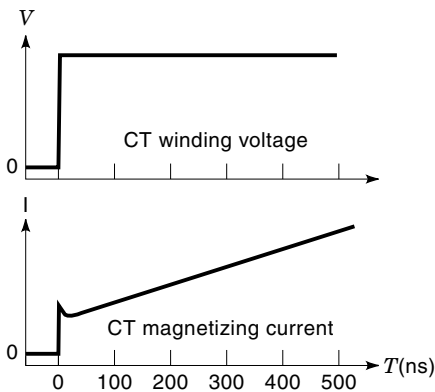


Figure 9. The pulse magnetization current of a ferrite has an initial step in current due to core losses. The effect is shown for a permeability of 10,000, where the core loss current exceeds the inductive magnetization current for about 200 to 300 ns.

magnetizing current in L_m , which adds to the reset current after the next pulse and increases the reset voltage. The residual magnetizing current may build up to a stable value before core saturation, which could be acceptable if the increase in magnetizing error current can be tolerated. Alternatives are to use a lower permeability core with more magnetizing current, or to provide a small core reset bias current.

The core loss resistance also creates an unexpectedly high magnetizing current for short pulses, as illustrated in Fig. 9 for a typical 10,000 permeability ferrite. For a pulse duration of 100 ns, the magnetizing current is four times higher than would be calculated from the low-frequency permeability of the ferrite alone. Under these conditions, the open circuit CT kick back voltage can be no higher than $\frac{1}{3}$ of the forward voltage during the pulse.

The high duty cycle PCT is not limited to unipolar pulse currents. The antiparallel diodes D_1 and D_2 of Fig. 10 allow sensing of bipolar current pulses, while still allowing for resetting of residual core magnetizing currents during the zero current interval between pulses. Alternatively, the bipolar input current pulses can be rectified to unipolar output voltage pulses with the circuit of Fig. 11.

The secondary volt-second mismatch between the two polarities of input pulse cannot be too great, or the core will not be able to reset between pulses with the reset voltage limited to a diode drop. More reset voltage is available with p - n diodes than Schottky diodes, and several diodes could be used in series to allow greater reset voltages. A drawback is that the diode voltage also adds to the load voltage and magnetizing current errors.

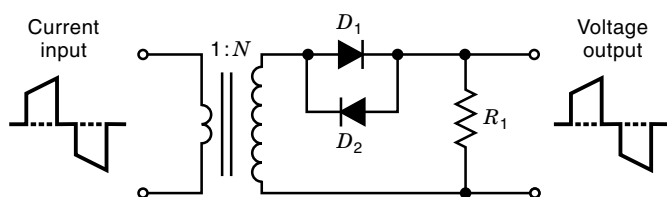


Figure 10. Antiparallel CT secondary diodes allow bipolar pulse currents to be monitored.

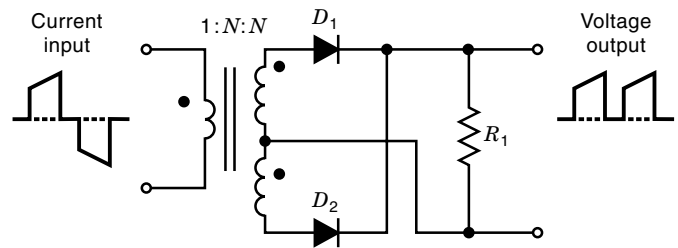


Figure 11. A center tapped current transformer secondary can be used to rectify bipolar current pulses to unipolar output voltage pulses.

It must always be borne in mind that PCT core resetting requires that the input current be zero for a finite interval. A pulse current waveform like that shown in Fig. 12 cannot be accurately sensed.

DC + AC CURRENT TRANSFORMERS

Dc current-sensing capability can be added to an ac current transformer by placing a Hall effect element in a small air gap in the CT core (4), as sketched in Fig. 13. A Hall effect element is typically a small semiconductor chip, with current input and voltage output electrodes on opposite edges, which is placed in a magnetic field with the flux normal to the surface. The output Hall effect voltage is proportional to the input current I_e and the flux density.

An amplifier A_1 senses the Hall voltage produced by a current in the primary, and produces a current in a secondary winding to cancel the core flux produced by the primary winding. Operation is similar to the flux canceling circuit of Fig. 5 in most regards, except that the flux-sensing tertiary winding is replaced with the Hall effect element. Since the Hall effect responds to dc as well as ac flux, the frequency response extends to dc. The necessary gap in the core of the CT, however, degrades the performance of the device. In practice, the accuracy is limited to a few tenths of a percent.

As in the flux-canceling circuit of Fig. 5, the bandwidth can exceed that of the amplifier, as the circuit can revert to conventional ac CT operation at high frequencies, which may range from tens of kilohertz to tens of megahertz.

Other designs eliminate the flux canceling winding in Fig. 13, and just amplify the output of the Hall effect element directly for current indication. The advantages are lower cost and power consumption, at the expense of reduced bandwidth and accuracy.

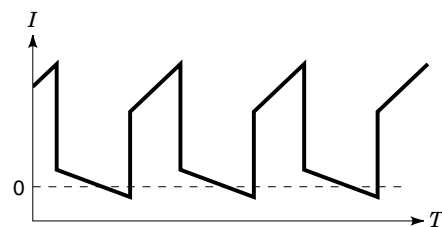


Figure 12. A pulse current waveform with no finite $I = 0$ interval, such as that shown here, cannot be monitored with a pulse current transformer.

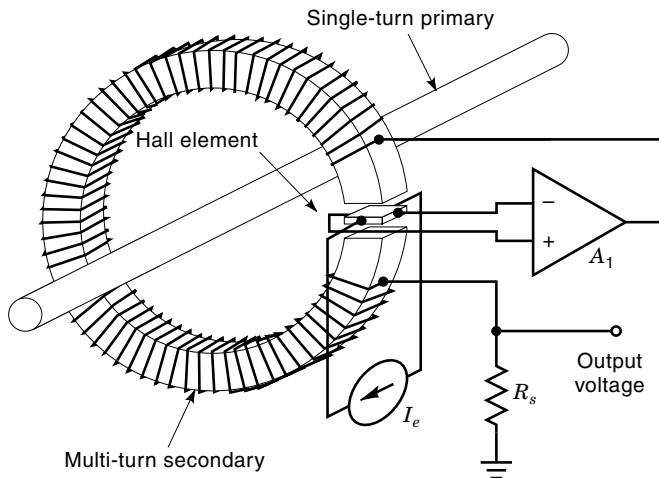


Figure 13. A Hall effect element is placed in an air gap in the core of a flux-cancellation CT circuit, to extend the low-frequency response to dc.

PRECISION DC CURRENT TRANSFORMERS

Although the conventional CT cannot measure dc current directly, a number of ingenious techniques have been developed (4–7) which use magnetic transformers to sense and measure dc current, with accuracies of parts per million in some cases. Typically, these approaches use two to four magnetic cores, one or two of which are driven to the verge of saturation or beyond by an ac source. A winding with the dc current to be measured links all cores, as does a dc compensation or canceling winding. Any dc amp-turn imbalance between the measured and compensation windings causes an asymmetrical response in the cores with an ac flux, which is sensed by a feedback and control circuit which adjusts the compensation current to remove the imbalance. The measured current is known from the compensation current and the turns ratios of the windings. Some of these circuits are suitable for measuring ac as well as dc current. A detailed description of these specialized approaches is beyond the scope of this overview article.

AIR-CORED CURRENT TRANSFORMERS (ROGOWSKI COILS)

The construction of classical Rogowski coils (8) is similar to that of ac current transformers, but with a nonmagnetic or “air” core. The “secondary” winding is usually in the shape of a toroid, with the return conductor brought back inside the winding to minimize spurious response to ambient magnetic fields, as shown in Fig. 14.

The output of the Rogowski coil is proportional to the rate of change of the enclosed magnetic field, or to the dI/dt of a conductor through the middle of the toroid. The response of a Rogowski coil is theoretically uniform to any current inside the hole of the toroid, and zero to any current outside the toroid. In practice, the response is usually slightly nonuniform to internal currents, with a slight spurious response to external currents, depending on the uniformity of the coil winding. These nonideal responses are typically within a few percent for well-constructed coils.

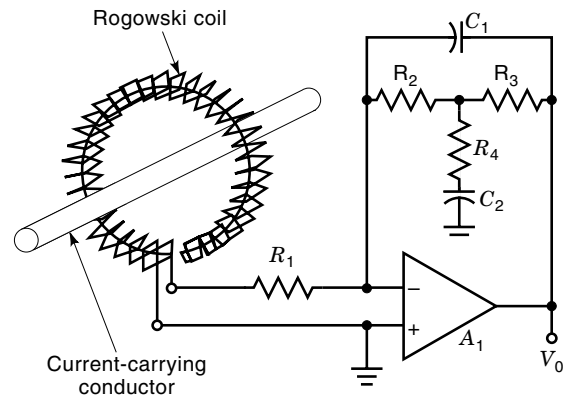


Figure 14. The output of a Rogowski coil is proportional to the derivative of the sensed current, and requires an integrator to reconstruct the primary current waveform.

The output voltage may be used directly to monitor the amplitude of sinusoidal currents of a constant frequency, but the voltage is usually integrated to produce a flat response over a range of frequencies (9,10). If phase shifts in the integrator are well controlled, the time domain fidelity of the waveform can be within one percent over a broad range of frequencies (11).

Rogowski coils have several advantages over conventional CTs. Although they cannot measure dc current, their performance is not affected even by very large dc currents. They cannot be saturated by large ac currents, and actually become increasingly advantageous as ac current increases. Conventional CTs require secondary turns to increase as measured currents rise to keep the output current constant, while Rogowski coils require fewer turns at higher currents, to produce a constant output voltage. Thus Rogowski coils can be much smaller than CTs at high current levels.

Magnetic Cores for Rogowski Coils

Rogowski coils tend to become less useful at low currents and low frequencies, where an excessive number of turns becomes required to produce a useful output voltage. This can be somewhat overcome by using a magnetic core, which increases the output in proportion to the permeability. However, the effective permeability must be stable and well defined, and insensitive to frequency, temperature, and dc and ac flux density over the operating conditions. This stability does not occur with solid ferromagnetic cores, but can be achieved with a core (such as ferrite) with one or more discrete air gaps. Unfortunately, the discrete gaps will increase sensitivity to external fields and primary conductor placement, unless a large number of equally spaced air gaps are used.

The best alternative is to use a “distributed gap” material. The most suitable types are molypermalloy powder (MPP) cores, which are available in permeabilities of 14 to 300. Achievable stabilities vs. flux range from better than $\pm 1\%$ at $\mu = 14$ to about $\pm 4\%$ at $\mu = 300$, although the initial standard tolerance of $\pm 8\%$ on permeability usually requires calibration for reasonable accuracy. The -3 dB permeability rolloff frequency ranges from about 600 kHz for $\mu = 300$ to about 20 MHz for $\mu = 14$.

Powdered iron cores are not recommended, in general, because they typically have relatively low permeabilities and

exhibit large changes in permeability with ac flux density. Possible exceptions are the unannealed carbonyl iron powder cores, which have stable permeabilities of about ten or less, but are useable to hundreds of megahertz.

Note that the output voltage of a CT (for a given load resistance) is proportional to N_p/N_s , while the output voltage of a Rogowski coil is proportional to $N_p \times N_s$. The sensitivity to low currents with either device can be increased by increasing the number of primary turns.

Design of Rogowski Coils

If the secondary inductance of the Rogowski coil is known, the output voltage for a sinusoidal primary current is

$$V_o = 6.28 f L_s I_p (N_p/N_s) \quad (6)$$

where V_o = output voltage, in V; f = frequency, in Hz; L_s = secondary inductance, in H; I_p = primary current, in A; N_p = number of primary turns (usually $N_p = 1$); and N_s = number of secondary turns.

For an arbitrary current waveform

$$V_o = L_s (N_p/N_s) (dI_p/dt) \quad (7)$$

where V_o represents the instantaneous output voltage, in V and dI_p/dt is the time derivative of primary current, in A/s. The inductance of a toroidal winding (or any cored inductor) is

$$L = (1.257 \times 10^{-6}) (N^2 A_e \mu) / l_e \quad (8)$$

where L = inductance, in H; N = number of turns; A_e = magnetic core area, or winding area for air core, in m^2 ; μ = relative core permeability ($\mu = 1$ for air); and l_e = magnetic path length, in m.

Replacing L in Eqs. (6) and (7) with the formula in Eq. (8),

$$V_o = (7.896 \times 10^{-6}) (f I_p N_s N_p A_e \mu) / l_e \quad (9)$$

for sinusoidal currents, and

$$V_o = (1.257 \times 10^{-6} N_s N_p A_e \mu) (dI_p/dt) / l_e \quad (10)$$

for an arbitrary primary current waveform.

BIBLIOGRAPHY

1. R. Stark, M. Schwendtner, and G. Steinmuller, *Electronic measuring instrument arrangement for measuring electrical a-c quantities*, US Patent No. 3,916,310, 1975.
2. C. Waters, Current transformers provide accurate, isolated measurements, *Power Conversion & Intelligent Motion*, Ventura, CA: Intertec International, **12** (12): December 1986.
3. C. Waters, Clamp-on current monitors achieve wide bandwidth, *Power Conversion & Intelligent Motion*, Ventura, CA: Intertec International, **20** (5): May 1994.
4. *Isolated current and voltage Transducers: Characteristics—applications—calculations*, Geneva, Switzerland: LEM Corporate Communications, 1996.
5. *Guideline Precision Instrumentation*, Smith Falls, Ontario, Canada: Catalog of Guideline Instruments, Ltd., 1980.
6. K. Nakamura et al., *Direct current detecting device using saturable reactors*, US Patent No. 4,286,211, 1981.
7. J. R. Leehey, *Current Sensor*, US Patent No. 4,482,862, 1984.
8. W. Rogowski and W. Steinhaus, Die messung der magnetische spannung, *Arch. Electrotech*, **1**: 141–150, 1912.
9. W. F. Ray and R. M. Davis, *Improvements in Current Measurement*, U.K. Patent GB 2,259,150 B, 1995.
10. W. F. Ray and R. M. Davis, Wideband ac current transducer measures high current, *Power Conversion & Intelligent Motion*, Ventura, CA: Intertec International, **20** (9), September 1994.
11. B. Carsten, Integrator for inductive current sensor, US Patent No. 5,539,354, 1996.

BRUCE W. CARSTEN
Bruce Carsten Associates, Inc.

INSTRUMENT TRANSFORMERS

An instrument transformer (IT), like all ac transformers, consists of a ferromagnetic core on which a primary winding and a secondary winding are applied. The windings may be isolated from each other (double wound), or one winding may be a tapping off the other (auto wound). The basic difference between a power transformer and an IT is that the latter is used with little or no load connected to its secondary winding, and, as such, an accurate transformation of voltage or current can be attained.

In the electrical supply industry, ITs are widely used to extend the range of an instrument such as a voltmeter, an ammeter, a wattmeter, a watthour meter, or an item of protection or control equipment. In addition to its use as a range-extending device, a double-wound IT may also serve to isolate an instrument from a high-voltage circuit.

It is usually impractical to manufacture precision indicating instruments with voltage ranges above about 500 V, or current ranges above about 20 A. As the electrical supply industry is required to measure or detect voltages and currents very much higher than these values, a convenient way to do so is to use an IT to transform such voltages and currents down to measurable values. Also, an IT may be required to transform low voltages or low currents up to values which can be accurately measured on available indicating instruments.

TRANSFORMER RATIO

An IT used as a ratio standard is probably the most accurate and stable of all electrical ratio devices. To a first order of accuracy, the ratio is dependent only on numbers of turns wound on the ferromagnetic core.

In the case of a voltage transformer (VT), the external load (known as the *burden*) connected to the secondary terminals must be of high impedance. When this is so, a voltage applied to the primary winding causes a voltage to be induced in the secondary winding that is nearly equal to the applied voltage divided by the ratio of primary to secondary turns. In other words, the voltage on each winding is nominally proportional to number of turns.

In the case of a current transformer (CT), the external load connected to the secondary terminals must be of low impedance. When this is so, a current passed through the primary winding causes a current to be induced in the secondary winding that is

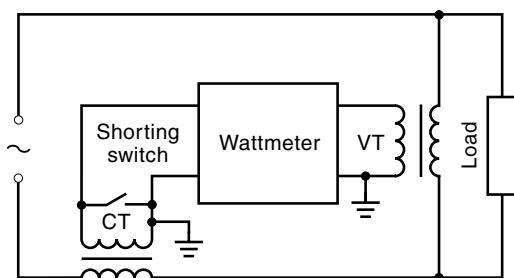


Figure 1. Use of instrument transformers to extend the range of a wattmeter. The voltage transformer and current transformer avoid the need to apply high voltages and currents to the wattmeter itself.

nearly equal to the primary current multiplied by the ratio of primary to secondary turns. In other words, the current passing through each winding is nominally inversely proportional to number of turns.

USE OF INSTRUMENT TRANSFORMERS

When a VT is used to extend the range of a voltmeter or voltage circuit of a wattmeter, the voltage to be measured must be connected to the primary winding of the VT. A connection must then be made from the secondary terminals of the VT to the voltmeter. Where the VT is double-wound, it would be normal (for safety reasons) to connect an earth to one secondary terminal. Any exposed metal casing of the VT must always be earthed.

When a CT is used to extend the range of an ammeter or current circuit of a wattmeter, the circuit must be opened at the point where the current is to be measured, and the primary winding of the CT must be connected across this opening. A connection must then be made between the secondary terminals of the CT and the ammeter. Where the CT is double-wound, it would be normal (again for safety reasons) to connect an earth to one secondary terminal. For additional safety, a short-circuiting switch is usually connected across the secondary winding of a CT. This should be kept closed until all connections are complete. The reason for this is that if appreciable current is passed through a primary winding of a CT with an open circuit on its secondary, dangerously high peak voltages are likely to be induced in the secondary (see Fig. 1).

VOLTAGE TRANSFORMER CONSTRUCTION

A VT consists typically of a closed, oil-filled metal tank into which the core together with its windings have been inserted. The winding terminations are typically brought out to terminals through porcelain insulators which are secured to the top of the tank (see Fig. 2). In most parts of the world, the secondary-voltage rating is standardized at 110 V. A VT may have more than one primary- or secondary-voltage tapping if it is intended for multiple ratio use. VTs intended only for laboratory use may incorporate Bakelite or other plastic insulators; or, instead of being oil-filled, their whole assembly may be encapsulated in an epoxy molding. VTs rated at voltages below about 1 kV are rarely oil-filled or epoxy-encapsulated.

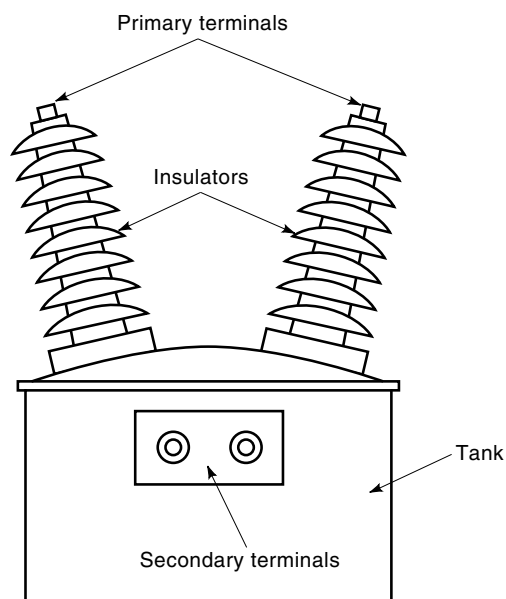


Figure 2. Typical voltage transformer in a metal tank, with outdoor-type primary-terminal insulators.

A VT may be designated according to international standards (1) as “unearthed,” in which case two equal-size insulators (as in Fig. 2) will be used for its primary terminations. Such a VT will be suitable for connection between phases of a three-phase system. A designated “earthed” VT will be designed for one end of its primary winding to be suitable for connection to a potential close to earth only.

During manufacture, so as to grade the insulation most efficiently, it is most common to apply the lowest-voltage winding of a VT closest to the core and then work outwards to the higher-voltage windings.

Most modern VTs, except those designed for special laboratory use, incorporate a ferromagnetic core made from cold-rolled silicon iron (also referred to as *grain-oriented silicon steel*, or GOSS). This ferromagnetic material is chosen because it is relatively cheap and has a high saturation flux density of about 1.8 T. The use of a more expensive, high-permeability, nickel-iron core could result in the VT having lower errors of transformation but, with a saturation flux density of only about 0.4 T, might cause the VT to be prohibitively large in size. A third alternative core material, known as *amorphous metal* or *glassy metal*, can be formulated to combine high saturation flux density with high permeability, but this material is usually too expensive for ITs.

The actual core construction of a VT typically consists of laminations as in a power transformer, with the windings applied on cylindrical formers through which the laminations pass. Such a construction is usually referred to as a “shell iron” VT.

Where specially high accuracy of transformation is required, a ring-type core might be used, where the core is wound up from a long length of strip as in a clock spring. Then the windings must be applied toroidally round the core. The advantage of such construction is that the windings are likely to be coupled more closely, and any component of error due to flux leakage will be minimized. Also, as the iron is used more efficiently, a lighter construction should result. Such construction (usually referred to as “toroidal”) is more expensive than the use of laminations.

CAPACITIVE VOLTAGE TRANSFORMERS

VTs with voltage ratings greater than about 150 kV would often become too large and expensive to be manufactured by conventional means. In such cases it is common to use what is called a *capacitive voltage transformer* (CVT). Such a device is constructed in two parts. The first part consists of a capacitive divider with a low-voltage tapping at about 10 kV. The second part (known as the *electromagnetic unit*) consists of a conventional double-wound VT with its primary winding connected across the low-voltage tapping of the capacitive divider. An additional feature of the CVT is that the input inductance of the electromagnetic unit is arranged to resonate (at rated frequency) with the capacitive divider. This in turn enables the capacitive divider to be much smaller in size than would otherwise be the case for a given accuracy. A disadvantage of a CVT is that its errors are frequency dependent.

INDUCTIVE VOLTAGE DIVIDERS

A special type of VT, intended only for laboratory use, is known as an *inductive voltage divider* (IVD). An IVD is basically a step-down autotransformer with multiple secondary tapings arranged in decades. Intended for very high accuracy use, IVDs are made with the highest-permeability cores. Also, all windings of each decade are stranded together in a bunch or rope so that they are very well coupled with each other. A typical IVD has six decades so that the input voltage can be subdivided to a resolution of one part per million. IVDs are rarely made for an input voltage rating much greater than 100 V at 50 Hz.

CURRENT TRANSFORMER CONSTRUCTION

As the magnetizing forces (ampere turns) generated by the primary and secondary windings of a CT are nearly equal and are opposite in phase, the net ampere turns—and hence the working flux density—are normally very low. Also, as the permeability of cheap, silicon-iron cores drops away to almost zero at very low flux density, it is normal to construct CTs intended for accurate measurements with nickel-iron cores, which retain their high permeability down to very low excitation. However, protection CTs, where freedom from saturation of the core under overload conditions is more important than accuracy of transformation, usually incorporate silicon-iron cores.

Care needs to be taken when handling nickel-iron cores because mechanical shock is liable to greatly reduce their permeability. For this reason, many manufacturers will enclose their cores in a protective case so that the pressure of the windings themselves on the core do not cause loss of permeability.

In most parts of the world, the secondary-current rating for CTs intended for measurement use is 5 A, while for protection CTs it is 1 A. Laboratory standard CTs often have secondary tapings of both 1 A and 5 A, although the latter is more common.

The normal method of constructing a CT is to apply the secondary winding first, so that it couples the core closely. Thus, secondary-winding leakage-reactance is minimized. Although it is also desirable that the primary winding closely couples the core, leakage reactance in the primary, no matter how large, is

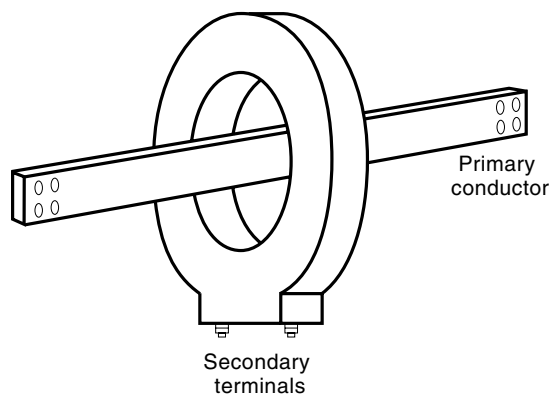


Figure 3. Typical current transformer for insertion onto a busbar, where the busbar itself forms a single-turn primary winding.

simply in series with the wiring to the load and is of no consequence as far as errors of the CT are concerned.

The most common type of CT encountered in the electricity supply industry is the “window” or “through” type, where a secondary winding is applied on a toroidal core, and the aperture of the core (the window) is left open so that one or more cables can be fed through it by hand to form the primary winding (Fig. 3).

When a window-type CT is used on a high-voltage transmission line, the clearance in the aperture must be large enough for there to be no possibility of a breakdown between the primary and secondary windings. Alternatively, solid insulation must be applied in the aperture.

In the highest-quality CT designs, terminals are made from plain copper, and all internal connections are soldered. Many manufacturers, however, use nickel- or chromium-plated terminals, often with detrimental effects. Research has shown (2) that the hard oxides on these plating materials can cause partial rectification at the contact faces, which in turn causes dc to flow in the windings. This component of dc will cause the permeability of the core to change or even cause saturation of the core. Plain, copper-to-copper connections, however, even when badly oxidized, rarely cause any trouble. Plain brass is a reasonable alternative metal to use for terminals, although when oxidized and poorly tightened, brass-to-copper connections can cause partial rectification.

TRANSFORMER ERROR

Due to various losses, the actual transformation ratio of an IT is never exactly equal to the ratio as defined by numbers of turns. Therefore, where the highest accuracy of measurement is required, the error in the transformation ratio must be determined and allowed for. The following are six effects which may give rise to errors in ITs, or which may cause the actual ratio to be unstable.

Effect of Loading

In a well-designed IT, the greatest contribution to error is that due to loading. The larger the volt-amperes delivered to the external burden connected to the secondary terminals, the greater

will be the losses internal to the IT, and hence the greater the errors.

Effect of Insulation-Leakage Currents

In applications where the voltage applied to an IT is high, or where a high voltage is induced in one of the windings, insulation-leakage currents may contribute appreciably to the errors. ITs that fall into this category are VTs with voltage ratings above about 30 kV, CTs with current ratings below about 0.5 A, and ITs operating at high frequency. In most cases it is the capacitance of the insulation material that has the major effect. Up to a point, capacitance leakage can be a good thing, because this current tends to cancel out the magnetizing current, which in turn, particularly in a CT, can reduce the component of error due to loading. However, excessive capacitance leakage currents can become a serious liability as far as the errors of the IT are concerned.

Effect of Partial Coupling and Flux Leakage

When the turns of the primary and secondary windings do not closely couple each other, such as when one or both windings are not uniformly distributed around the circumference of the ferromagnetic core, some lines of magnetic flux generated by one winding will not link all turns of another winding. This in turn will fractionally change the effective turns ratio. In a VT, this effect usually gives rise to the secondary voltage being less than nominal, while in a CT it usually gives rise to the secondary current being greater than nominal. This effect is usually associated with VTs, where, for ease and cheapness of construction, the primary and secondary windings are not concentric with each other or with window-type CTs. This effect should not be confused with the circumstances which might arise in a window-type CT where leakage fluxes become so large that, in some regions of the core, local saturation occurs. This in turn will cause the secondary current to become much lower than nominal (3).

Turns Ratio Error

If, during manufacture, the number of turns on any winding is wrongly applied, this will give rise to an error in the actual ratio.

Bad Contacts

In an otherwise-well-designed IT, errors can occur due to bad contacts in the internal or external circuit. These errors are usually associated with CTs where partial rectification at contact faces of connections causes a small direct current to flow in one of the windings. This in turn causes the permeability of the core to change temporarily, and hence causes a change in the errors. The slow deterioration of bad internal connections of an IT can take tens of years before they start to cause serious instability.

Partial Discharges

As a result of air voids or bubbles that may be present in the insulating materials of an IT, ionization of the air in these voids can occur when high enough voltages are present. When the ionization voltage is reached, alternate breakdown and

recovery of the air path in the void will rapidly occur. This effect is known as *partial discharges* (PDs). As a result of these PDs, a high-frequency voltage will be superimposed on the windings. Errors due to this effect are almost entirely associated with badly designed or constructed, high-voltage VTs. As often as not, the apparent error that these discharges cause is due to the behavior of the instrument connected to the IT and is not due to the IT itself, and therefore any calibration made of an IT exhibiting appreciable PDs could be meaningless. An additional effect that PDs can have on an IT is to degrade the insulation slowly with time and eventually cause its complete breakdown.

DEFINITIONS

Transformation Ratio

The transformation ratio of an IT (usually just referred to as “ratio”) is universally defined as the actual primary voltage or current divided by the actual secondary voltage or current. For this definition to have some useful meaning, the voltages or currents must be their root-mean-square (rms) values. This fact is not always made clear in definitions made by some standards organizations, although it will be assumed throughout this article.

In order to express the error in the ratio of an IT, some organizations define this as the quotient of “true ratio” (as defined above), divided by “nominal ratio” (as stated on the rating plate of the IT). Some organizations (particularly in the United States) refer to this quotient as the “ratio correction factor” (RCF). According to the above definition, a perfect IT would have an RCF of unity.

Turns Ratio

The turns ratio of a VT is universally defined as primary-winding turns divided by secondary-winding turns. Turns ratio of a CT is defined by some standards organizations as primary-winding turns divided by secondary-winding turns, but by others as the inverse of this. When using formulae quoted by standards organizations, metrologists should therefore be careful that they check on how the turns ratio of a CT is defined.

Voltage Error of a VT

According to international standards (1), the error in the ratio of a VT is defined as follows:

The error which a transformer introduces into the measurement of a voltage and which arises when the actual transformation ratio is not equal to the rated transformation ratio. The voltage error, expressed in percent, is given by the formula

$$\text{voltage error \%} = \frac{k_n U_s - U_p}{U_p} \times 100 \quad (1)$$

where

k_n is the rated transformation ratio

U_p is the actual primary voltage

U_s is the actual secondary voltage when U_p is applied under the conditions of measurement.

According to the above definition, a perfect VT would have a voltage error of zero. Although not clearly stated in the above definition, U_p and U_s are the magnitude of their values (i.e., they are not vector quantities).

Current Error of a CT

According to the same international standards, the error in the ratio of a CT is defined as in Eq. (1) above except that the word "voltage" is replaced by "current" and the symbol U is replaced by I .

Phase Displacement

In some applications, particularly when an IT extends the range of a wattmeter or watthour meter not working at unity power factor, it is necessary to know the displacement in phase between the primary and secondary voltages or currents. According to international standards, phase displacement is defined for a VT as follows:

"The difference in phase between the primary and secondary voltage vectors, the direction of the vectors being so chosen that the angle is zero for a perfect transformer. The phase displacement is said to be positive when the secondary voltage vector leads the primary voltage vector. *Note.* This definition is strictly correct for sinusoidal voltages only."

Substitute the word "current" for "voltage" in the definition of phase displacement for a CT. There are slight variations in wording between various world organizations in the definition of phase displacement (some organizations call it "phase angle" and others call it "phase error"), but numerically they are all the same.

Most standards organizations give the preference that phase displacement is expressed in centiradians because this unit is logically equivalent to voltage error or current error expressed in percent. Throughout the world, however, many electricity authorities still prefer to express phase displacement in "minutes" because this was the preferred unit used up until about 1970.

From the above, it is clear that IT errors are defined in polar coordinates. In circuit analysis and in technical papers, it is often more convenient to express IT error in rectangular coordinates in the form " $a + jb$ ". The definition of IT error in rectangular coordinates is identical to the definition of voltage error or current error above, except that the voltages or currents concerned must all be complex quantities.

In these definitions, the sign of the quadrature component of error in rectangular coordinates is the same as the sign of the phase displacement in polar coordinates.

TRANSFORMER CLASS

It is the practice among most manufacturers to assign an accuracy class to an IT. In its simplest form, this class designation consists of a number, which indicates the limits of voltage error or current error in percent and the limits of phase displacement in centiradians which should apply during normal use. The range of voltage or current levels, or range of burden levels over which the limits of error should apply, varies from one type of IT to another and also varies from one standards organization to another. Commonly, VTs are required to be within class

between 90% and 110% of rated voltage; CTs between 10% and 125% of rated current; and all ITs between 25% and 100% of rated burden. Burden is usually expressed in "millisiemens" for a VT and in "ohms" for a CT, although before about 1970 it was normal to express burden for both VTs and CTs in volt-amperes (VA). It will therefore be noted that many older ITs have rated burden expressed in "VA" on their rating plate.

In addition to a number, a class designation may also include one or more letters. These letters, for instance "L," mean that the IT is intended for laboratory use; "M" would mean that the IT is for measurement use in the field; while "P" would mean that the IT is for protection use. There are also many subclasses of protection ITs which are designated with additional letters or numbers.

The main difference between measurement ITs and protection ITs is that the latter need to maintain their accuracy at voltages or currents many times greater than their rated voltages or currents. This is particularly the case in protection CTs where their class designation may require them to perform satisfactorily at up to 20 times rated current.

EQUIVALENT CIRCUIT OF AN INSTRUMENT TRANSFORMER

If an IT is analyzed (from basic electrical principles) as two mutually coupled windings each having coupled and uncoupled impedance, the equivalent circuit of Fig. 4 can be derived. For simplicity, this equivalent circuit shows an IT with a nominal ratio of unity.

In Fig. 4, the combined effect of less-than-perfect coupling between primary and secondary windings and the effect of turns-ratio error are shown as a tapping " a " off the secondary winding. Some textbooks like to show the effect of less-than-perfect coupling as an additional reactive impedance in series with one of the windings.

In reality, if no local saturation of the core occurs, it is rare for the effects of less-than-perfect coupling to directly cause errors greater than 0.01% or 0.01 crad at 50 Hz, even when the primary and secondary windings are concentrated in different regions of the core circumference. When a situation is reached in a window-type CT where the ampere-turns become so unbalanced that local saturation of the core occurs, errors are likely to become as high as 10% and 10 crad. At the other extreme, however, when high-permeability toroidal cores are used with uniformly distributed windings, the effect of less-than-perfect coupling would not be expected to cause errors much greater than one part per million at 50 Hz.

The circuit of Fig. 4 does not take into account insulation leakage effects. However, for most practical purposes, particularly for CTs, this can be represented by a single impedance shunting the primary terminals.

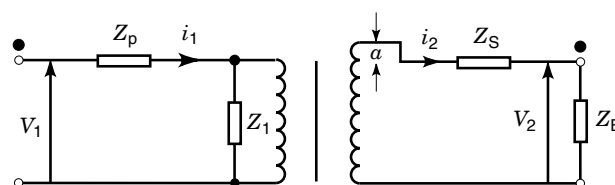


Figure 4. Equivalent circuit of an instrument transformer having a nominal ratio of unity.

TERMINAL MARKING CONVENTION

The universally accepted polarity convention for currents and voltages in an IT is shown in Fig. 4. When the instantaneous directions of the voltages and currents are as shown, the heavy dots represent a similarly marked terminal.

CALIBRATION OF INSTRUMENT TRANSFORMERS

The most common method of calibrating ITs is to compare them in a differential circuit with a standard IT of the same nominal ratio and of known error.

Voltage Transformer Testing Circuits

A basic circuit for the intercomparison of two VTs is that of Fig. 5. In this circuit, “V1” represents an ac instrument which indicates (in terms of V2) the magnitude and phase of the difference in voltage between the secondaries of the standard VT and the VT under test. An instrument designed to compare one VT with another is commonly called a *voltage transformer testing set* (VTTS).

The earthing arrangement should be specially noted in Fig. 5. For safety reasons, one of the secondary terminals of both VTs must be connected to earth. This means that there is no choice but to have the measuring circuit (V1 in Fig. 5), floating up at the secondary voltage of the VTs unless an isolating VT is connected between the “high” ends of the two secondary windings. The alternative of measuring the difference voltage between the “low” ends of the two secondary windings could lead to a large systematic error in the measurement. This is because any insulation leakage currents between the primary and secondary windings of one of the VTs, driven by the full primary voltage, will flow to earth through the measuring circuit. In most test circuits (except where two designated “unearthed” VTs connected between phases of a three-phase system are being compared), it will also be necessary to earth one primary terminal of both VTs.

Current Transformer Testing Circuits

A basic circuit for the intercomparison of two CTs is that of Fig. 6. In this circuit, “A1” represents an ac instrument which indicates (in terms of A2) the magnitude and phase of the difference in current between the secondaries of the standard CT, and the CT under test. An instrument designed to compare

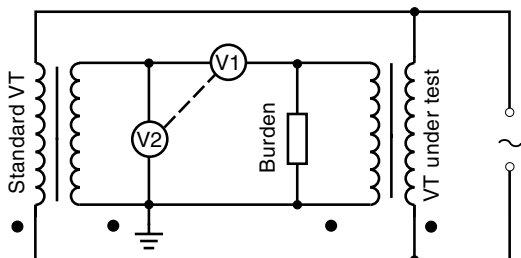


Figure 5. Differential circuit for intercomparing two voltage transformers. The same voltage is applied to both primary windings, while voltmeter V1 simply reads the difference between the two secondary voltages.

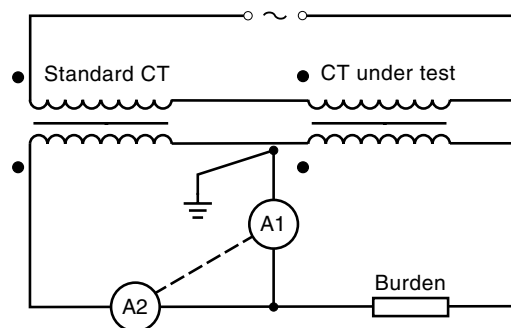


Figure 6. Differential circuit for intercomparing two current transformers. The same current is passed through both primary windings, while the ammeter A1 simply reads the difference between the two secondary currents.

one CT with another is commonly called a *current transformer testing set* (CTTS).

The earthing arrangement of a CT testing circuit is not as critical from the safety point of view as that of a VT testing circuit. In order to minimize measurement errors, however, it is important that one point on the primary circuit and one point on the secondary circuit are earthed. Figure 6 shows the common secondary terminals as earthed. This not only defines the potentials of both CT secondaries, but also eliminates any errors due to the measuring circuit not being at earth potential.

There is no best position for the earth on the primary circuit, however. This is because if one side of the supply is earthed, the primary winding of one or other of the two CTs will be floating above earth, which might cause its errors to change. If the common primary terminal is earthed, any leakage currents between one side of the supply and earth will cause the primary currents of the standard CT and CT under test to be unequal. In order to estimate systematic errors due to earthing arrangements in a CT testing circuit, it is necessary for the metrologist to investigate changes in error with all possible earthing arrangements for each individual CT under test and, from these, assign a component of uncertainty. When the highest accuracy of measurement is required, changes in errors due to different earthing arrangements can be eliminated by employing electrostatically screened standard CTs (4).

Coordinate Systems

There are basically two types of instrument transformer testing sets (ITTSs). The first is where the difference voltage or current is balanced manually against a voltage or current derived from an ac potentiometer. In this method, the out-of-balance voltage or current is displayed on a galvanometer or amplifier-driven cathode ray tube, and, at balance, the error quantities are read on the balancing dials. In the second type, the difference voltage or current is fed into a microprocessor-driven circuit which outputs the error measurement as a digital readout. In practically all commercial ITTSs, the circuitry is such that the difference voltage or current is resolved into rectangular components.

Provided that the error being measured is within about 0.1 % and 0.1 crad, the difference in indication between the measured

quantity in rectangular coordinates and the defined quantities in polar coordinates, is insignificantly small. If larger errors than these are indicated, corrections can be applied to convert from rectangular to polar coordinates according to the following formulae (5).

$$C_R = \sqrt{(1+a)^2 + b^2} - (1+a) \quad (2)$$

$$C_P = \left[\tan^{-1} \frac{b}{(1+a)} \right] - b \quad (3)$$

where

C_R is the correction to the indicated voltage error or current error in per unit

C_P is the correction to the indicated phase displacement in radians

a is the indicated voltage error or current error in per unit

b is the indicated phase displacement in radians

For all practical purposes, these formulae may be approximated to the following:

$$C_R \approx \frac{1}{2}b^2 \quad (4)$$

$$C_P \approx -ab \quad (5)$$

Some manufacturers of ITTSs do not indicate in their literature whether or not these corrections must be applied. With the microprocessor-controlled type of ITTS, it is common for these rectangular-to-polar coordinate corrections to be applied automatically as part of the software.

Calibration Procedure

As a preliminary procedure before calibrating CTs, both the standard CT and CT under test should be demagnetized. This is necessary because during previous use, one of the CTs might have been subjected to switching surges or a component of dc in one of its windings, either of which might polarize the core. This in turn will change the magnetizing current and hence change the errors. Demagnetization can be achieved by applying a voltage to the winding with the highest number of turns with all other windings open circuited. The voltage should be obtained from a variable-voltage source and raised slowly until saturation of the core occurs. Saturation can be detected on an ammeter in series with the winding to which the voltage is applied. Before saturation, little or no current will be seen to flow, but on the inception of saturation a sudden increase of current will occur. The voltage should then be lowered slowly to zero again. It is extremely important that demagnetization is not obtained by applying voltage to the winding with fewest turns, because if it is, then high peak-voltages might be induced in the winding with the highest turns. This in turn could break down insulation or cause electric shock.

The primary voltage or current in an IT testing circuit should never be fed directly from a voltage-regulating device or electronic amplifier. Small components of dc are likely to emanate from either such device which will change the characteristics of the ITs in the test circuit. To block such currents from the test circuit, an isolating transformer should always be used as a buffer.

It is rarely necessary to demagnetize VTs before a test because the very action of applying rated voltage to their primary winding is enough to eliminate any residual polarization that might be present in the core.

When calibrating ITs using a differential circuit, first apply a minimal amount of voltage or current while checking that polarities are correct. If full-load excitation is applied with incorrect polarity on one of the ITs, double full-load voltage or current will be applied to a measuring circuit, and this could cause a lot of damage. If all is well with polarities, then momentarily apply full-load excitation (without taking any readings) to eliminate any remaining polarization of the core.

If the calibration is to be made over a range of voltages or currents, start from the lowest excitation required, and finally, after working upwards to the highest excitation, repeat the reading at the lowest excitation. The difference in readings between the first and last settings will then be an indication of any component of uncertainty due to instability.

When calibrating VTs using a manually balanced VTTS, it may be discovered that when the excitation reaches a certain value, instability is detected on the balance-indicating device. If this instability becomes worse as the excitation is raised further, it is a probable sign that PDs are present either internally to one of the VTs, or externally to the VTs in the form of corona. In either case, if the effects cannot be eliminated, the ITTS's readings may have little useful meaning.

When calibrating CTs, it may be discovered that when a certain excitation is reached, the CTTS's reading slowly drifts with time. If this instability becomes worse as excitation is raised further, it is a probable sign that a small component of dc has entered into the circuit due to bad contacts. Upon detecting such instability, it will be necessary to check the tightness of all contacts and demagnetize all CTs in the circuit again; otherwise the readings may have little useful meaning. If the effect recurs, it will be necessary to search for nickel- or chromium-plated terminals in the circuit and, in extreme cases, remove the plating from the actual contact faces.

AUTOMATIC INSTRUMENT TRANSFORMER TESTING SETS

There is at least one serious shortcoming with all automatically balancing, digital-readout ITTS. That is, any instability that might be present in either the standard IT or IT under test, such as the inception of partial discharges, will not be evident (as it would be on a testing set balanced with the use of a galvanometer or cathode ray tube), and, as a result, a false indication may occur. Users of such ITTSs should therefore make alternative arrangements to detect instabilities, such as connecting a cathode ray oscilloscope across the secondary terminals.

APPLYING CORRECTIONS FOR INSTRUMENT TRANSFORMERS

Standard Transformers

When an IT is calibrated by comparing it with a standard IT, the errors of the standard, in both magnitude and phase, must be added to the indicated errors to obtain (within the uncertainties) the true errors of the IT under test.

Transformers Used in Wattmeter Circuits

When a VT and CT are used to extend the range of a wattmeter or watthour meter, corrections must be made to the indicated value of watts or watthours for the errors of the ITs. These corrections are according to the following formula:

$$\text{Correction} = -(a + b) + (\beta - \alpha) \tan(\cos^{-1} pf) \quad (6)$$

where

a is the voltage error of the VT in per-unit

b is the current error of the CT in per-unit

α is the phase displacement of the VT in radians

β is the phase displacement of the CT in radians

pf is the power factor

Note that all the terms in Eq. (6) are algebraic quantities including power factor. When the power factor is leading, the sign of the power factor (from the point of view of this formula) is positive; and when the power factor is lagging, the sign of the power factor is negative. Note also that this formula is correct for sinusoidal quantities only.

Some organizations use a convention where a positive sign is assigned to the value of a lagging power factor. Therefore metrologists should be careful that the power-factor sign convention is stated before they use any formulae given for wattmeter corrections. In ITTSs, however, the universally adopted convention is for a positive sign to be synonymous with a leading phase displacement and vice versa.

UNCERTAINTIES IN THE ERRORS OF INSTRUMENT TRANSFORMERS

The errors of an IT will change by small amounts with small changes in excitation level, frequency, burden, and temperature. Therefore when calibrating or using an IT, separate investigations should be made to ascertain the coefficients for these parameters so that components of uncertainty can be applied for whatever arbitrary limits are to be assigned to likely changes in these parameters. As a guide, these components of uncertainty can be estimated as follows.

Excitation Level

If the calibration is made over a range of different voltages or currents, a curve of error against excitation level can be plotted; and from the slope of the curve at any excitation, a component of uncertainty can be estimated. In most cases it will be found that the slope of the curve is greatest at the lowest excitation; and therefore from simple inspection of the figures at the two lowest levels of excitation, a component of uncertainty can be estimated. An exception to this rule might occur in an IT which is so constructed that at the highest level of excitation, the core flux density is close to saturation. In such cases it will be found that the errors (particularly the phase displacement) will change very rapidly as full excitation is reached. With such an IT, a separate uncertainty may need to be given for different excitation levels.

When an indicating instrument in the secondary circuit is used to set the voltage or current in the primary circuit, care should be taken (when doing so) to take into account the error

in ratio of the IT. If this is not done, an additional uncertainty will result.

Frequency

Most electricity authorities in Western countries guarantee their frequency to be within $\pm 0.2\%$ of nominal at all times except under fault conditions, so this is a good starting point for estimating a component of uncertainty for frequency. It is not usually convenient or even possible to make deliberate changes in frequency so as to estimate a frequency coefficient of the IT. Theory states, however, that the change in error due to change in frequency in a VT will be the same as for the same percentage change in voltage level. Similarly in a CT, the change in error with change in frequency will be the same as for the same percentage change in total burden. Note that total burden of a CT is an impedance equal to the externally connected burden plus the internal resistance and leakage reactance of the secondary winding. For the purpose of estimating the internal burden of a CT, the dc resistance can usually be taken, because the leakage reactance should be negligible.

Burden

The greatest component of uncertainty during both the use of an IT and during its calibration is usually that due to errors in applying its correct burden. It is usually easy to make a deliberate small change in burden during calibration to ascertain the IT's burden coefficient.

Temperature

It is not usually convenient or possible to change the temperature of an IT during calibration to ascertain its temperature coefficient. However, calculations can be made of the temperature coefficient from measured internal-circuit parameters of the IT and certain assumptions. These assumptions are as follows:

1. Winding resistances changes by $+0.4\%$ per $^{\circ}\text{C}$.
2. Excitation reactance changes by -0.1% per $^{\circ}\text{C}$.
3. Excitation resistance does not change with temperature.
4. Leakage reactances are negligible compared with winding resistances.

Other Components of Uncertainty

Other components of uncertainty that should be considered during the calibration of an IT are as follows:

1. The uncertainty in the standard IT against which the IT under test is calibrated.
2. The uncertainty in the ITTS's indication.
3. Uncertainties due to rounding off of both the reported errors and the reported uncertainty.
4. The effect of different earthing arrangements.
5. The effect of primary conductor configuration in a window type CT.
6. Possible change in errors due to self-heating (usually negligible).
7. Instabilities.

VOLTAGE TRANSFORMERS CONNECTED PHASE-TO-PHASE

A standard VT is normally calibrated with one of its primary terminals at earth potential. However, if such a VT needs to be used with its primary winding connected between phases of a three-phase system, the additional errors that will result can be estimated. These additional errors are based on the assumption (verified from practical experiment) that a VT has a component of error proportional to the voltage (in magnitude and phase) of the midpoint of its primary winding above earth.

A VT intended for use with its primary winding connected between phases of a three-phase system can be calibrated as follows. First calibrate the VT with one of its primary terminals earthed (say the "A" terminal), and then repeat the calibration with the other terminal earthed (say the "B" terminal). The formulae below can then be used to calculate the additional errors (from those measured with the "A" terminal earthed) for the case where the VT is connected between phases of a three-phase system of phase sequence A-B-C.

$$\Delta VE = -\frac{(VE_A - VE_B)}{2} - \frac{(PD_A - PD_B)}{2\sqrt{3}} \quad (7)$$

$$\Delta PD = -\frac{(PD_A - PD_B)}{2} + \frac{(VE_A - VE_B)}{2\sqrt{3}} \quad (8)$$

where

ΔVE is additional voltage error (in per unit)

ΔPD is additional phase displacement (in radians)

VE_A is voltage error (in per unit) with terminal "A" earthed

VE_B is voltage error (in per unit) with terminal "B" earthed

PD_A is phase displacement (in radians) with terminal "A" earthed

PD_B is phase displacement (in radians) with terminal "B" earthed

Corrections for the above additional errors are difficult to apply with confidence, particularly because the phase sequence of the supply must be known with certainty. If the phase sequence is the reverse of A-B-C, the signs of the second terms in Eqs. (7) and (8) must also be reversed. As the additional errors are usually no more than a few parts per million, it is probably safer to apply them as an additional component of uncertainty rather than risk applying the wrong sign to a correction.

CALIBRATION OF THREE-PHASE VOLTAGE TRANSFORMERS

When one phase of a three-phase VT is calibrated against a single-phase standard VT, all three phases must be energized during the calibration; otherwise a systematic error will be introduced. Experience has shown that this is necessary even if the three-phase VT consists of three single-phase VTs mounted in the same tank. Also, the phase sequence of the supply must be noted as part of the calibration, because if the phase sequence is reversed, significant changes in the VT's errors are likely to occur.

ABSOLUTE CALIBRATION OF INSTRUMENT TRANSFORMERS

When it is not appropriate to calibrate an IT by a differential method against a standard IT, such as in a laboratory that is the highest authority in a country, an absolute method must be used.

Most absolute methods of VT calibration involve comparing a VT with a capacitance divider in a bridge circuit, with either (a) the capacitors themselves being calibrated by some other absolute method or (b) the relative values of the capacitors being calibrated by a "buildup" method (6). In the United States and some other countries, the term "bootstrap" tends to be used instead of the term "buildup." A capacitance buildup usually involves measuring a number of three-terminal capacitors individually against one arbitrary reference capacitor and then reconnecting one or more of these capacitors in parallel in each arm of a bridge.

A method for the absolute calibration of VTs or CTs is to construct a special IT with a primary winding made up of sections which can be connected in various series-parallel arrangements. If one of the winding arrangements results in a ratio of unity, which can be calibrated absolutely in a differential circuit by comparing the primary directly with the secondary, it could be assumed that all the other ratios will have the same errors. Such an assumption would be conditional that errors due to insulation leakage currents had been eliminated, as would errors due to imperfect sharing of currents when the winding sections are connected in parallel.

An extension of the above can be used in the absolute calibration of CTs where two CTs of known errors, connected with their primaries in parallel and their secondaries in series, are used to obtain a known, composite CT of higher ratio (7).

An extension of either of the above can be used in the absolute calibration of CTs where a window-type CT is calibrated with multiple inserted primary turns and then assumed to have the same errors with fewer primary turns. For this assumption to be valid, it would be necessary to know that the CT is insensitive to primary-turns configuration.

COMPENSATED INSTRUMENT TRANSFORMERS

Compensation methods may be divided into the categories of (1) simple passive methods, (2) amplifier aiding, and (3) two-staging. Also, any combination of the three categories might be used together.

Simple Passive Compensation

The simplest and most common method of compensation used in measurement-class ITs is what is termed "turns compensation." For example, if the component of voltage error or current error (at rated burden) due to loading is -1% , all that is needed is to make the secondary turns 1% greater than nominal in a VT or 1% fewer than nominal in a CT for the error in ratio to be minimized. Of course, the amount by which the errors change with change of burden will not be improved, while the phase displacement will not be improved at all. Therefore this method of compensation is only efficient for those ITs used in the measurement of voltage or current, or in the measurement of watts where the power factor is close to unity.

Other methods of simple passive compensation (8,9) effectively supply the IT with a negative burden from a separate, adjustable source which cancels out the positive burden supplied to the load.

Amplifier Aiding

There have been large numbers of circuits devised where electronic amplifiers are used to reduce the errors of ITs. Basically, most amplifier aiding methods incorporate a unity-gain buffer amplifier on the output of the IT so that the IT itself is supplying zero burden and the amplifier is supplying the load (10); or the amplifier acts separately to supply the magnetizing current, and hence reduces the voltage drop in the windings (11). Few, if any, of the amplifier-aiding methods compensate for errors due to less-than-perfect coupling between windings, or turns-ratio errors, while most do not compensate for errors due to insulation leakage. A disadvantage of amplifier aiding is that a separate supply must be brought into the IT to power the amplifier.

Two-Staging

The theory behind the two-stage IT was first proposed by Brooks and Holtz as early as 1922 (12). Two-staging involves an IT with two cores, although not all two-core ITs are two-stage ITs. Basically, the primary and secondary windings of a two-stage IT are applied equally to both cores, and an auxiliary winding (often called the *magnetizing winding*) is applied to just one of the cores (Fig. 7). In a two-stage VT, the auxiliary winding has the same number of turns as the primary winding and is connected in parallel with the primary. In a two-stage CT, the auxiliary winding has the same number of turns as the secondary winding and is connected in parallel with the secondary. Toroidal cores are usually used in two-stage ITs. After the auxiliary winding is applied to the first core, the two cores may be stacked together axially before the main windings are applied. Alternatively, the second core may be constructed in such a way that it totally encloses the first core. If the latter core configuration is used, the second core is often referred to as a *magnetic shield*.

Mathematically, it can be shown that the errors of a two-stage IT can be reduced almost to zero at zero external-burden. This will be so no matter what the internal winding impedance. If the external burden connected to an IT is not zero, or at least is not small, the application of two-staging will result in an

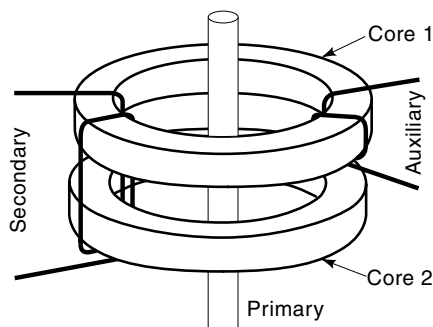


Figure 7. Winding arrangement of a two-stage current transformer.

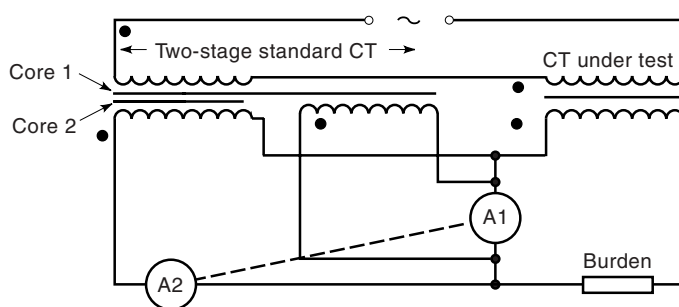


Figure 8. Two-stage current transformer in a differential calibration circuit. The two-core standard is in effect two current transformers, where that on core 2 does the work, and that on core 1 compensates for the errors.

IT with larger errors than would have resulted if the same amount of iron and copper had been used in a single-stage IT. Two-staging does not reduce components of error due to insulation leakage impedance, turns-ratio error, or less-than-perfect coupling between the windings.

From the above it might appear that two-staging would be of very little practical use. However, by various circuit additions, good use can be made of the low-error potential of a two-stage IT. For example, in a two-stage VT, an unloading addition to the circuit can be added as described in Ref. 13.

In a two-stage CT, a unit-gain buffer amplifier can be connected between the auxiliary-winding output and the secondary terminals. With such a system, the component of error due to loading would remain negligible for any external burden up to the limit of output of the amplifier. Also of note is that the amplifier would only need to cope with the small error component of current whereas a single-stage, amplifier-aided CT would require the amplifier to cope with the full secondary current. Some manufacturers make two-stage laboratory-standard CTs amplifier-aided as described above.

Another practical example of the use of a two-stage CT is in a differential calibration circuit (14). If the auxiliary winding is connected directly into the difference current part of the circuit as shown in Fig. 8, the burden applied to the auxiliary winding is practically zero, and the same reduction in error will be obtained as with amplifier aiding.

A special example of a two-stage CT in a differential calibration circuit is the so-called "current comparator" (15). With such a system, ammeter A1 in Fig. 8 is replaced by a current injection circuit derived from A2, and the magnitude and phase of this injected current is adjusted until zero flux is detected in the core to which the auxiliary winding is applied. This zero flux is detected by measuring zero voltage across an open-circuited winding (Fig. 9).

It should be emphasized that the very small error potential of the standard CTs shown in Figs. 8 and 9 will not be realized if the CT is taken out of the differential circuit.

OTHER SPECIAL DESIGNS OF INSTRUMENT TRANSFORMERS

Balancing Windings

In window-type CTs with ampere-turn ratings greater than about 2000, the magnetizing forces of the primary and

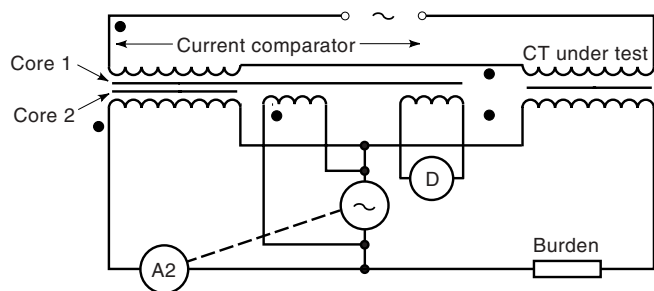


Figure 9. Basic circuit of a current comparator. This is fundamentally the same as the circuit of Figure 8, with the additional feature that the flux in core 1 is forced to be zero.

secondary windings are very large. Although these forces are almost equal and opposite, and almost cancel each other out where both windings are uniformly distributed around the core circumference, any small imbalance of primary conductor configuration will cause large leakage fluxes to flow across air paths (3).

A method of reducing these leakage fluxes, and hence making the CT insensitive to the primary-conductor configuration, is with the use of balancing windings. Such windings consist of a number of equal sections (usually at least eight) all with the same number of turns, each occupying a separate sector of the core circumference, all of them open circuited, and all of them connected in parallel. The effect of these windings is that any imbalance of ampere turns around the core circumference will cause circulating currents to flow between the balancing-winding sections. This in turn will decrease the flux in those regions of the core where the magnetizing forces are greater, but will increase it where the forces are lesser.

For these balancing windings to be effective, there must be little voltage drop in the windings themselves. In order to achieve this aim, practical experiment has shown that the sum of the weight of copper in all the balancing-windings sections should be at least equal to the weight of copper in the secondary winding. Also, the resistance of the paralleling links should be no more than one-tenth of the resistance of any of the balancing-winding sections.

Another way of achieving ampere-turn balance is to enclose the whole of the core and winding assembly in a toroidal, conducting metal envelope made in such a way that it does not form a short-circuited turn. This in effect is an infinite number of single-turn open-circuited windings connected in parallel. Such an arrangement is often referred to as a "magnetic shield," although this is not the same thing as a magnetic shield already described as used in a two-stage IT. Magnetic shields as described above are usually made from copper or aluminum, and again their weight should be at least equal to the weight of copper in the secondary winding.

Transformer Windings with Large Numbers of Turns

When the number of turns on the primary winding of an IT is greater than about 2000, insulation leakage currents between winding layers are likely to contribute a major component of

error in the IT. A method of reducing this component of error could be to increase the insulation between winding layers, but a more efficient method is to divide the winding in question into series-connected sections, with each section confined to a separate sector of the core circumference. By so doing, this component of error can be reduced by a factor of up to n^2 , where n is the number of sections (13,16). In order to gain the full effect of this winding method, however, extra insulation must be applied between the winding in question and any other winding or electrostatic screen.

BIBLIOGRAPHY

1. International Electrotechnical Commission publication, IEC 60044, *Instrument Transformers*.
2. P. J. Betts, Nickel and chromium plated terminals in current transformer circuits, *Proc. IEE*, **132**, Pt. As (4), 178–180, 1985.
3. P. J. Betts, Flux leakage in current transformers, *J. Electr. Electron. Eng. Australia*, **8** (1): 24–29, 1988.
4. P. J. Betts, Electrostatically screened current transformers in measuring circuits, *Aust. Phys.*, 156–159, September 1981.
5. J. R. Fiander, The interpretation and combination of instrument transformer errors, CSIRO Division of Applied Physics Australia Technical Memorandum No. 39, 1986.
6. W. K. Clothier and L. Medina, The absolute calibration of voltage transformers, *Proc. IEE*, **104**, Pt. A, 204–214, 1956.
7. P. J. Betts, W. K. Clothier, and H. A. Smith, Method for the absolute calibration of current transformers, *Proc. IEE*, **129**, Pt. A (5), 322–327, 1982.
8. P. J. Betts et al., Construction details of instrument transformers for dual bridge power comparator, CSIRO Division of Applied Physics Australia Technical Memorandum No. 65, May 1990.
9. A. Hobson, *J. IEE*, **91**, Pt. 2 (20), 147–168, April 1994.
10. G. E. Beard, Single-stage amplifier-aided current transformers possessing small ratio errors at 60 Hz, *IEEE Trans. Instrum. Meas.*, **IM-28**: 141–146, 1979.
11. D. L. H. Gibbings, A circuit for reducing the exciting current of inductive devices, *Proc. IEE*, **108**, Pt. B, 339–343, 1961.
12. H. B. Brooks and F. C. Holtz, The two-stage current transformers, *Trans. Amer. Inst. Electr. Engr.*, **41**: 382–393, 1922.
13. P. J. Betts, Self-calibratable voltage transformer with part-per-million accuracy, *Proc. IEE Sci. Meas. Technol.*, **141** (5): 379–382, 1994.
14. P. J. Betts, Two-stage current transformers in differential calibration circuits, *Proc. IEE*, **130**, Pt. A (6): 324–328, 1983.
15. N. L. Kusters and W. J. M. Moore, The compensated current comparator; a new reference standard for current-transformer calibration in industry, *IEEE Trans. Instrum. Meas.*, **IM-13**: 107–114, 1964.
16. P. J. Betts, A laboratory standard current transformer with a 100:1 step-up ratio, *J. Electr. Electron. Eng. Aust.* **4** (3): 179–186, 1984.
17. A. Wright, *Current Transformers: Their Transient and Steady State Performance*, London: Chapman and Hall, 1968.

PETER J. BETTS

Commonwealth Scientific and
Industrial Research Organisation,
Division of Telecommunications and
Industrial Physics

GENERAL-PURPOSE INSTRUMENTATION AND MEASUREMENT

MULTIMETERS

The term multimeter is used to describe many different types of instruments. Multimeters are just what the name describes, meters that can measure multiple quantities with the same instrument. The most basic multimeters measure voltage, and current both ac and dc, and more advanced digital multimeters (DMMs) measure voltage, current, resistance, frequency, and temperature, and they can trend this information. Top of the line DMMs now have waveform capture capabilities, digital storing capabilities, and printing functions. Some models can communicate with personal computers (PCs). This allows users to store and recall setups, save measurement data to disk, and perform postprocessing functions on the data. This communication feature enables the use of measurement data in documents used for reporting and verification testing.

Technicians, engineers, facilities managers, and repair technicians use multimeters to diagnose a multitude of problems with equipment and power systems. These tools are an invaluable part of any field service personnel's toolbox. Making simple continuity checks is a snap with instruments that use an audible signal to alert the user of resistance values less than 20 Ω .

This article presents a historical account of multimeters, volt-ohm meters (VOMs), and the techniques used to make measurements in electronic and electrical circuits.

HISTORY OF METER MOVEMENTS

Primitive humans were in awe of lightning. Lightning was used as a source of fire for heat and cooking. The early Greeks were the first known to question the phenomenon of lightning. Although for years the Greeks had known about the attraction of particles and the forces of magnetic substances, it wasn't until approximately 700 BC that one Greek astronomer offered an explanation. This astronomer, Thales of Miletus, suggested that magnetic substances had a soul and were alive (1).

Throughout human history in other places there have been scientists who experimented with electricity and magnetism. In 376 BC the Chinese developed the first compass using magnetism. In the late eighteenth century Benjamin Franklin demonstrated that lightning was really an electrical phenomenon. In 1796 Alessandro Volta developed the first battery. In 1820 Hans Oersted discovered the relationship between current and magnetism. Using this information, André Marie Ampère showed the relationship between current flow and magnetism. He actually proved that a current flowing through a wire acted in the same manner as a magnet. In 1827 Georg Simon Ohm discovered one of the most basic and useful principles of electrical engineering, the relationship between voltage, current, and resistance, known as Ohm's law.

In 1881 Jacques d'Arsonval developed and patented a moving coil based on Oersted's principle known as the moving-coil galvanometer (see Fig. 1). The basic principle of operation of

the d'Arsonval meter movement is much like that of a permanent magnet dc motor. As current passes through the windings of the moving coil, north and south poles are set up on the coil, also known as the armature. These poles interact with the poles of the permanent magnet and cause the needle to deflect. The direction of rotation depends on the direction of current flow. Therefore the polarity of the meter movement is crucial, and the terminals on the movement are marked accordingly. The spring that is attached to the moving coil is used to retract the needle to the "zero" position after current is removed (2).

Before we begin the discussion of DMMs, we must first look at their predecessor, the volt-ohm meter, or VOM. The VOM is an analog device that is capable of measuring several different electrical quantities: voltage, current, and resistance. A majority of VOMs incorporate the use of a d'Arsonval meter movement. A simplified diagram of the d'Arsonval meter movement is shown in Fig. 1.

This type of meter responds to the average value of the current that passes through the moving coil. The amount of current passing through the meter movement is dependent on the design and calibration of the instrument. The d'Arsonval meter movement responds to current flow, not the quantity that the scale indicates. In other words, when the meter is being used to measure ohms, the scale may read 100 Ω , but the meter is actually responding to the average amount of current being passed through the moving coil.

This type of meter is rated with a full-scale deflection, usually measured in microamps, that relates the amount of current needed to make the needle move from "zero" to its maximum value. This value is important when designing a meter that incorporates the use of a d'Arsonval meter movement. Resistors placed in series with and in shunt with the meter movement allow the measurement of greater currents and voltages than is indicated by the full-scale deflection of the meter movement.

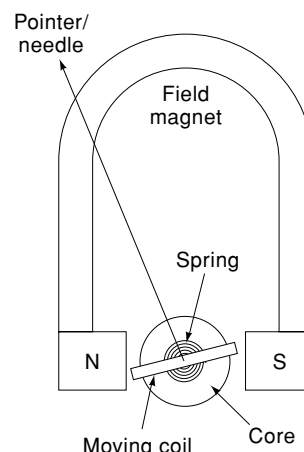


Figure 1. Basic diagram of the d'Arsonval meter movement.

The following sections discuss the use of the d'Arsonval meter movement and its uses in various types of meters. There are three basic measurements performed with VOMs and DMMs: current, voltage, and resistance. Basic design theory for each measurement is addressed.

Direct Current (dc) Meters

Perhaps the most common use for the d'Arsonval meter movement is in the dc ammeter. Figure 2 illustrates the most basic type of dc ammeter. A resistor is placed in shunt with the meter movement in order to limit the amount of current that passes through the meter movement. The value for R_{shunt} will vary depending on the desired full-scale reading of the ammeter.

Given a meter movement with a full-scale deflection of 10 mA and an internal resistance of 100 Ω , the value of R_{shunt} can be calculated as follows for an ammeter with a full-scale reading of 10 A.

The voltage drop across the meter movement at full scale would be 10 mA times 100 Ω , which gives 1.0 V. The voltage across the shunt resistor is the same as the voltage across the meter movement. The amount of current that must pass through the shunt resistor is the total full-scale reading (10 A) minus the current that passes through meter movement (10 mA). This value is 9.99 A. Therefore the value for R_{shunt} would be

$$R_{\text{shunt}} = \frac{V_{\text{shunt}}}{I_{\text{shunt}}} = \frac{1.0 \text{ V}}{9.99 \text{ A}} = 0.1001 \Omega \quad (1)$$

The power requirements of the shunt resistors must be taken into account when designing ammeters. The physical and electrical size of the shunt resistor plays an important role in the overall physical design of the instrument. The power requirement for the shunt resistor in our example was found by taking $I^2 R$ to be approximately 10 W.

The previous example illustrates how the d'Arsonval meter movement was used to make an ammeter capable of measuring current from 0 A to 10 A using a meter movement with a full-scale deflection of 10 mA. It is desirable, however, to have an instrument capable of measuring various ranges of current. Figure 3 is an example of such a meter. The meter in Fig. 3 incorporates the use of the Ayrton shunt.

This type of design allows for multiple scales in one meter. The values for the shunt resistors R_a through R_d are calculated in a similar manner to the shunt for the single-range meter. The actual values for the resistors that comprise the shunt are calculated starting from the most sensitive range to the least sensitive range. In the case of the meter shown in Fig. 3, the instrument is switched to its most sensitive range. The value

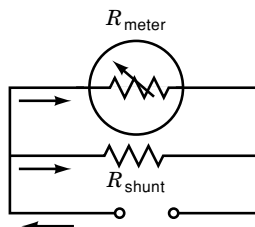


Figure 2. Schematic diagram for basic dc amp meter.

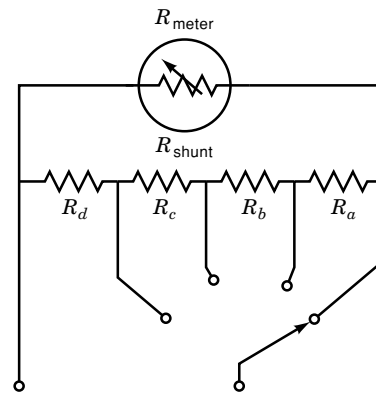


Figure 3. Schematic diagram for a multiple-range dc ammeter using the d'Arsonval meter movement and the Ayrton shunt.

for the shunt resistor for that range is $R_a + R_b + R_c + R_d$. The equations will not be derived for determining the values of the individual resistors. Instead, the reader is referred to Ref. 2.

Direct Current (dc) Voltmeter

As mentioned previously, another use for the d'Arsonval meter movement is for measuring voltage. However, some modification to the meter movement is necessary to make the meter practical. If the meter movement configuration in our previous example were used as is, it would only be able to measure 1 V. By placing resistors in series with the meter movement, as opposed to in parallel as in the ammeter, the deflection of the meter movement can be made to represent voltages of various magnitudes. Figure 4 is a schematic diagram of a simple voltmeter. Proper selection of the series resistors, sometimes called multipliers, R_a , R_b , and R_c results in a three-range voltmeter as shown. Given the internal resistance (100 Ω) and full-scale deflection (10 mA) of the meter movement, the values for the series resistor can be calculated. The three ranges of the meter are 2 V, 20 V, and 200 V.

The current through the meter will be the same as the current that passes through the series resistor. Therefore, for the 2 V range, we calculate the total resistance of the meter and series resistor to be 200 Ω (2 V divided by 10 mA). The value of the series resistor R_c is 100 Ω (200 Ω – 100 Ω). It follows that the values for R_b and R_a are 1900 Ω and 19,900 Ω , respectively.

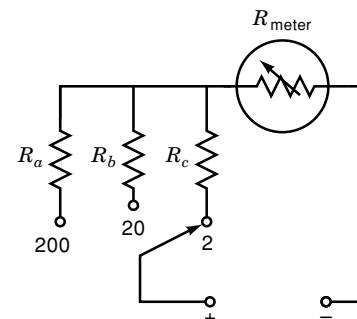


Figure 4. Schematic diagram of a multiple-voltmeter incorporating the use of a d'Arsonval meter movement.

Another commonly used way to calculate the series resistors is to use the sensitivity of the meter movement. The sensitivity of the meter movement is the inverse of the full-scale current rating of the meter movement. To determine the resistor size, simply multiply the desired voltage range by the sensitivity of the meter movement; then subtract the value of the internal resistance of the meter. In our example, the sensitivity of the meter movement is 1 divided by 10 mA. This gives a sensitivity of 100 Ω/V . Therefore the value for R_c , the series resistor for the 2 V range, would be

$$R_c = \left[100 \frac{\Omega}{\text{V}} \cdot 2 \right] - 100 \Omega = 100 \Omega \quad (2)$$

For the 20 V range the value for R_b would be calculated as

$$R_b = \left[100 \frac{\Omega}{\text{V}} \cdot 20 \right] - 100 \Omega = 1900 \Omega \quad (3)$$

Values for other series resistors can be found in the same way. The only limitation on the magnitude of the measured voltage is the dielectric strength of the series resistors used in the multiplication.

When the direct current voltmeter is used to measure the voltage across a component in a given circuit, the voltmeter is in parallel with the circuit and changes the total resistance of the circuit. The resistance in the circuit is now smaller than either of the two resistances, and the voltage measured is less than the actual voltage across the test component. For this reason it is desirable to make the resistance of the voltage meter much larger than the resistance of the circuit under test. In other words, the sensitivity of the voltmeter should be high, on the order of 20 k Ω/V . Quality voltmeters typically have an input impedance of 1 M Ω or higher.

The Ohmmeter

As discussed earlier the d'Arsonval meter movement measures the amount of current that passes through its coil. Using Ohm's law, an instrument that measures resistance can be developed according to the d'Arsonval meter movement principle. Figure 5 gives the basic schematic diagram of an ohmmeter that incorporates the d'Arsonval meter movement. There is one major difference in the diagram of the ohmmeter compared with the voltmeter and ammeter—a voltage source. The ohmmeter requires a voltage source, since it is used to measure a passive element or circuit. Unlike the voltmeter and ammeter, there is no current flow in the circuit unless a source is supplied. A

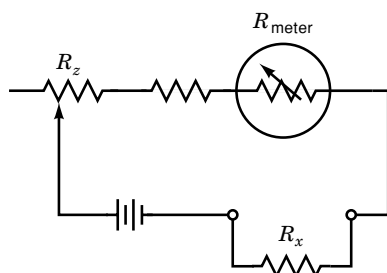


Figure 5. Schematic diagram for an ohmmeter that incorporates the use of a d'Arsonval meter movement.

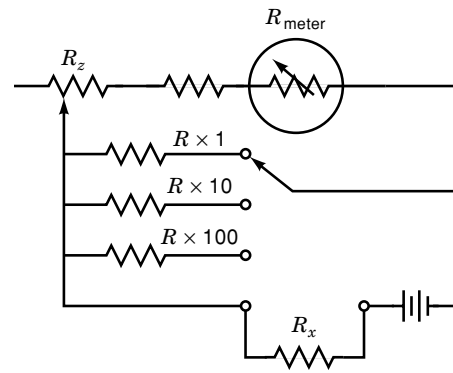


Figure 6. Schematic diagram for a multiple-range ohmmeter that incorporates the d'Arsonval meter movement.

potentiometer or variable resistance is used to “zero” the meter. The potentiometer compensates for fluctuations in the source voltage, lead resistance, and meter resistance changes due to temperature fluctuations.

The principle of operation of the ohmmeter is basically the same as that of the voltmeter or the ammeter. That is, it relates the deflection of the needle to the quantity under test. In this case the deflection of the needle represents the resistance of the element.

Multiple-Range Ohmmeters

Figure 6 is a schematic diagram of a multiple-range ohmmeter. By using some type of switch, usually a rotary switch, different ranges on the ohmmeter can be easily selected.

The Multimeter. The multimeter was developed because scientists, engineers, technicians, and electricians needed to make different types of measurements. The multimeter is simply a meter that is made up of several of the meters previously discussed. Multimeters typically measure dc volts and amperes, ac volts, and dc resistance. Each of these meters uses the same basic d'Arsonval meter movement for its operation, and therefore it is reasonable to assume that one instrument, using the same meter movement, can measure several quantities. Multimeters come in various shapes and sizes with different functions and accuracy. Figure 7 illustrates a differential, laboratory grade multimeter used for measuring both ac and dc quantities.

Up to this point we have discussed the use of the d'Arsonval meter movement with dc voltage and current readings. The d'Arsonval meter movement can also be used for ac measurements. The following sections address the issues of concern in using a d'Arsonval meter movement for ac measurements.

Alternating Current (ac) Meters

Measuring ac quantities can be more difficult than measuring dc quantities. With dc, the value of the voltage or current under test remains constant with time. However, with ac, the value of the voltage or current changes with time. Since the d'Arsonval meter movement responds to the average value of a signal, meter manufacturers are faced with a design dilemma. For dc, the average value of the signal is equal to the rms or effective value. For ac, the average value of the signal is equal



Figure 7. Photograph of the Fluke 803 differential dc-ac voltmeter. Reproduced with permission from Fluke Corporation.

to zero (assuming that only one frequency is present and that the negative and positive half-cycles are equal).

The rms value of a signal is the value that causes the same heating effect in a circuit element as a dc signal. The rms value of a signal, or the effective value of the signal, is found by squaring the signal, then computing the average value or the mean, and then taking the square root of the result (6).

Manufacturers have developed different methods for dealing with ac quantities when using the d'Arsonval meter movement. For example, two such methods are half-wave and full-wave rectification. Figure 8 illustrates a sinusoid, its half-wave rectified signal, and its full-wave rectified signal.

Half-Wave Rectification. When a diode is placed in series with the meter movement, the ac signal seen by the meter

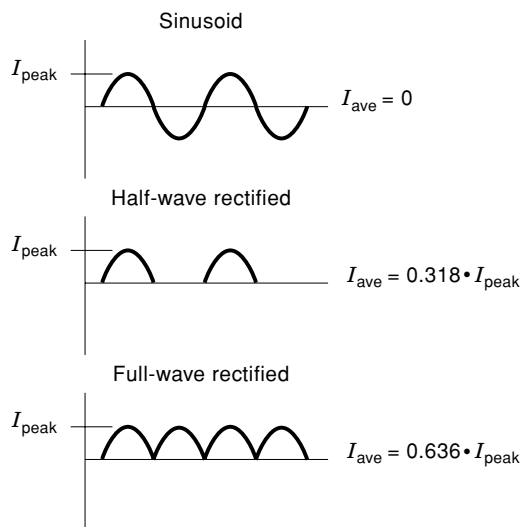


Figure 8. Illustration of the average value for three ac signals.



Figure 9. Example ac voltmeter incorporating full-wave rectification technique. Photo by Bill Dabbs.

movement is half-wave rectified. This produces an ac signal with an average value of 0.318 times the peak value of the signal. By this principle, the scale on the meter itself can be calibrated to read correctly for a purely sinusoidal signal.

If the same meter movement and circuit is used as in Fig. 4, the sensitivity of the meter would be approximately 45% of the dc voltmeter's sensitivity. Therefore a new circuit is developed so that the full-scale deflection of the meter is obtained when a 2 V, 20 V, or 200 V rms ac signal is applied.

Full-Wave Rectification. By using full-wave rectification, the sensitivity of the ac voltmeter can be improved to 90% of the dc voltmeter. The average value of a full-wave rectified ac signal is twice the value (0.636 times) of a half-wave rectified signal.

Both half-wave and full-wave rectification techniques are only valid for pure sinusoidal waveforms. For nonsinusoidal signals, ac voltmeters employing half-wave or full-wave rectification will give erroneous readings. These types of signals will be addressed later in the article.

There are other types of meter movements that are used for both ac and dc measurements:

- Electrodynamometer movements
- Iron-vane meter movements
- Thermocouple meter movements

However, these types of meter movements will not be discussed here. The reader is directed to Ref. 2 for further reading on these meter movements. Figure 9 is a picture of an ac voltmeter that incorporates the full-wave rectification technique.

DIGITAL MULTIMETERS—ELECTRONIC METERS

With the development of the semiconductor in the 1950s came a new breed of instrumentation, the digital multimeter (DMM). Semiconductor technology has advanced greatly since the 1950s and has allowed the development of smaller, more accurate meters. Today's DMMs measure the same quantities as their predecessors, the analog meters. Similar to analog meters, the DMM can be used to measure many quantities



Figure 10. Example of a typical pocket DMM. Photo by Bill Dabbs.

including:

- ac voltage and current
- dc voltage and current
- dc resistance
- Frequency
- Diode characteristics
- Minimum, average, and maximum values
- Capacitance

Manufacturers have developed small pocket multimeters that incorporate basic measurement functions like voltage, current, and resistance. Figure 10 is an example of a small pocket multimeter that measures voltage, both ac and dc, calculates dc resistance, and has an audible indication of continuity for resistance under 20 Ω . Other multifunction DMMs are also available. Any quantity can be measured with a DMM provided that the proper transducer is used. Quantities like temperature, vibration, and pressure are easily measured with a DMM and a transducer. The transducer is used to convert the mechanical quantity into either a voltage or a current. The developed voltage or current is measured by the DMM, and the result is displayed. Using the ratio between the mechanical quantity and the electrical quantity, the actual quantity can be calculated. Several meter manufacturers offer instruments with transducers scaled in such a way that when the DMM displays volts they are directly proportional to the quantity being tested. In other words, 158 mV may be equivalent to 158°F. An example of a popular high-quality DMM is illustrated in Fig. 11.

Theory of Operation

Most digital meters function in a similar way. The signal under test is reduced to an appropriate level for the circuit that will be measuring the signal. This is normally accomplished by using a voltage divider network much like the ones in analog meters. However, once the signal has been transformed, the similarities between analog meters and digital meters end. Digital meters incorporate a digitizer, or analog-to-digital converter (ADC). The function of the ADC is to change the analog



Figure 11. Photograph of the Fluke 87 III true rms multimeter. Reproduced with Permission from Fluke Corporation.

signal into a digital word. Many DMMs incorporate the use of a $3\frac{1}{2}$ digit ADC.

Number of Digits

Most DMMs have a digit or value associated with them. This value is the number of digits that can be displayed. A $3\frac{1}{2}$ digit display can display values from 0 to 1999. Similarly a $4\frac{1}{2}$ digit display can display numbers ranging from 0 to 19,999. This value serves another purpose, apart from indicating the number of digits a meter can display. The number of digits also gives an indication as to the number of “counts” that an instrument uses. A $3\frac{1}{2}$ digit instrument has 1999 counts, and a $4\frac{1}{2}$ digit instrument has 19,999 counts. The count is the highest number that can be displayed on the instrument. Figure 12 is a typical $4\frac{1}{2}$ digit display.

DMM Resolution

The resolution of DMMs is typically given in the specification sheet of the instrument. The resolution of an instrument is the smallest change in a measured value to which the instrument will respond. For an analog instrument, the resolution is one-minor scale division. For digital instruments, the digitizer or ADC determines the resolution. The range that is selected on the instrument determines the resolution in measured units. An example of this calculation follows.

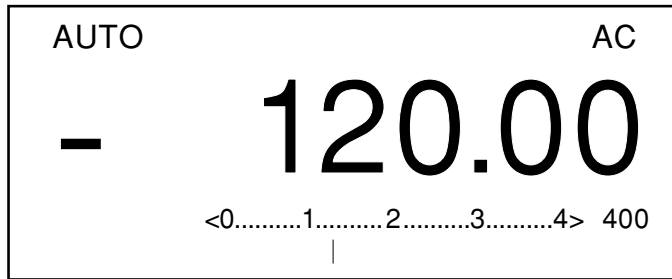


Figure 12. Typical $4\frac{1}{2}$ digit display for a DMM.

This example calculates the resolution of a $3\frac{1}{2}$ digit DMM. The instrument for this example is a $3\frac{1}{2}$ digit, 2000 count DMM on the 200 V range:

$$\text{Resolution} = \frac{\text{Range}}{\text{Count}} = \frac{200 \text{ V}}{2000 \text{ counts}} = 0.1 \text{ V} \quad (4)$$

If the number of counts for this instrument were 20,000, the resolution of the instrument on the 200 V scale would then be 0.01 V.

Sensitivity

Sensitivity of an instrument is the smallest level to which the instrument can respond. This value is usually specified by the least detectable change on the lowest range. For an analog instrument on the 1 V scale with 100 divisions, the sensitivity would be 0.01 V (10 mV). For a digital instrument with $4\frac{1}{2}$ digits on the lowest range of 200 mV, the sensitivity would be 0.01 mV. An example follows.

The instrument is an analog meter with 100 divisions on the scale. The lowest range on the instrument is the 1 V scale. Therefore the sensitivity of the instrument can be calculated as

$$S = \frac{1 \text{ V full scale}}{100 \text{ divisions}} = 0.01 \text{ V} = 10 \text{ mV} \quad (5)$$

To determine the sensitivity of a digital meter, we need to know the number of counts used and the range at which the instrument is set. For our example, suppose we have a 20,000 count DMM that is set to the 200 V range. The sensitivity or maximum resolution on the 200 V range would be 0.01 V. The equation used to determine the sensitivity is

$$S = \frac{\text{DMM range}}{\text{Number of counts}} = \frac{200 \text{ V}}{20,000 \text{ counts}} = 0.01 \text{ V} = 10 \text{ mV} \quad (6)$$

Accuracy and Uncertainty

Accuracy and uncertainty are complementary terms. An accuracy of 99.9% has an uncertainty of 0.1%. Accuracy is a measure of “closeness” to a known or given value. If an instrument has an accuracy of 1%, this means that the measurement will be within plus or minus 1%. Accuracy is given in two ways.

- Percent of full scale
- Percent of reading

There can be quite a difference between these two methods of determining accuracy. The following example illustrates the difference.

Given an instrument with a full-scale reading of 200 V and an accuracy of 1% of the full-scale value. The accuracy is $\pm 1\%$ of the 200 V, or 2 V. A reading of 100 V could actually be 102 V or 98 V. This produces an error of 2%. A reading of 10 V could actually be 12 or 8 V. This produces an error of 20%.

An instrument with an accuracy of $\pm 1\%$ of the reading produces drastically different results than a meter with an accuracy of $\pm 1\%$ of the full-scale value. Measuring the same 100 V would result in a reading between 99 V and 101 V. This is an error of 1%. A reading of 10 V could actually be between 9.9 V and 10.1 V. This also is an error of 1%.

The accuracy depends on temperature and other environmental conditions. Manufacturers often state accuracy in typical values, not the worst case values.

Voltage

Ac, dc, rms. The most important factor to consider when selecting and using a DMM is the method of calculation used in the meter. All of the commonly used meters are calibrated to give an rms value for the measured signal. A number of different methods are used to calculate the rms value. The three most common methods are

- Peak method
- Averaging method
- True rms method

Peak Method. A peak-detecting DMM uses a peak detection circuit to determine the peak of the signal. The circuitry in the meter determines the peak value and divides the result by 1.414 (square root of 2) to obtain the rms value. The rms value of a pure sinusoidal signal is related to the peak value by a factor of 1.414. This is also known as the crest factor of a signal.

Average Responding. An average-responding meter determines the average value of a rectified signal, as does the d’Arsonval meter movement. For a clean sinusoidal signal, the average value is related to the rms value by the constant, $k = 1.11$. This value is used to scale all waveforms measured. As with the d’Arsonval meter movement, the average value of a sinusoid is 0.636 times the peak value. Therefore the rms value of the rectified signal can be found by multiplying the average value $(0.636 \times \text{peak}) \times 1.11$.

True rms. The rms value of a signal is a measure of the heating which will result if the voltage is impressed across a resistive load. One method of detecting the true rms value is to actually use a thermal detector to measure a heating value. Modern digital meters calculate the rms value by squaring the signal on a sample by sample basis, averaging over a period, and then taking the square root of the result. Figure 13 illustrates a good-quality true rms DMM.

In today’s electronic environment, it is important to use instruments that read true rms quantities (Fig. 14). Electronic equipment is nonlinear. That is to say, the current wave shape does not follow the voltage wave shape. As mentioned previously, several techniques used to measure rms quantities



Figure 13. Photograph of the Fluke 87 III true rms multimeter. Reproduced with Permission from Fluke Corporation.

are only valid for purely sinusoidal signals. When electronic equipment is present in a circuit, the current drawn by the load contains harmonic components, currents of frequencies other than the 60 Hz fundamental. Table 1 shows the error produced by three different meter types when measuring nonsinusoidal quantities.

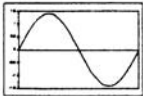
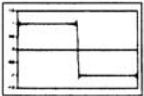
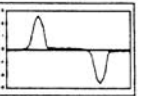
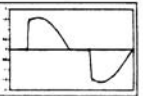
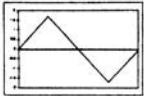
For the true rms method, the instrument displays the correct value for the rms quantity under test. However, the peak detection method and the average responding method produce an error when measuring quantities that are not purely sinusoidal.

Transducers

Many DMMs on the market today can measure current. The maximum current that can be measured directly is usually less than 10 A rms. In most cases it is necessary to interrupt, or break, the circuit under test in order to measure current. DMM manufacturers have recognized this as a problem and now offer current probes or transducers (CTs) for use with their

equipment. These CTs are usually clamp-on type devices that allow you to clamp around a conductor without interrupting the circuit. The CTs also extend the range of the DMM and allow the user to measure current up to 1000 A rms. The CT uses the same principles as a regular transformer except that the turns ratio is designed to give a low-level voltage or current on the secondary winding of the CT that is proportional to the current passing through the primary winding (the jaw of a clamp on CT). As with transformers, CTs are only capable of transforming ac signals. The electric field around a current-carrying conductor changes with time for an ac signal. Therefore the flux and the resulting voltage induced in the secondary of the CT coil change also. For dc, the flux does not change, and no voltage is induced in the secondary of the CT. For dc measurements, a Hall effect current probe must be used. Hall effect CTs work off the magnetic field produced by current flowing through a conductor. These devices often require a separate voltage source to drive the circuits internal to the current probe. Figure 15 illustrate several different models and styles of current probes that can be used with DMMs.

Table 1. Methods for Measuring Voltages and Currents in Multimeters

Meter Type	Circuit	Sine Wave	Square Wave	PC Current	Light Dimmer	Triangle Wave
						
Peak method	Peak/1.414	100%	82%	184%	113%	121%
Average responding	Sine avg. × 1.1	100	110	60	84	96
True rms	rms converter	100	100	100	100	100

Source: Reference 4.



Figure 14. Photograph of the Fluke 8060A true rms multimeter. Reproduced with permission from Fluke Corporation.

Several manufacturers of DMMs incorporate the clamp-on CT into the design of the DMM. This type of DMM is capable of measuring currents into the hundreds of amps as well as voltage, resistance, and other electrical quantities. Figure 16 is an example of this type of instrument.



Figure 15. Example of a clamp on DMM. Photo by Bill Dabbs.

SAFETY

As with any instrument used to measure electricity, some safety concerns must be addressed. This section deals with several safety considerations that must be kept in mind when using multimeters. It is not intended to be the final word on safety, nor is it intended to supersede any manufacturer's safety practices or recommendations. Several standards are cited, and the reader is encouraged to review these standards prior to purchasing and using any instrument.

When working with electrical circuits it is ultimately your responsibility to ensure your safety. Working on dead or de-energized circuits whenever possible is recommended. However, certain situations do not allow measurements to be performed on dead circuits. If working on live or energized circuits, remember the following:

Never work alone. Always work in pairs in case you accidentally come in contact with the circuit you are measuring. Your partner can disconnect the circuit or knock you away from the live circuit.

Always use protective gear. You should always wear eye protection and insulated gloves with leather protectors. This equipment should be tested annually to ensure that its electrical withstand capability has not degraded. Never wear jewelry, watches, or other metallic articles of dress when working on electrical circuits. Any electrical conductor coming in contact with a live circuit can have disastrous results. Flame resistant clothing is also a good idea.

When performing measurements on live circuits, follow these basic steps to ensure your safety:



Figure 16. Example of common current probes available for use with DMMs. Photo by Bill Dabbs.

Table 2. Effects of Current on the Human Body

<i>Current Applied to Skin</i>	<i>Effect of Prolonged Contact</i>
6 A or more	Sustained myocardial contraction followed by abnormal rhythm. Temporary respiratory paralysis. Burns on contact area.
0.1 to 2–3 A	Ventricular fibrillation. Respiratory center intact.
30 to 50 mA	Pain, fainting, exhaustion. On prolonged duration, death often results.
16 mA	“Let Go” current, muscle contraction.
1 mA	Threshold of perception, tingling.
<i>Current Applied to Myocardium</i>	<i>Effect of Prolonged Contact</i>
100 μ A	Ventricular fibrillation.
10 μ A	Recommended maximum leakage current for patient connected biomedical equipment.

1. Always connect the ground lead first then the hot lead. Removal of the test leads is in reverse order. Remove the hot lead first then the ground lead.
2. Use the three-point test method. When performing measurements, it is important to ensure that the test equipment is functioning properly. First, perform a similar measurement on a known source prior to connecting the instrument to the test circuit. Second, perform the measurements on the circuit under test. Third, test the meter on the known circuit to verify its operation after the measurement.
3. Use one hand. Whenever possible, only use one hand when connecting and disconnecting test leads from a live circuit. This lessens the chance that you will come in contact with a live circuit. More important, it lessens the chance that you will receive an electrical shock across your chest and through your heart.

When the human body comes in contact with live electrical circuits, current flows through the body. If this current flow is across the chest and heart, the result can be disastrous. Table 2 lists different levels of current and their effect on the human body. Using the information in Table 2, we can calculate a threshold for a hazardous voltage. Under dry conditions the approximate skin resistance across a human body is 1000 Ω . Using Ohm's law, a voltage of 35 V would produce a current of 35 mA. From Table 2 it can be seen that this would cause loss of breathing and possible death. When the skin is wet, the hazardous voltage can be as low as 10 V.

Voltage Rating

All multimeters are tested against and are rated in accordance with safety standards. The International Electrotechnical

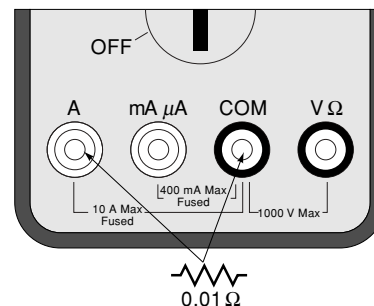
Commission (IEC) has several safety standards by which instruments are tested. One standard, IEC 1010, is a test procedure that uses three criteria to determine a multimeter's voltage rating. They are the steady-state voltage, peak voltage, and source impedance. These three criteria are used to describe the meters true withstand capability. Table 3 lists the different levels of overvoltage categories and the values used for each criterion.

Another concern with safety is the test leads. If an instrument is rated CAT III 600 V, but the leads are only CAT II 600 V, the leads become the actual rating of the measurement system. In other words, don't overlook the test leads. Choose test leads that are certified in the same category and voltage rating as the meter or higher. Test leads are also rated by IEC 1010.

Another safety concern is that the meter is connected properly. Many DMMs can measure both voltage and current. When a voltage measurement is made, there is a high input impedance (usually 10 M Ω) on the voltage circuit. This impedance conditions the voltage or reduces the voltage to appropriate levels for the circuitry within the instrument. However, when measuring current, the instrument is placed in series with the load under test. The burden, or input impedance, is typically 0.01 Ω or less. This low impedance doesn't affect the current that the instrument is trying to measure. The voltage developed across the burden is read, and using ohm's law, the measured current is displayed. In order to avoid confusion, meter manufacturers typically use a common jack (COM), a volts jack (V Ω), and an amps jack (A); see Fig. 17. When performing voltage measurements, the leads are connected to “COM” and “V Ω .” When performing current measurements, the leads are connected to “COM” and “A.” However, if the leads are

Table 3. Transient Test Values for Overvoltage Withstand Capabilities

Overvoltage Category	Working Voltage (volts)	Peak Impulse Transient (volts)	Test Source Impedance (ohms)
CAT I	600	2500	30
CAT I	1000	4000	30
CAT II	600	4000	12
CAT II	1000	6000	12
CAT III	600	6000	2
CAT III	1000	8000	2

**Figure 17.** Example DMM connection jacks.

left connected to “COM” and “A” and a voltage measurement is performed, the low impedance of the burden is effectively a short circuit. This short circuit occurs even if the selector switch or dial of the meter is turned to volts. For this reason fuses must protect the amp terminals, and one must make sure to use the fuses specified by the manufacturer. These fuses are generally high energy and have an interrupting capacity to protect users (4).

SUMMARY

Multimeters are used worldwide by many people and in different professions. Since their invention, they have proved to be invaluable tools for technicians, electricians, engineers, and homeowners. Multimeters can be used to troubleshoot equipment, verify proper operation of equipment, and determine values of electrical components. Advances made in multimeter technology have made them more affordable and economical for practically any individual to own.

When used properly, multimeters serve the user well. Safety is of the utmost importance, and meter manufacturers take great pains to ensure that they produce safe products for the consumer. International committees have developed standards that rate multimeters. Consumers should use these ratings to ensure that the multimeter they are using meets or exceeds the requirements of the work being performed.

ACKNOWLEDGMENTS

The author would like to thank Daniel Brooks for supplying several of the instruments used for the photographs in this article and also Bill Dabbs for his time and expertise in taking the photographs.

BIBLIOGRAPHY

1. T. Henry, *Electrical History*, Tom Henry, 1996.
2. L. D. Jones and A. F. Chin, *Electronic Instruments and Measurements*, New York: Wiley, 1983.
3. C. J. Melhorn and M. F. McGranaghan, Interpretation and analysis of power quality measurements, *IEEE Trans. Ind. Appl.*, **31**: 1363–1370, 1995.
4. *ABCs of Multimeter Safety*, B0317UEN, Fluke Corporation, 1996.
5. *IEEE Recommended Practice for Power and Grounding Sensitive Electronic Equipment*, New York: Inst. Electr. Electron. Eng., 1992.
6. J. D. Irwin, *Basic Engineering Circuit Analysis*, 2nd ed., New York: Macmillan, 1987.

CHRISTOPHER J. MELHORN
Electrotek Concepts, Inc.

DIGITAL MULTIMETERS

Analog multimeters were created in the early part of the twentieth century. The most common form measured dc and ac voltage, resistance, and dc current, and these became known as volt-ohm-milliammeters or VOMs.

In 1955, the first digital voltmeter was marketed by Non Linear Systems, Inc. This meter scaled inputs and then digitized them, displaying the results in a numeric digital display.



Figure 1. Modern full feature DMM (Fluke 87 III).

Digital measurement and display technologies developed in parallel with digital integrated circuit technology until finally, in the late 1960s, portable digital versions of the VOM appeared. This form of meter shown in Fig. 1 has since become known as the digital multimeter (DMM).

THE ELEMENTS OF A DMM

A DMM is made up of several basic functional elements shown in Fig. 2.

1. Signal conditioning
2. Analog/digital conversion (A/D)
3. Numeric digital display

Signal Conditioning

The analog-to-digital converter (ADC) in most DMMs converts a dc voltage input to a digital form for display. Ac or dc voltage, current, and resistance are the most common parameters to be measured. Now many DMMs also measure frequency, and additional functions, such as temperature, have also appeared.

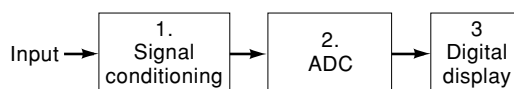


Figure 2. Basic DMM block diagram shows the basic flow of information from the analog input signal through the various analog signal conversion circuits which convert the measured quantity to a dc voltage equivalent. Then the ADC translates this dc signal to digital form, and the display system shows the resultant value with appropriate annunciation, such as measurement units of the original input signals. Not shown are the necessary, but not directly related, parts of the DMM. These include the power supply and mechanical packaging.

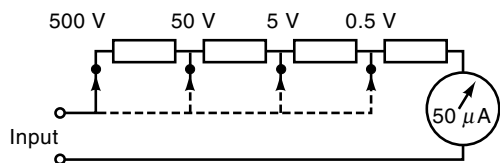


Figure 3. VOM input multiplier—dc volts. The basic input configuration for dc voltage input to an analog VOM. The major point of this diagram is that the input resistance varies as a function of range to maintain current through the meter movement within its limited dynamic range. This configuration leads to the common VOM sensitivity specification in ohms per volt. A 20,000 Ω/V meter draws 50 μA from the input at full scale indication on any dc voltage range.

First these inputs must be switched and conditioned to present an equivalent dc voltage of the appropriate range to the ADC or a sealed input to the counting circuits.

Dc Voltage

It is necessary to scale a wide range of dc voltage inputs to the limited range of the ADC. A resistive divider and switching are generally used for this function. One of the major differences between the classical analog VOM and the DMM is in the way this scaling occurs.

In the analog VOM, the input circuit is designed to provide full-scale current to operate the d'Arsonval meter movement (refer to Fig. 3). A typical meter has 50 μA full-scale sensitivity. The input sensitivity of such a meter is specified by calculating the series resistance per volt at 50 μA , $R = 1/0.00005$ or 20,000 Ω/V . A 5 V range therefore has an input resistance of 100,000 Ω , and a 500 V range has an input resistance of 10 M Ω as more resistors are added to the input multiplier.

The DMM, on the other hand, operates with a fixed resistive divider across the input (typically 10 M Ω) regardless of the range chosen (see Fig. 4). In this case the input resistance is a divider with output taps switched to present a voltage, usually less than 1 V, to the ADC. For this reason, the DMM uses much less current (is more sensitive) on the lower voltage ranges than the typical VOM, thereby reducing circuit loading errors.

Another difference between the VOM and the DMM is in the rejection of ac noise which is present along with the dc voltage to be measured. In the VOM, the mechanical damping provided by the d'Arsonval meter movement averages this noise and presents a stable reading even when the noise is

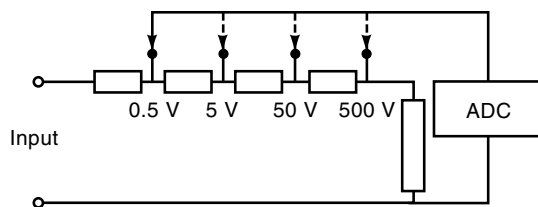


Figure 4. DMM input divider—dc volts. The constant input resistance input divider network in the typical DMM. The total resistance at the input terminals is relatively high, usually 10 M Ω . Output taps are selected to present a small voltage (less than 1 V) to the ADC even when the applied input is 500 V or more. The same divider is often used for AC voltage scaling. In this case, shunt capacitance should be specified.

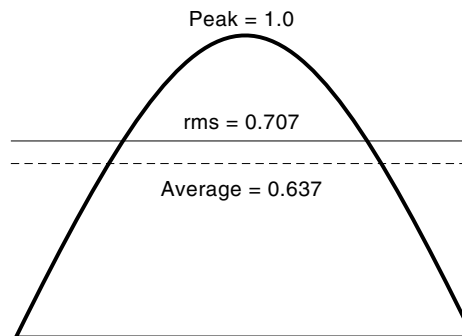


Figure 5. The positive half-cycle (180°) of a sine wave and the relationship between peak, rms, and average values. The simplest and least expensive DMMs have always used the rectified average value scaled to equivalent rms for display. This relationship is true only for the sine wave whereas most 'real-world' measurements are of waveforms that range from square to narrow pulse with varying rise and fall times. Therefore, this method is usually inaccurate to some degree, and it is difficult to predict the amount or direction of the error. A similar problem occurs when the peak value is scaled to the rms. The end message: If the rms value is important, measure the true rms.

a significant portion of the signal being measured. The DMM input circuit, combined with the ADC method used must filter to provide equal or better performance. See a discussion of noise rejection in the section on A/D conversion.

Ac Voltage

For ac voltage, the signal must be scaled and then converted to a dc equivalent value before sending it to the ADC. Following are two popular methods of conversion (refer to Fig. 5).

Average Sensing-Rms Indicating ac Measurements. In the analog VOM, ac to dc conversion is accomplished by using a bridge rectifier in the meter circuit to provide a single polarity current that deflects the meter needle. This current has significant ripple at twice the frequency of the measured ac signal, but this is effectively filtered (or averaged) by the mechanical damping of the meter movement. In the DMM, an operational rectifier circuit is used, but the output must be filtered to reduce the ripple before it is applied to the ADC.

This rectified average value is equal to $2/\pi$, or 0.637 times the peak value for a sine wave. The desired ac voltage (or current) value is the rms, or heating value, which is equal to $1/\sqrt{2}$, or 0.707 times the peak of a sine wave. Therefore average analog or digital sensing meters, scale the rectified signal by $\pi/(2 \times \sqrt{2})$, or 1.111 times the rectified average value. This ratio of rms to average value is known as the form factor of the input waveform. Although this method works for a sine wave of a single frequency, significant errors occur when the waveform is distorted because these ratios are no longer true if harmonics are present.

True Rms ac Measurements. A more accurate conversion provides the true rms value—equivalent to a dc voltage which would heat a resistor to the same temperature. With true rms conversion, this is true regardless of waveform. Most DMMs, which offer true rms capability, use an analog computing circuit to perform the necessary calculation (Fig. 6).



Figure 6. An early 1980s true rms DMM (Fluke 8060A).

Rms stands for root-mean-square, which is expressed mathematically as $\text{rms} = \sqrt{(\sum i^2/n)}$, where i is the instantaneous value and n is the number of points i (1 to n), which are squared and summed for use in the calculation (see Fig. 7).

Some DMMs also allow measuring the peak value of the ac waveform.

Dc and ac Current. To measure current with a DMM, it is necessary to convert the current at the input to a voltage for use by the ADC. This is done through a series of switched resistors, called shunts. A small-valued resistor in the input is placed in series with the current to be measured and Ohm's law ($E = IR$) defines the small voltage proportional to the input current which is then measured.

Ac current uses the same method except that the voltage across the shunt is routed through an ac-dc voltage converter before going to the ADC.

Burden Voltage. The voltage drop across the current input resistance of a DMM connected in a series circuit is known

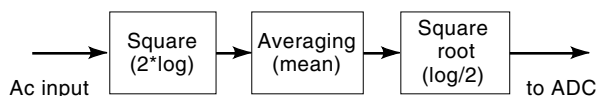


Figure 7. True rms ac-to-dc conversion. The functional blocks of the basic analog circuit that produce a dc voltage proportional to the rms value of a complex waveform. The logarithmic conversion for squaring is accomplished by using the current versus voltage relationship of a forward-biased base-emitter junction of a transistor. Factors which affect the performance of this conversion process include frequency response, crest factor, and slew rate. Commercial converters are available. Analog Devices (ADI) in Norwood, MA, has several products. Their rms converter website is <http://products.analog.com/products.html/list-gen.117.html>

as burden voltage and is subtracted from the voltage normally available to the load. Therefore the resistance added to the measured circuit should be kept as small as possible to minimize measurement error.

Resistance

Because the ADC measures only dc volts, it is necessary to create a voltage proportional to the resistance when that measurement is desired. The DMM input circuit must provide a dc current flowing through the resistor and then measure the resulting voltage. Ohm's Law provides the necessary relationship ($R = E/I$). Of course, it is necessary to know the applied current value for this method to be useful. Many DMMs provide a calibrated constant current source for this purpose. Many others use reference resistors in series with the resistor being measured, and then measure the unknown and reference voltage drops. When combined with the known reference resistor values, this voltage ratio technique also yields the desired resistance value.

ANALOG-TO-DIGITAL CONVERSION

There have been several schemes used for the ADC conversion in DMMs. Some of them read voltage by using successive approximation, a technique which measures the voltage directly with a digital-to-analog converter (DAC) and voltage comparator. Many others rely on time-counting techniques. The most common of these by far is dual-slope integration.

Dual-Slope Integration

This process uses an operational amplifier configured as an integrator [Fig. 8(a)]. First the integrator is set to zero. Then the input voltage (V_{input}) is applied for a precise period of time controlled by a counter. The input is then switched to a reference voltage of the opposite polarity ($V_{\text{reference}}$), and the time required to integrate back to zero is counted. When done correctly, this process relies only on the stability of the integrator capacitor, the precision and stability of the counter, and the accuracy of the voltage reference to achieve the desired result.

The process just described is represented by the voltage versus time diagram in Fig. 8(b). V_1 represents the integrator voltage output at the end of the fixed period when an input voltage is applied. T_1 represents the time required for the integrator to return to zero when the appropriate reference voltage is selected. A similar case is shown for a larger input voltage (V_2) where a longer time (T_2) is required to return to zero.

In the dual-slope ADC, the time required to deintegrate (return to zero) is directly proportional to the integrator voltage at T_0 , which, in turn, is directly proportional to the input voltage.

NOISE REJECTION

Normal-Mode Noise

Common implementations of the dual-slope ADC use the unique noise rejection characteristics of the process to minimize the effects of the major noise source, the 50 Hz or 60 Hz power system. The inherent filtering characteristic of this

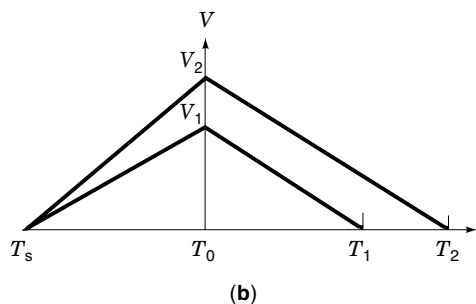
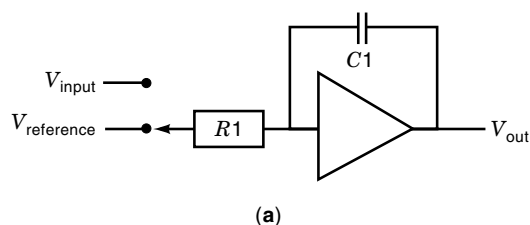


Figure 8. (a) The basic elements of the analog integrator circuit of the dual-slope integrator. The circuit uses an inverting operational amplifier to convert an input current (through input resistor $R1$) to a voltage across the integrator capacitor $C1$. From an initial condition when the voltage across $C1$ is zero, $R1$ is connected to V_{Input} , and integration occurs for a fixed period of time. This is the first slope of the dual-slope process. The polarity of the charge on $C1$ is determined, and a reference voltage, $V_{\text{Reference}}$ is selected. The purpose of the second slope integration is to determine the time needed to return the voltage on $C1$ to zero. This time is proportional to the voltage on $C1$ at the end of the first slope which, in turn, is proportional to the applied input voltage. (b) The voltage versus time diagram.

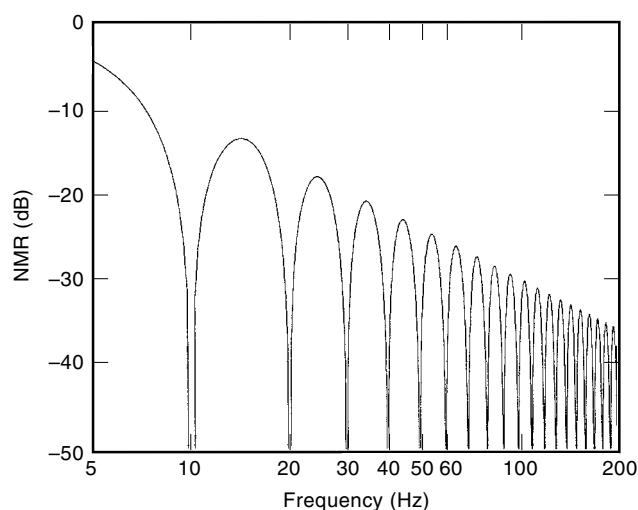


Figure 9. Dual-slope integrator normal-mode rejection (NMR)—100 ms integration time. The rejection characteristic of input frequencies whose periods are directly related to the inverse of the signal integration time. The 100 ms time used for this example equals a period of one cycle at 10 Hz. Therefore there is near infinite rejection of 10 Hz and its integral multiples.

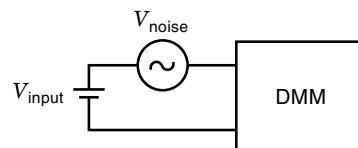


Figure 10. Normal (series)-mode noise. The interfering signal (noise) is a part of the measured signal seen by the meter. It may be ripple in the output of a power supply or may be induced in the signal leads connecting the input source and the DMM.

converter, set to integrate for 100 ms, is shown in Fig. 9. This integration time results in near infinite rejection of noise at integral multiples of 10 Hz and because exactly 5 cycles of 50 Hz and 6 cycles of 60 Hz occur in 100 ms, this technique addresses power system noise directly.

The presence of noise directly affects the performance of a DMM by causing a reading error or instability. The noise may be an integral part of the signal being measured but is outside the passband of the measuring system. In the preceding case, the circuit measures dc voltage, and the noise is at the power frequency. This is referred to as normal-mode or series-mode noise (Fig. 10).

In Fig. 9, normal-mode noise rejection is expressed as a ratio of peak display disturbance divided by the in-band signal being measured. For example, if the DMM is reading 5.000 V with noise up to ± 0.050 V, then $0.05/5.0$ yields a normal-mode rejection ratio (NMRR) of 0.01—expressed in decibels as 40 dB NMRR.

Common-Mode Noise

If the noise is present between a third point and the two measurement connections in common, then it is known as common-mode noise (Fig. 11).

The common-mode rejection ratio (CMRR) is calculated similarly to the ratio for normal-mode noise, the peak error or instability displayed in the measurement reading divided by the nominal reading value.

These examples have shown dc voltages affected by ac noise. However, the same principle also applies to ac measurements. For example, when measuring the ac voltage in a three-phase power system, neither input terminal is at ground, and a large common-mode signal with respect to ground is present. Now this frequency is the same as the frequency of the voltage being measured. The result may be a measurement that is stable, but either higher or lower than the actual phase-to-phase voltage. Reversing the test leads shows this possible “turnover”

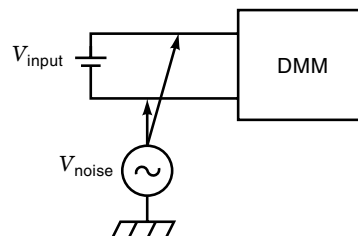


Figure 11. Common-mode noise. The interfering signal (noise) influences both the input and common leads of the DMM input. This type of noise results when the circuit being measured is floating with respect to local “ground.”

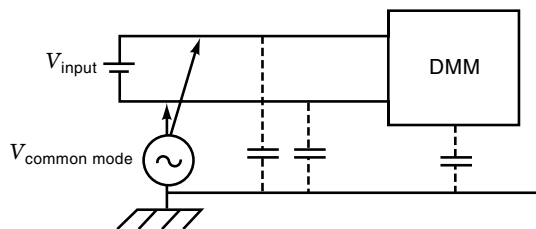


Figure 12. Common-mode capacitive leakage effects. The common lead of the meter is usually more closely associated with shields and power supply reference in the meter and therefore has the lowest impedance for possible ground loops. The voltage input is the high (typically 10 M Ω) input. The net effect is for the common-mode signal to appear as an equivalent normal-mode noise input in this unbalanced circuit.

error. The shielding and filtering characteristics of the DMM become important in minimizing the effect of leakage capacitance (Fig. 12).

Dc voltage measurements also show this effect with a dc common-mode voltage. In this case, leakage resistance is the issue.

DIGITAL DISPLAY

The final element of the DMM is the digital display. Several display devices are in use now including vacuum fluorescent, light-emitting diode (LED), and liquid crystal display (LCD) technologies. The LCD is the most common display in use today, primarily because it requires the least power to operate. Typical digital displays show the measured value using seven segments per digit plus decimal points and polarity indicators. More capable DMMs also provide bar graph and measurement unit annunciation elements. Typical displays are shown in Fig. 13.

DISPLAY RANGE, RESOLUTION, AND ACCURACY

Range

The effective range of a DMM display is described in several different ways:

1. In counts (2000 counts, 3000 counts, etc.). The count reference comes from the fact that most DMMs use a COUNTER circuit to drive the display circuitry.
2. As a number of digits (3 1/2, 4 3/4, etc.). Referring to a display as “3 1/2 digits” or “4 3/4 digits” indicates that the leading, or most significant, digit in the display never exceeds the numeral 2, 3, or 4. A 3 1/2 digit meter might

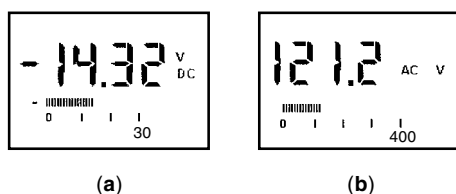


Figure 13. (a) Dc volts on 30 V range. (b) Ac volts on 400 V range.

operate to 2.999, a 4 3/4 digit might operate to 3.999, etc. Attempts have been made to make the reference scientific by calculating the base 10 logarithm of the maximum number. For example, 1999 has 3.3 digits, 2999 has 3.5 digits, and 39999 has 4.6 digits.

3. By showing the actual measurement limits in volts, etc. (–2.999 to +2.999). This is the simplest and least ambiguous way to describe the effective measurement range of a DMM.

Resolution

The resolution of measurement may be specified in least significant digits (LSD), counts, percentage of range, or absolute units (0.01 V), or it may be implied in the range specification. When in the 2.999 V range, for example, the implied resolution is 0.001 volts.

From a practical standpoint, usable resolution is sometimes less than display resolution. Reading instabilities may dictate that the usable resolution of a meter is only ± 0.002 V or 0.003 V even though it has a display resolution of 0.001 V.

Accuracy

An accuracy statement for a DMM includes many factors. Among these are resolution, linearity, long-term stability, ambient temperature, and humidity.

Single-Term Accuracy. The simplest DMM accuracy specification uses one term, such as 3% of range, for each function. This applies over all temperature and humidity conditions for some period of time. One year is a common period. This is the usual way to specify the accuracy of analog VOMs whose ranges may be organized in 1, 2, 5 or 1, 3, 10 sequence. For example, $\pm 3\%$ of full scale on the 10 V range (± 0.3 V) means that the accuracy of a 10 V reading is 3% of the reading whereas that of a 3 V reading on the same range is 10% of the reading. The user can switch to a 3 V range in this case and return to 3% of reading accuracy at full-scale.

In the DMM, however, the digital display ranges are in decades, such as a 3, 30, 300 sequence. Ranging occurs near one-tenth of range. For the following descriptions it will be helpful to refer to Fig. 14.

Single-Term DMM Accuracy. A reading on the 30.00 range displays 3.00 (10 mV resolution) before ranging allows a change to 3.000 where 1 mV resolution is possible. A specification of 1% of full scale (± 0.3 V) becomes 10% of the reading at 3 V before the range is changed. This leads to an accuracy statement which combines a percentage of the reading term and a resolution-based (percentage of range) term.

Two-Term Accuracy. The input divider, amplifier gain, and reference voltage accuracies are combined in a percentage of reading term because they all contribute to the actual value displayed. The amplifier offset, noise, and least significant digit considerations for the ADC and display are combined in a percentage of range term stated as counts, digits, or percentage of range. In the 30 V range example before, such a specification might read as follows: $\pm(1\% \text{ of reading} + 2 \text{ counts})$.

At 30 V, the resulting maximum error is 1% of 30.00 (0.3) plus 2 counts (2×0.01 or 0.02) for a total of 30.00 ± 0.32 V.

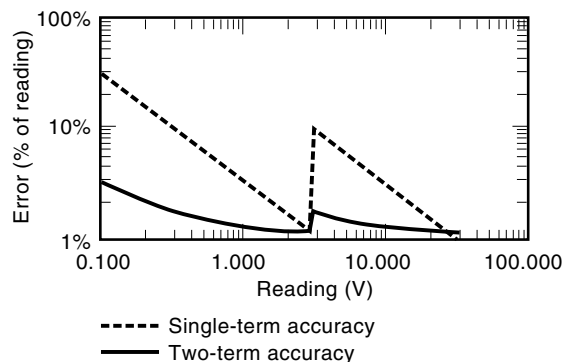


Figure 14. Single-term (dashed line) versus two-term (solid line) accuracy. The graph assumes that range change occurs exactly at the 3 V point. In fact, when an instrument is autoranged, the decision to down range comes at about 90% of the next lowest range. This is at 2.7 V in this example. The reason for the overlap is to minimize “hunting” of the ranging circuit because of slight instability or noise near the ranging point.

Taken as a percentage of the reading, this equals about 1.07% of the reading. At 3 V in the same range, the error is 1% of 3.00 (0.03) plus the same two counts (0.02) for a total of 0.05 V. As a percentage of reading this is now only 1.6% of the 3 V input, far better than the single-term (% of full-scale) accuracy described previously.

Other Factors To Be Considered

Resolution and linearity are included in the basic accuracy specification for most DMMs. Most specifications indicate the time (6 months, 1 year, etc.) for which the specification applies. This takes care of the long-term stability consideration. An accuracy specification may also indicate a temperature range and humidity level which are included in the basic statement. Additional specifications may be furnished to indicate added error terms for operating temperatures and humidities outside the basic range stated.

APPLICABLE STANDARDS AND RATINGS

International Safety Standards for DMMs

IEC 61010 establishes international safety requirements for electrical equipment for measurement, control and laboratory use. There are four categories, CAT I, CAT II, CAT III, and CAT IV ranging from electronic to outdoor electrical power distribution, respectively. There are also maximum voltage ratings that indicate the maximum service voltage to be applied to a meter which is so rated. A meter intended for use at interior single-phase receptacles might carry a notation of CAT II, 300 V, for example. Independent testing organization, such as UL, CSA, TUV, certify that a DMM meets the appropriate IEC standards. Overload ratings are also useful. These ratings indicate maximum values of voltage or current that may be applied to a DMM input without causing damage.

Standards for DMM Measurements and Accuracy Statements

Work began in 1961 to develop a standard that covers digital measuring instruments. At the end of that decade, American

National Standard Requirements for Automatic Digital Voltmeters and Ratio Meters, ANSI C39.6-1969, was approved. It was published by ANSI in 1970. The last revision of this standard was approved in 1982 and published in 1983 as the American National Standard for Electrical Instrumentation-Digital Measuring Instruments, ANSI C39.6-1983. Although this standard has been withdrawn and is no longer in effect, many of the definitions of terms, testing methods, and specification presentations for DMMs today are based on the examples set in this standard.

BIBLIOGRAPHY

1. *Standard Handbook for Electrical Engineers*, 13th Edition, New York: McGraw-Hill, 1993.
2. *Reference Data for Radio Engineers*, 5th Edition, Indianapolis, IN: Sams, 1974.
3. *Special Linear Reference Manual*, Analog Devices, Inc., 1992.
4. *American National Standard for Electrical Instrumentation*, Digital Measuring Instruments, ANSI C39.6-1983.
5. *User's Manual, Model 87 True RMS Multimeter*, Fluke Corporation, 1988, pp. 34–41.
6. *Instruction Manual, Model 8060A Digital Multimeter*, Fluke Corporation, 1988.

CHARLES B. NEWCOMBE
Fluke Corporation

OHMMETERS

An ohmmeter measures direct current (dc) electrical resistance (in ohms), a basic electrical quantity. Most commonly, it is one function of a digital multimeter (DMM), which combines voltage (V), current (I), and resistance (R) measurements in one instrument. These three quantities are related through Ohm's law, $V = RI$. Multiple ranges (by factors of 10) are needed to cover the wide range of resistance normally encountered in electrical/electronic devices (0.01 Ω to 10 M Ω). More specialized meters cover much lower and higher ranges (below $10^{-6} \Omega$ and above $10^{15} \Omega$).

Generally dc resistances are highly linear; that is, they obey Ohm's law to within measurement precision. This is especially true of commercial resistors and of metal conductors such as copper. Three exceptions are (1) temperature-sensitive resistors, which are heated by the power dissipated in the resistor; (2) semiconductor devices such as diodes, which may have very nonlinear $I - V$ properties; and (3) ionic conductors, including biomedical electrodes and aqueous solutions.

Resistance may be measured by applying a known voltage (constant voltage source) across the device under test and measuring the current that flows, or by applying a known current (constant current source) and measuring the voltage developed across the device. The current and voltage are limited to avoid damage to the device being tested (e.g., 5 V or 10 mA). Ohm's law is used to convert to resistance with the appropriate scaling factors. For example, a 1 mA current source will produce a 1 V drop across a 1 k Ω resistor so that a voltmeter calibrated as an ohmmeter (e.g., 1 k Ω at 1 V) will read the resistance out directly.

Digital ohmmeters have an analog input section with an operational amplifier to supply a constant current (Fig. 1). The

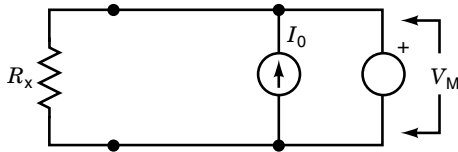


Figure 1. Digital ohmmeter.

amplifier adjusts its output so that the voltage drop produced by the current, across a known internal resistor, is equal to a reference voltage.

Switching the internal resistor changes the current (source) range and thus the resistance range. The operational amplifier will produce a constant current only up to a maximum voltage, which is chosen to be above the corresponding full scale of the readout. In any case, it is less than the amplifier supply (e.g., battery) voltage.

Loading effects of the voltmeter may limit the upper resistance range, as indicated in Fig. 2. An ideal voltmeter would have an infinite resistance and draw zero current. FET input amplifiers approximate this requirement. The meter input resistance R_{mi} and/or input leakage current provides a shunting path to the current provided by the current source so that the effective resistance R_x is reduced. Input impedances over $10^{10} \Omega$ and input under 0.1 nA are readily achievable with FET-type operational amplifiers. Most commercial DMM and DVMs (digital voltmeters) have a built-in dual-ramp-type analog-to-digital converter and also drivers for three or four digital (seven segment) displays.

Older design ohmmeters use a moving coil milliammeter or d'Arsonval meter as an analog indicator. As illustrated in Fig. 3, they may use operational amplifiers or other amplifiers. There are two general types: shunt and series. With the series type, the unknown resistance R_x is connected in series with a battery (e.g., 1.5 V) and the milliammeter. A variable resistor R_2 compensates for variable battery voltage. It is adjusted so that the meter is full scale when the leads are shorted ($R_x = 0$). The scale marked on the meter face is nonlinear, a disadvantage, since the current is inversely proportional to the resistance. At the low-current end, the resistance is infinite. This causes increasing crowding and inaccuracy at the high-resistance end of the scale. The resistance range can be reduced by introducing one or more current shunting resistors at position R_3 . A multiple position switch SW_i is used for multiple ranges.

With a shunt-type ohmmeter, the unknown resistance R_x is in parallel with the voltmeter readout (Fig. 3b), but the voltage source and standard resistor ($R_s = R_1, R_1 + R_2, \dots$) is in series. The gain resistor is adjusted so that the meter is full scale at $R_x = \infty$. Half scale corresponds to $R_x = R_s$ and 0 V to 0 Ω . The scale is obviously also nonlinear. The amplifier can be replaced

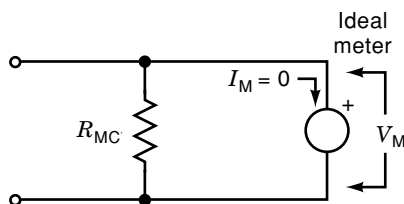
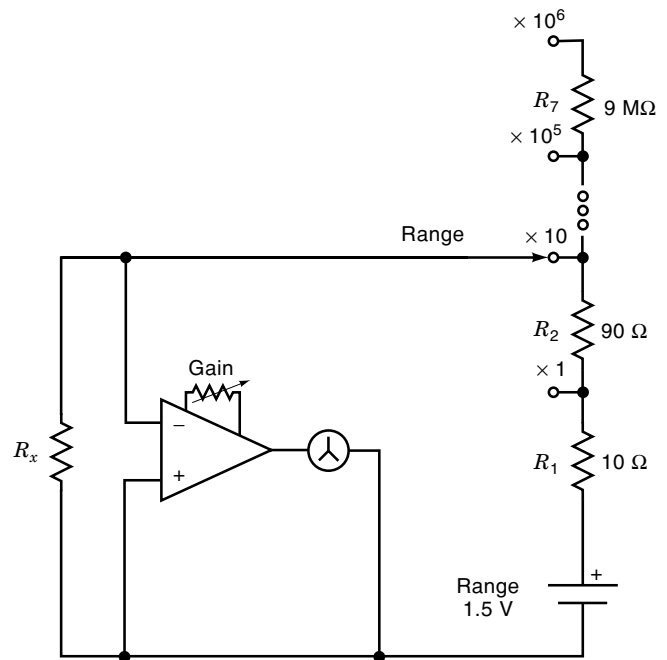
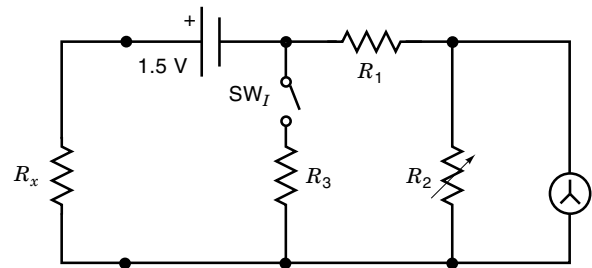


Figure 2. Voltmeter loading effect.



(a)



(b)

Figure 3. Analog indicator ohmmeters: (a) shunt type and (b) series type.

by a milliammeter acting as a finite-resistance voltmeter, but the sensitivity is less.

Low-resistance measurements may be inaccurate because contact and lead resistances become a significant fraction of the measured resistance R_x . Contact resistances in the 0.01 Ω to 0.1 Ω range are common. Often they are not reproducible. To overcome this, a four-probe or Kelvin method of measurement may be used (Fig. 4). A constant current source I_0 supplies the current through the resistor under test. The current passes through the "current" probes or leads with presumably low, but otherwise unknown resistances (R_{I1}, R_{I2}). Assume that the total load on the current source ($R_x + R_{I1} + R_{I2}$) is within its linear output voltage range; the source will deliver the specified current I_0 independent of the load, even if the contact and lead resistances change. The voltage drop as measured across the resistance and also, if unloaded, at the voltage probes or leads V_x is exactly $I_0 R_x$. The differential or instrumentation

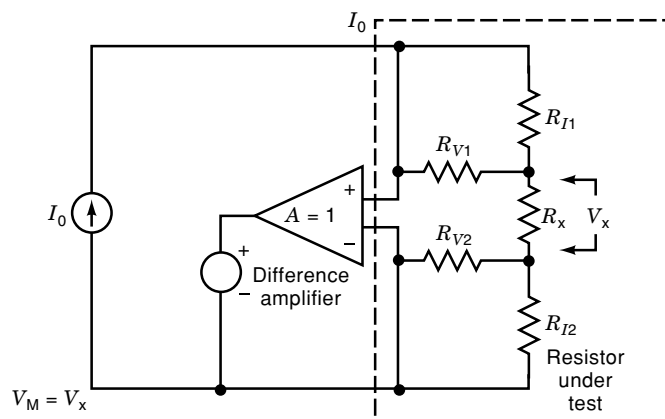


Figure 4. Example of a four-probe or Kelvin method of measurement.

amplifier, which must have a high common mode rejection and high input resistance, transfers this balanced voltage to the measurement voltmeter, which is unbalanced with respect to ground (one input grounded). Thus $R_x = V_x I_0$, independent of contact resistance. Some ohmmeters (over 20 M Ω) intended for very high resistance such as insulation resistance use a relatively high source voltage (e.g., 100 V or more). This is because insulators will absorb (soak) charge for some time after the potential is applied so that a stable reading in a reasonable time is not obtained unless higher voltages are used. Care must be taken in this case not to damage the device being tested by overvoltage.

Reading List

- L. Bobrow, *Fundamentals of Electrical Engineering*, 2nd ed., Oxford, UK: Oxford University Press, 1996, pp. 975–977.
- D. Finch and C. Christiansen, *Electronic Engineers' Handbook*, 3rd ed., New York: McGraw-Hill, 1989, pp. 17–14 to 17–20.
- R. Graf, *Encyclopedia of Electronic Circuits*, vols. 1, 3, Blue Ridge Summit, PA: TAB Books, 1991, I-549, III-540.
- L. D. Jones and A. F. Chin, *Electronic Instruments and Measurements*, New York: Wiley, 1983.
- Keithley Instruments, *Low-Level Measurements*, 4th ed., Cleveland, OH: Keithley Instruments, 1992.
- R. Northrop, *Introduction to Instrumentation and Measurements*, Boca Raton, FL: CRC Press, 1997, pp. 390–395.
- R. Pallas-Areny and J. G. Webster, *Sensors and Signal Conditioning*, New York: Wiley 1991, chap. 3.
- M. B. Stout, *Basic Electrical Measurements*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1960.

DAROLD C. WOBSCHELL
D. D. L. CHUNG
State University of New York at
Buffalo

ELECTROMETERS

An electrometer is a sensitive, high input impedance instrument used for specialized measurements of dc voltage, dc current, and resistance. An electrometer may also measure electrical charge (a coulombmeter). Superficially, an electrometer

is an elegant dc multimeter. In reality, an electrometer is a specialized voltage and current measurement device possessing exceptionally high impedance. This very high input impedance, more than any other technical criterion, is the distinguishing characteristic of instruments in the electrometer class. The similarity to a dc multimeter makes electrometer use fairly intuitive, although a thorough appreciation of performance determinants (e.g., noise and settling time) and careful control of materials, lead routing, and other fixturing details are essential to obtaining the full accuracy and resolution within the capabilities of these exceptional instruments.

Few instruments in measurement science can exceed the electrometer's dynamic range of units. A quality electrometer is capable of presenting very small [atto-amperes (aA) (10^{-18} A)] and very large [petaohms (P Ω) (10^{15} Ω)] units of measure unfamiliar even to experienced equipment users. The general reader will find the standard prefix definitions in Table 1 helpful.

The term electrometer is found in the earliest literature of the electrical sciences—it was employed in the contemporary sense by none other than James Clerk Maxwell (1). Instruments capable of measuring very small currents are sometimes called electrometers, although such usage may be a misnomer (this will be discussed below in “Applications, Current Measurements”).

Interestingly, the *IEEE Standard Dictionary of Electrical and Electronics Terms* has no definition of an electrometer, although it tells the reader an electrometer tube is “a vacuum tube having a very low control-electrode conductance to facilitate the measurement of extremely small direct current or voltage” [(2), p. 425]. Lack of specificity is not uncommon, as similar findings are evident in some standard handbooks for electrical and electronics engineers (3), physicists (4), and in textbooks (5–7). For purposes of the following discussion, an electrometer is taken to be an instrument which, as a consequence of possessing exceptionally high input impedance, may be used to make sensitive measurements of voltage, current, and charge.

However ill-defined an electrometer may be in a formal dictionary sense, an electrometer has certain characteristic features. Foremost among the electrometer's defining features

Table 1. Standard Prefix Definitions

Symbol	Prefix	Exponent
y	yocto-	10^{-24}
z	zepto-	10^{-21}
a	atto-	10^{-18}
f	femto-	10^{-15}
p	pico-	10^{-12}
n	nano-	10^{-9}
μ	micro-	10^{-6}
m	milli-	10^{-3}
—	—	—
k	kilo-	10^3
M	mega-	10^6
G	giga-	10^9
T	terra-	10^{12}
P	peta-	10^{15}
E	exa-	10^{18}
Z	zetta-	10^{21}
Y	yotta-	10^{24}

is exceptionally high input impedance. Input impedances are measured in hundreds of teraohms $[(T\Omega) (10^{12} \Omega)]$ or even in petaohms in better quality commercial equipment. The electrometer's high input impedance is responsible for many of the instrument's operating characteristics. Other significant features of an electrometer are the ability to measure small voltages, small currents, and low-level electrical charges. These features alone, however, do not necessarily place an instrument in the electrometer class, for many higher quality, general purpose instruments can measure signals in the microvolt and submicroampere range *if source impedance is low*.

Voltage or current measurement is so commonplace as to be unremarkable in routine laboratory work. This situation prevails because most sensors and virtually all amplifiers present low output impedance to the external world. Instrument makers have gone to considerable pains to make measurement equipment simple to use in very general situations. Even inexpensive instruments have input impedances in the megohm range (clearly, this discussion excludes voltage and current measurements for audio, RF, and other specific applications where impedance matching is a major consideration or where electrometer time constants are prohibitively long). Laboratory quality instruments present relatively high input impedances to the typically low impedance sources producing signals to be measured.

Some measurement applications require an instrument with exceptionally high input impedance because the source itself has unusually high output impedance. Examples of such high impedance signal sources are large-ratio voltage dividers used in high-voltage measurements, ECG scalp electrodes, pH probes, ionization chambers, large resistances, photodiodes, dielectric measurements, and certain semiconductor measurements. The electrometer's high input impedance is required to prevent or reduce circuit loading by the measuring instrument, that is, undesired voltage or current dividers established between source output impedance and measurement device input impedance.

The electrometer's high input impedance is difficult to establish and maintain, is influenced by many subtle error sources, and is also the ultimate reason why electrometers are fundamentally dc or, at best, very low frequency instruments. Shunt capacitance at the instrument input is a physical reality that cannot be entirely eliminated, and the interaction of input shunt capacitance and source resistance establishes the frequency limit of the instrument. This frequency limit is so low as to make an electrometer a dc instrument, and the associated input settling time is a consideration of considerable importance.

The significance of input impedance is easily appreciated in the context of voltage measurements (for example, as in an oscilloscope). If we wish to determine the voltage across a circuit impedance Z_c , a very high impedance measurement circuit with impedance Z_m , produces little or no circuit disturbance. An ideal measurement circuit with infinite input impedance (infinite resistance with zero capacitance) would produce no circuit disturbance at all. The circuit impedance in parallel with the ideal measurement circuit produces a net impedance that is the circuit impedance itself, that is,

$$Z_p = \frac{Z_c Z_m}{Z_c + Z_m} \quad (1)$$

and

$$\lim_{Z_m \rightarrow \infty} Z_p = Z_c \quad (2)$$

Although measurement device impedance may be very large, it is not infinite for a physically realizable instrument. Thus, as source impedance increases, the error produced by the measurement device's finite input impedance becomes an increasingly greater contribution to total measurement error. A measurement device of the electrometer class closely approximates the theoretical idealization of infinite input impedance near dc, thereby reducing circuit-loading effects and improving voltage-transfer accuracy.

Many general instrument users experience input impedance problems only at higher frequencies, when shunt capacitance of common instrument inputs can no longer be neglected. The most familiar example of this phenomenon is high-frequency signal attenuation produced by the shunt capacitance of an oscilloscope probe. The -3 dBV response frequency and net impedance of the probe's parallel RC input are well known. As operating frequency increases, probe capacitance reduces the oscilloscope's net input impedance. At sufficiently high frequency, probe capacitance begins to influence and then dominate circuit response. Said another way, as impedance decreases with frequency, the oscilloscope probe places an increasingly greater load on the circuit being measured. The same effects are observed in the interplay between generalized source and measurement impedances, making discussions of oscilloscope probe losses instructive to the general reader as analogous background information (8–10).

The need to avoid input circuit loading is a major consideration requiring the use of an electrometer (the other is noise). Consider the simple circuit of Fig. 1, which represents the low-frequency impedances of a pH electrode (R_e) and the measuring instrument (R_m). Let the voltage actually produced by the pH electrode be V_e , and the voltage actually transferred to the voltage measurement circuit be V_m . By simple voltage division, the low-frequency signal transferred to the input of the voltage measurement elements is given by

$$V_m = \left(\frac{R_m}{R_e + R_m} \right) V_e \quad (3)$$

If, for example, a pH electrode has $100 \text{ M}\Omega$ source impedance (11), a common voltage measurement instrument with $10 \text{ M}\Omega$ input impedance would introduce a grossly unacceptable

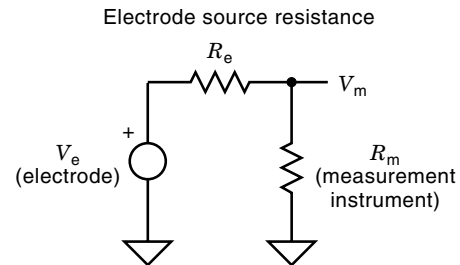


Figure 1. pH electrode source resistance and measurement circuit input resistance form voltage divider.

measurement error, even though the same 10 M Ω impedance would be thoroughly satisfactory for general laboratory work. An input impedance of 1 G Ω would be only just acceptable, forming a voltage divider transferring roughly 90% of the electrode signal voltage (at best) to the voltage-measurement elements. An input impedance of nearly 10 G Ω is required if the measuring instrument is to transfer 99% of the pH electrode voltage to the voltage-measuring circuit elements.

Establishing and maintaining input impedances in the high G Ω and T Ω ranges is no simple task. Minuscule error currents, ordinarily negligible and themselves difficult to measure directly, become serious errors. Tiny shunt capacitances in combination with such large impedances severely limit the electrometer's time-domain performance. Thermal effects, humidity, dust, circuit board contaminants, vibration, thermocouple effects, material properties, and other considerations all weigh heavily in electrometer design. The old political adage applies: the devil is in the details.

APPLICATIONS

The electrometer's superficial similarity to a dc multimeter makes fundamental instrument application simple and intuitive. Although electrometer application basics are easy to appreciate, it is not simple to obtain the greatest possible accuracy and resolution from the instrument. Error sources can be very subtle, and test fixturing is unusually important. The following material provides an introduction to the use and application of electrometers and related instruments. Little design-related information is included. Excellent introductions to design considerations are available in (17) and in the instrumentation and measurement literature.

Noise Analysis

Electrometer performance analysis is tied intimately to noise analysis. The performance of any measurement instrument is ultimately limited by noise. Noise analysis is especially important to understanding electrometer applications properly because electrometers routinely tread the edge of the theoretical noise limit. A few major results from noise analysis are given below. The more advanced reader will wish to consult other references for a detailed treatment. Noise analysis is covered in a great many texts, and only a few representative examples are cited. Communication theory is a comfortable context for many readers (12–14), more specific discussions [see (15,16)] are useful for their treatment of specific noise mechanisms.

External noise sources, including the interaction of test articles and cabling with electrical, magnetic, and propagating electromagnetic fields, can be attenuated, minimized, or eliminated by test fixture design and shielding. External noise effects are not to be trivialized and are discussed later, but there is often some redress for external noise effects. Internal noise sources, such as thermal noise (Johnson noise) of resistors, shot noise from movement of charge carriers in amplifier input bias currents, and low-frequency ($1/f$) noise, are major noise sources within the electrometer itself. These internal noise sources exist in any physical device and ultimately establish the noise floor (hence, the ultimate sensitivity limit) of the electrometer, even if all external noise sources were reduced to zero.

These same noise mechanisms also exist in source impedances. Proper appreciation of the best theoretically possible noise limit is essential if the user is to obtain reliable measurements from any instrument.

Thermal noise and shot noise are the most common noise mechanisms governing the performance limits of measurement systems, and their significance to electrometer applications is considerable. Root mean square (rms) thermal noise current is given by

$$i_{T,rms} = \sqrt{\frac{4kT}{R}}(NBW) \quad (4)$$

where k is Boltzmann's constant (1.38×10^{-23} J/K), T is temperature in kelvins, R is resistance in Ω , and NBW is noise bandwidth in hertz. Root mean square shot noise is given by

$$i_{s,rms} = \sqrt{2qI_B}(NBW) \quad (5)$$

where q is electron charge (1.60×10^{-19} C), I_B is bias current in amperes, and NBW is noise bandwidth in hertz. These noise currents may be converted to equivalent noise voltages using knowledge of circuit impedances.

Consider the electrometer input stage. Thermal noise is an explicit function of absolute temperature. Because semiconductor device bias currents vary (ordinarily, they increase) with temperature, shot noise is also a function of absolute temperature and noise bandwidth. Noise signals produced by various mechanisms are often treated as random white Gaussian signals, allowing individual noise components to be combined as a quadrature sum. Although $1/f$ noise is not Gaussian because of its frequency dependence, $1/f$ noise is often included in an overall noise envelope, a piecewise linear fit including the effects of major noise sources. In the near-dc regime of the electrometer, $1/f$ noise is incorporated in the lumped specified noise performance (regardless of origin) of the instrument.

Although noise considerations are sometimes ignored in routine work, noise relationships are important determinants of electrometer performance and must be understood. The instrument designer understands that noise bandwidth and input bias currents must be minimized and that temperature should be maintained at a reasonable level. The instrument user must recognize that certain physical realities (e.g., noise) impose limits on instrument performance and that both system design and instrument specifications must conform to these realities. For example, a specification requiring or implying 1 nV resolution from a 1 MHz bandwidth source should be recognized for what it is: a practical impossibility. Although an electrometer generates measurements with apparent precision, the user can avoid embarrassment by interpreting instrument measurements in the light of the theoretical noise limits of the problem.

Time Constant

Electrometers are distinguished from other test instruments by their uncommonly high input impedance. As Fig. 2 demonstrates, the electrometer input circuit may be represented by a resistance R_m and a capacitance C_m . In a voltage measurement configuration, the signal source impedance R_s and the electrometer input impedance form a voltage divider connection.

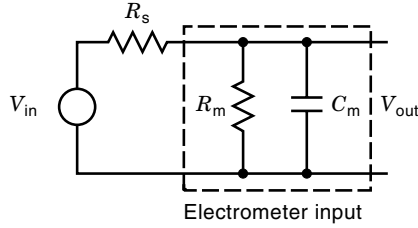


Figure 2. Equivalent circuit of source resistance and electrometer input impedance.

In this case, the voltage transfer function in the Laplace transform domain is obtained as

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{R_m}{R_s + R_m} \left[\frac{1}{1 + sC_m R_p} \right] \quad (6)$$

where

$$R_p = \frac{R_s R_m}{R_s + R_m} \quad (7)$$

The equivalent time-domain step response is the familiar first-order step response given by

$$V_0(t) = (1 - e^{-t/\tau})u(t) \quad (8)$$

where $u(t)$ is the unit step function and $\tau = C_m R_p$ is the time constant.

Although realizable electrometers do not actually possess infinite real input impedance, it is very frequently true that $R_m \gg R_s$. If R_m is sufficiently large with respect to R_s , as it often will be, then the result given in Eq. (6) simplifies to

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{1 + sC_m R_s}$$

and the result given in Eq. (7) simplifies to $R_p \cong R_s$. The time constant, τ , is determined by the source resistance and electrometer input capacitance according to $\tau = C_m R_s$ when this approximation for R_p is used.

Previous discussion related the dc divider ratio to the overall signal measurement or voltage transfer accuracy. However, there is an associated time-domain response as given previously which introduces significant error if signal sampling is done incorrectly. This time-domain consideration is often described as the settling time, that is, the amount of time required for an input step change to settle to within some arbitrary accuracy of the total step. Settling time is most often expressed in terms of the time constant τ . For example, if only a single time constant elapses between application of a step change and a measurement, evaluation of Eq. (8) reveals that only 63% of the step amplitude will be measured. Table 2 indicates the relationship between elapsed time constants and relative measurement accuracy. At least five time constants (5τ) must be allowed for settling if the step change is to be represented with 99% fidelity. In highly precise applications, those in which accuracies are measured in parts per million (ppm), settling times may be much longer. For example, fourteen time constants (14τ) are required for a settling error less than 1 ppm.

Table 2. Relationship Between Elapsed Time Constants and Relative Measurement Accuracy

Time Constants Elapsed ($\tau = RC$)	Fraction of Step Attained	Error (ppm)
τ	0.632121	367,879
2τ	0.864665	135,335
3τ	0.950213	49,787
4τ	0.981684	18,316
5τ	0.993262	6,738
6τ	0.997521	2,479
7τ	0.999088	912
8τ	0.999665	335
9τ	0.999878	122
10τ	0.999955	45
11τ	0.999983	17
12τ	0.999994	6
13τ	0.999998	2
14τ	>0.999999	<1

Thus, accuracy considerations require the passage of not one but many time constants before the instrument input settles. As shown in Table 2, a settling time of 4τ to 5τ , longer if possible, should be allowed to benefit from a good electrometer's basic accuracy. Settling time is a significant system constraint and is one of the reasons why an electrometer is a dc instrument.

Because an electrometer has very high input impedance, source resistance and input capacitance usually determine the input time constant. Response time is reduced only by reducing electrometer input capacitance or by reducing source resistance (which often is not possible). Various design and construction alternatives, such as driven guard cables, circuit board guard rings, and special mechanical structures are used to reduce input capacitance, thereby improving settling time (17). These techniques may increase net capacitance to ground.

Leakage Current

In the context of an electrometer, leakage current refers to tiny currents present in the instrument input leads. In an ideal instrument, leakage currents are identically zero. Readers familiar with operational amplifier design understand that similar leakage currents interact with circuit impedances to produce various categories of offset errors and other effects. This general problem is especially troublesome in electrometer design for at least two reasons.

First, electrometers are employed with signal sources having very large source impedance and, as noted already, the electrometer itself presents exceptionally high input impedance. Thus, instrument leakage current preferentially finds a path to signal common through the signal source impedance. In the voltage measurement mode, very small electrometer leakage currents through the relatively large signal source impedance can produce highly undesirable voltage offset components at the electrometer input. This is the familiar problem of input offset voltage production in operational amplifier design. In current and charge measurement modes, instrument leakage currents directly interfere with the current being measured.

Second, although the complication is less obvious, recall that shot noise is proportional to the square root of current. Thus, as leakage current increases, shot noise increases. Therefore, input leakage current produces two highly undesirable effects. It directly obscures or interferes with the parameter (voltage,

current, or charge) being measured, and it contributes additional noise (uncertainty) to the measurement.

Careful attention to device selection, component matching, thermal environment, and details of assembly and fabrication minimize instrument leakage current. Special signal processing techniques may be used to reduce the apparent effects of signal leakage current, although primary emphasis should be on error prevention rather than error compensation.

Ammeter

Electrometers are dc instruments optimized for sensitivity and accuracy at low (usually microampere and lower) current levels. Many contemporary commercial instruments have maximum input currents of 20 mA, making them of limited utility as general-purpose ammeters. Older instrument designs still in production have higher maximum input current ratings, although full-scale input current is still usually less than 1 A. An electrometer may be used to measure the voltage developed across an external current shunt with great accuracy. Although this approach extends the useful full-scale current range of the composite measurement system, long settling time, limited maximum input current, and cost usually contraindicate using an electrometer as a general-purpose ammeter.

Current Measurement. Current measurement applications include tasks, such as photodiode responses, photodiode dark current variations with temperature, semiconductor current measurements, and other measurement problems characterized by the need to measure small currents and also by the need to handle a wide dynamic range of signal currents. Commercial instruments in computer-controlled test sets accommodate the dynamic range problem by CPU-generated instrument range changes or by the use of logarithmic outputs. The literature contains descriptions of application-specific current measurement devices which approach the dynamic range problem in other ways.

Electrometers are widely employed for low-level current measurement. A device known as a *shunt picoammeter* is an electrometer variant with an input current shunt optimized specifically for low-level current measurements. The differences between an electrometer-based shunt picoammeter and an electrometer are not necessarily substantial, as the two instruments are fundamentally the same in most respects. Identifiable differences between an electrometer and a shunt picoammeter are associated with the shunt resistor and other optimizations for current measurement. The shunt picoammeter has a somewhat lower voltage burden, somewhat faster response time, somewhat reduced sensitivity, and lower cost (17) than an electrometer. This optimization is obtained at the expense of overall versatility.

One especially common and useful application circuit deserves special consideration in this discussion. This circuit, usually called a *transimpedance amplifier* or *feedback picoammeter* (also a current-to-voltage converter) is widely used to measure small currents in a multitude of applications, for example, in conjunction with photodiode detectors in computed tomography (CT) detectors and similar low-signal tasks. Because transimpedance amplifiers are used to measure very small currents, they and apparatus using them are sometimes erroneously described as electrometers, a practice difficult to discourage in the absence of a formal definition of an

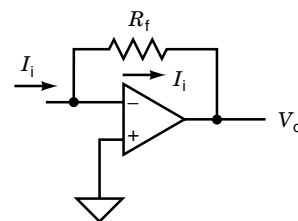


Figure 3. Transimpedance amplifier used as picoammeter.

electrometer. Not all authorities agree that the transimpedance amplifier configuration constitutes an electrometer. One school of thought maintains that the transimpedance amplifier is a picoammeter but *not* an electrometer, although papers can be found in the engineering literature which make the equivalence. Nevertheless, the transimpedance amplifier configuration clearly provides useful analog signal processing of very small currents.

The transimpedance amplifier uses an inverting amplifier connection, as shown in Fig. 3. Assuming an ideal operational amplifier having zero input current, the amplifier's output produces an output current equal exactly to the input current I_i appearing on the inverting input, thereby satisfying Kirchhoff's current law. To do so, the amplifier output voltage V_o , must satisfy the relationship

$$V_o = -R_f I_i$$

The feedback resistance R_f , is often nothing more than a single resistance. In many cases, a plurality of feedback resistances are switch selectable. The choice of multiple feedback resistances is advantageous when a wide dynamic range is anticipated.

Several features should be apparent. The input of the amplifier is a virtual ground, making the input impedance of the device a great deal less than what would normally be considered useful for electrometer applications. It also requires that the signal source behave like an ideal current source. Current errors are dominated by the input bias current of the amplifier, although these errors are relatively low with contemporary FET-input devices. The transimpedance configuration is highly desirable when something other than near-dc operation is required. The -3 dBV frequency is established primarily by the feedback resistor R_f and capacitance associated with the feedback resistor.

Charge Measurement

Electrometers measure charge in at least two ways. Charge may be determined by current integration using a feedback picoammeter with a standard capacitor rather than a resistor in the feedback loop, or charge may be determined by accurately measuring the voltage produced when the charge is transferred into a known (standard) capacitor. In both types of measurement, charge measurement is made possible by the electrometer's exceptionally low input bias currents. The virtual absence of instrument bias current introduces little error into the integration of an external current. In the case of stored charge measurement, the near-zero instrument input bias current removes or deposits almost no charge from or into a known external capacitance, allowing charge to be determined directly from the relationship $Q = CV$.

Table 3. Instrument Minimum Resolvable Inputs and Maximum Full-Scale Inputs

Digits	Model	Volts	Amps	Coulombs	Ohms	R_{in} (max)
<i>Electrometer^a—Minimum Resolvable Input / Maximum Input</i>						
Analog	610C	10 μ V/100 V	1 fA/300 mA	1 fC/10 μ C	1 Ω /100 T Ω	100 T Ω
4 $\frac{1}{2}$	614	10 μ V/20 V	10 fA/2 mA	10 fC/20 nC	1 Ω /200 G Ω	50 T Ω
4 $\frac{1}{2}$	617	10 μ V/200 V	100 aA/20 mA	10 fC/20 nC	100 m Ω /10 P Ω	200 T Ω
4 $\frac{1}{2}$	642	20 μ V/11 V	10 aA/100 nA	800 aC/100 pC	N/A	10 P Ω
4 $\frac{1}{2}$	6512	10 μ V/200 V	100 aA/20 mA	10 fC/20 nC	100 m Ω /200 G Ω	200 T Ω
5 $\frac{1}{2}$	6517A	10 μ V/200V	100 aA/20 mA	10 fC/2 μ C	10 Ω /100 P Ω	200 T Ω
<i>Nanovoltmeter^a—Minimum Resolvable Input / Maximum Input</i>						
Analog	155	± 1 μ V/ ± 1 kV				100 M Ω
6 $\frac{1}{2}$	182	1 nV/3 mV				10 G Ω
7 $\frac{1}{2}$	HP 34420A	100 pV/100 V				10 G Ω
<i>Picoammeter^a—Minimum Resolvable Input / Maximum Input</i>						
3 $\frac{1}{2}$	480		1 pA/2 mA			
4 $\frac{1}{2}$	485		100 fA/2 mA			
5 $\frac{1}{2}$	486/487		10 fA/2 mA			
3 $\frac{1}{2}$	HP4140B		1 fA/pA (est)			

^a Model numbers given are manufactured by Keithley Instruments, Cleveland, OH, unless otherwise specified.

In a closely related application, electrometer measurements can be used for experimentally estimating small unknown capacitances by determining the charge transferred from a standard capacitor to an unknown capacitor under conditions in which charge must be conserved.

Voltmeter

Voltage measurement is among the most common applications for electrometers. Just as manufacturers produce an electrometer variant for optimized low-level current measurement (the picoammeter), manufacturers also produce an electrometer variant optimized for low-level voltage measurement, commonly called a *nanovoltmeter*. Although optimized for small voltage measurements, nanovoltmeters ordinarily do not possess the exceptionally high input impedance associated with true electrometers because measuring exceptionally small voltages assumes lower source impedances. If this were not the case, source thermal noise would easily obscure the measurement. Nanovoltmeter optimization sacrifices the ability to work with the highest possible source impedances. It is usually not cost-effective to use an electrometer or nanovoltmeter as a general-purpose voltmeter.

Resistance Measurement

Electrometers are especially useful for measuring very high resistances or resistivities. This ability usually requires a current source in the electrometer, a configuration sometimes called a “source-measurement” combination. The electrometer measures current driven through the unknown resistance and also measures the resulting voltage developed across the unknown.

Sensitivity and Resolution

The sensitivity of an instrument is the smallest quantity that can be measured. Resolution, a closely related concept, is the smallest quantity that can be displayed. The resolution of an electrometer, as that of any other instrument, depends on full-scale amplitude and quantization accuracy (a digital display is assumed). At this time, most commercial instruments have

4 $\frac{1}{2}$ digit displays and a 20,000:1 ratio between full-scale signal and the minimum resolvable quantity corresponding to that full-scale signal. On the instrument’s most sensitive scale, sensitivity and resolution become essentially the same thing. In the discussion which follows, sensitivity is used in the sense of the smallest measurable quantity capable of display on the most sensitive scale.

The state of the art progresses rapidly. The minimum resolvable inputs (best resolutions) and maximum full-scale inputs shown in Table 3 are typical now of contemporary commercial instruments.

It is worth remembering that the quantity of charge associated with a single electron is roughly 0.160 aC, thus, a sensitivity of 100 aA corresponds to a current of roughly 629 electrons per second. The notion of electrical current is a macroscopic concept in which large numbers of charged particles (electrons) are in motion per unit time. At current levels measured in hundreds of attoamperes, measurement specialists often begin to speak of “electron counting,” for the notion of electrical current as a macroscopic behavioral average begins to break down. Current fluctuations become more pronounced. At the most fundamental level, each electron carries a 0.160 aC packet of electrical charge. When only relatively few electrons are in motion, the concept of electrical current increasingly becomes a statistic of average rates of arrival. Similarly, the ability to resolve electrical charge in steps of 10 fC corresponds to charge quantization in steps of slightly less than 63,000 electrons.

Commercial electrometers usually have voltage sensitivities of 10 μ V out of 200 mV full scale. On resistance measurements, resolutions usually range from 100 m Ω (out of a 2 k Ω full scale) to 1 G Ω (out of 200 T Ω full scale). Resistance measurements are most often implemented using a constant current source to drive the device under test with Kelvin connections to the voltmeter section. Resistance is a calculated value that takes advantage of the electrometer’s ability to make high-resolution voltage and current measurements.

Nanovoltmeters are electrometer-like instruments optimized for measuring small voltages. Nanovoltmeters do not possess the exceptionally high input impedance of a true electrometer, although they can hardly be regarded as low input impedance devices. At the present time, a high-quality

standard commercial instrument is sensitive to 1 nV out of a full-scale signal of 3 mV.

Because a great many small-current applications do not require the expense, versatility, and exceptional sensitivity of a full-blown electrometer, some instruments are optimized for small current measurements and reduced cost. These instruments are called picoammeters. As their name suggests, the available sensitivity is usually 1 pA, although at least one commercial picoammeter instrument has an advertised sensitivity of 1 fA. Picoammeters are exceptionally useful for laboratory work, as they accommodate the current measurement requirements of a large percentage of precision applications.

Nullmeter

One common early use of electrometers was to indicate null, or zero voltage, in precision bridges. This application is still important, and variations of the concept are found in servo loops and certain other automatic control applications where it is desirable to drive a loop to a null condition.

Oddly enough, nulling applications are somewhat vulnerable to an infrequently considered, yet surprisingly difficult, type of error, namely, the difficulty in determining exactly what is zero volts. A small piece of copper wire (a shorting bar) across the input terminals does not provide a completely satisfactory answer to this seemingly simple question, because such a simple approach is vulnerable to subtle thermocouple and low-level motion effects.

In response to this need and for calibration checks, instrument manufacturers have special electrometer accessories known as *low thermal shorting plugs*. These assemblies are indeed little more than sophisticated shorting bars, but they are important because they allow the user to verify that an indicated zero volt reading or null actually represents a zero volt differential when thermal and mechanical effects are controlled. The material of all connecting wires needs to be known.

Circuit of Electrometers

Complex networks of more than one electrometer require good measurement practices. The particular sensitivity of electrometer-class instruments demands careful consideration of the physical locations and connections of apparatus. Unanticipated input capacitance increases, low-level ground loops, thermal gradients, poor isolation practices, and lack of good shielding from stray fields are all common problems in low-level test applications. An electrometer in the test environment does not, by itself, make such problems more acute, but the consequences of poor test practices may be more severe because of the high impedance of the signal source (the fact that an electrometer is employed suggests that the signal source is a high-impedance source as well).

Many general instrument users are aware that thermocouple effects perturb delicate dc measurements. Most general users are surprised to learn that simple movement and vibration of the wire interconnecting test articles and test apparatus are a significant problem due to the triboelectric effect, the generation of small voltages as a conductor moves within an insulator. Likewise, a wire moving or swinging in a magnetic field gives rise to undesired signals by electromagnetic induction, just as in an electric generator. Insulator deformation can result in piezoelectric behavior which generates error signals. Such considerations as relative humidity, dust, and cleanliness

of the apparatus often have major impact when the investigator wishes to partition signal currents into current bins smaller than one fA. Dust adhering to a fingerprint across two circuit traces represents a surprisingly low impedance (relative to impedances of several hundred T Ω or higher), especially on a humid day. For these reasons, some commercial instruments have ancillary displays of temperature and relative humidity on the electrometer face.

Measurement Standards

The ability to measure fundamental electrical quantities, namely, voltage, current, and charge, with such sensitivities is quite remarkable. Various circuit elements, such as voltage references, digitizers (A/D converters), amplifiers, and related functional devices, can be isolated and refined during the normal course of engineering development. The process of insuring that measurements made worldwide respond uniformly to the same potential differences is one aspect of the problem of measurement standards. The actual primary standards themselves, for example, the standard volt, are the province of various national laboratories and international standard-setting bodies [in the United States, the National Institute of Science and Technology (NIST)].

Development and maintenance of standard units is a fascinating study in its own right but is outside the scope of this article. The instrument user is normally unconcerned with measurement standards other than to ensure that the instrument is properly calibrated periodically. Measurement standards involve a multitiered system of primary standards (at the national and international level), secondary standards (calibrated to the primary standard), and working standards calibrated to secondary standards. For example, various national and international bodies determine what absolute potential difference constitutes the standard volt and maintain the physical artifacts producing the standard volt. The instrument manufacturer maintains standards traceable to the primary physical standard and makes sure that all subsequent operations (quantizations or otherwise) ultimately agree with the physical standards.

ADVANCED TOPICS

The manufacturer specifies performance of an electrometer instrument which may be considered to begin at the input connector. Mechanical and electrical details inside the instrument are carefully designed to maintain high input impedance and to reduce error contributions from a great many sources. Conceptually, the electrometer is a highly sophisticated dc multimeter, and the operating concepts are intuitive. The electrometer is an instrument capable of operating at the theoretical limit of source thermal noise, and there is little the user can do to improve on the instrument's performance. There is, however, much the user can do to degrade the instrument's performance if the external test apparatus is carelessly made. The advanced user ordinarily does not struggle with circuit details on the electrometer side of the input connector. The instrument manufacturer does that, but the advanced user expends effort to connect the test article to the instrument while minimizing stray leakage paths, electrostatic effects, piezoelectric effects, triboelectric effects, inductive coupling, thermocouple effects, and a great many other low-level problems.

It is not possible to anticipate all problems that arise. Each application is different in its details, which is where the problems hide. Major considerations are introduced only briefly here. An excellent and detailed practical discussion is found in (17).

Input Guarding

Input guarding is the use of a driven guard structure around the input signal lines. This may be a guard conductor in a cable or guard ring traces on a circuit board. The practice is common in electrometer applications, and a guard connector is ordinarily available at the instrument input. The principle is to control guard ring potential so as to minimize potential difference between the input signal lines and surrounding structures, thereby nulling leakage current, although capacitance to ground increases. This is an important consideration, because instrument response time (settling time) is directly affected by input capacitance. Uncontrolled input capacitance introduces measurement error by allowing insufficient settling time for the measurement cycle.

Cable Selection and Preparation

In general, it is probably best to mount test cables rigidly so that capacitance and leakage are constant and controlled. This is not always possible. Specialty cables are available which provide shielding in the usual sense and also minimize triboelectric and other motion-related noise. Cable preparation is extremely important if the full benefits of specialty cables and connectors are to be realized.

Materials Selection

Materials employed as substrates, insulators, and stand-offs in the test apparatus substantially affect measurement quality. In large part, this dependence on materials is more pronounced in electrometer applications because source and instrument impedances are very high. Materials must be selected with the utmost care, and this often involves compromises between material properties, machining qualities, cost, and other factors.

Cleanliness

Thoroughly appropriate material selection can be defeated by lack of cleanliness. Surface contamination from skin oils, soldering residues, dust, and similar sources produces stray leakage paths which interfere with the intended signal path. This is especially true in relatively high humidity, and the combined effect of variable or elevated humidity and a contaminated surface can be very, very serious. Wear clean, lint-free gloves when you work on the apparatus or be prepared to clean it thoroughly after the work is complete.

Mechanical Rigidity

Mechanical rigidity of the test apparatus is important for several reasons. Triboelectric effects arise from the relative motions of a conductor and associated insulator. If there is no relative motion, there is no triboelectric effect. Additionally, some materials are more piezoelectric than others, and eliminating motion eliminates the opportunity for piezoelectric signal generation.

Electromagnetic Shielding

Electromagnetic shielding is employed to eliminate or minimize the effects of external (relative to the test article) electromagnetic fields. This topic is covered in great details in many sources (16). Low-frequency magnetic shielding is difficult to implement, but a reasonable number of options are available if the test cell volume is relatively small. Test leads and the test article together constitute an inductive loop. The effects of inductive signal pickup can be minimized by reducing loop area and by reorienting the loop geometry of the test apparatus relative to the interfering field source. Electrostatic shielding is ordinarily much simpler to implement.

Temperature Control

Temperature control is important because temperature is a consideration in thermal noise, semiconductor leakage current, thermocouple effects, and other similar phenomena. Semiconductor leakage currents, in particular, are strong functions of temperature. Measuring an offset voltage for later subtraction from a reading includes the tacit assumption that temperature (hence, offset) remains constant over the measurement interval. Thermal gradients across semiconductor devices can introduce unpredictable offset behavior, and for this reason most instruments have a specified warm-up period to guarantee full accuracy. Thermal gradients also arise from air flow around the test apparatus, and thus air flow should be controlled. Every connection in the test apparatus is a thermocouple junction, and temperature affects each of them.

Humidity Control

Humidity is a problem less obvious than temperature, but it also must be controlled. Humidity changes are especially troublesome because the insulating properties of dielectrics often depend on humidity. If someone handles a circuit board with bare hands and leaves a fingerprint across the measurement terminals, the resulting high (but finite) impedance leakage path strongly depends on relative humidity. Electrometers may have input impedances of 100 T Ω or more, the signal source being measured may have a source impedance measured in many G Ω . A leakage path, even one measured in hundreds of M Ω , can have a very detrimental impact on measurements. If a leakage path is a strong and highly variable function of relative humidity, as is true of many contamination paths, the effect can be very serious yet erratic.

Some upper end electrometers have built-in humidity indicators. This gives the operator an indication of the humidity inside the instrument, a consideration useful in obtaining the highest possible instrument accuracy, but humidity may very well exert a more serious influence on the test apparatus.

Light

The photoelectric effect was discovered early in the history of semiconductors. The photoelectric effect can be a problem during semiconductor measurements, especially measurements made at the die level. Diodes are especially light-sensitive, but care is warranted in all cases. If there is any question about the light-tightness of a device package, be aware that ambient light falling on the wrong junction can be a problem.

BIBLIOGRAPHY

1. J. C. Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., New York: Dover, 1954, vol. 1, pp. 326–340.
2. C. J. Booth (ed.), *The New IEEE Standard Dictionary of Electrical and Electronics Terms*, 5th ed., Institute of Electrical and Electronics Engineers, New York: IEEE Press, 1993.
3. D. G. Fink and J. M. Carroll, *Standard Handbook for Electrical Engineers*, 10th ed., New York: McGraw-Hill, 1968, pp. 3–20.
4. E. U. Condon and H. Odishaw, *Handbook of Physics*, 2nd ed., New York: McGraw-Hill, 1967, pp. 4–57.
5. S. Seely, *Electronic Circuits*, New York: Holt, Rinehart & Winston, 1968, pp. 659–671.
6. M. Stout, *Basic Electrical Measurements*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1960, pp. 506–507.
7. L. Page and N. Adams, Jr., *Principles of Electricity*, 3rd ed., Princeton, NJ: D. Van Nostrand, 1958, pp. 64–68.
8. *ABC's of Probes*, Tektronix, Inc., Beaverton, OR, 60W-6053-4.
9. *Active Probes: Their Unique Characteristics and Applications*, Tektronix, Inc., Beaverton, OR, 60W-6883.
10. *The Effect of Probe Input Capacitance on Measurement Accuracy*, Tektronix, Inc., Beaverton, OR, 60W-8910-0.
11. L. Cromwell et al., *Biomedical Instrumentation and Measurements*, Englewood Cliffs, NJ: Prentice-Hall, 1973, p. 56.
12. M. Schwartz, *Information Transmission, Modulation, and Noise*, New York: McGraw-Hill, 1959.
13. R. S. Simpson and R. C. Houts, *Fundamentals of Analog and Digital Communication Systems*, Boston: Allyn and Bacon, 1971.
14. R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise*, Boston: Houghton Mifflin, 1976.
15. C. D. Motchenbacher and F. C. Fitchen, *Low-Noise Electronic Design*, New York: Wiley, 1973.
16. H. W. Ott, *Noise Reduction Techniques in Electronic Systems*, 2nd ed., New York: Wiley, 1988.
17. *Low Level Measurements*, 4th ed., Cleveland, OH: Keithley Instruments, Inc., 1993.

MICHAEL F. GARD
Subsite Electronics

BALANCES

Balances are usually considered devices that weigh. Here, we will take a more general approach because even the ordinary beam balance in school laboratories uses one force balanced against another, so that static equilibrium of the device is maintained. More generally, then, a balance is a device that determines an unknown force by balancing a known or calibrated force. For example, lever-arm balances use a known mass at a known position to measure an unknown mass at a known location. Spring balances use the displacement of a spring with known elastic properties to measure unknown forces. A modified version of a spring balance is the cantilever-type, where an elastic arm supported at one end bends under load. Another related class of balances is the torsion balance, in which a thin torsion fiber suspends a beam that twists because of forces solely in a horizontal plane. The amount of twist per unit torque is determined by the elastic properties of the fiber. The principles governing the various balance types are discussed more thoroughly in the following subsections.

Many commercial balances available today have automated recording systems, so that force signals are converted either to electrical voltage or current. Specialized precision balances for measuring masses of minute quantities, investigating chemical reactions, or studying gaseous-phase processes have also benefited from automation. Automation allows continuous monitoring of experimental variables as well as measuring in environments such as vacuums. Instrumentation typical of automated balances is discussed in the next section.

The last section describes recent designs for microbalances and vacuum microbalances. Before the 1960s, a vacuum microbalance typically consisted of a lightweight beam, counterbalanced by an electromagnetic force. As mass is added to one end of the beam, feedback on the beam deflection maintains equilibrium, usually electromagnetically. This feedback signal, proportional to the mass increment, is read as the mass measurement. Although similar microbalances are still in use, microbalances based on quartz-crystal oscillators are more common today, where mass increments translate into resonance frequency shifts. Recent designs of microbalances are described later.

Lever-Arm, or Beam Balances

By far, these are the most common balances in academic laboratories. Beam balance weigh an unknown mass through cancellation of the gravitational torque caused by placing the mass a known distance from the fulcrum of an extended beam. Figure 1 depicts such a balance, in which the canceling torque is provided by a known mass placed a known distance away from the fulcrum.

Because this balance has components found even in more modern automatic beam balances, its operation is discussed further here. The unknown mass is to be placed in pan P_1 , and the balance weights, or standards, are then loaded into pan P_2 until the needle indicator I swings equally about the center of scale S . The center of gravity of the system falls above knife-edge K_2 in equilibrium. Knife-edges are used to lower friction that may cause the swinging motion to stick in a position other than true equilibrium.

Because masses less than 10 g are difficult to handle (and it is undesirable to handle them because any small deposits change the mass), small weights are usually incorporated in a rider system, denoted in Fig. 1 by R . A small mass of 10 g is loaded by the rider onto beam B_2 at any position closer to the fulcrum than the pan. In the balance shown, beams B_1 and B_2 are equal length, so that loading a 10 g mass at the end of B_2 gives a reading of 10 g. Thus, loading the 10 g mass at a position 0.1 of the distance out from knife-edge K_2 gives a reading of 1 g. Multiple-beam analytic balances use multiple rider systems, so that the standard weights are never handled physically.

Beam balances also use an arrestment device, labeled A in Fig. 1. This device allows attaching the weighing pans rigidly to the supporting structure, so that the pans are released slowly near equilibrium, after loading is accomplished. Releasing A sets the balance into swinging motion, so that the position of the indicator may be monitored.

Two other common features are not shown explicitly in this figure: taring and damping. Taring is necessary because inevitable differences in the masses of pans P_1 and P_2 would cause a nonzero reading even with the balance unloaded.

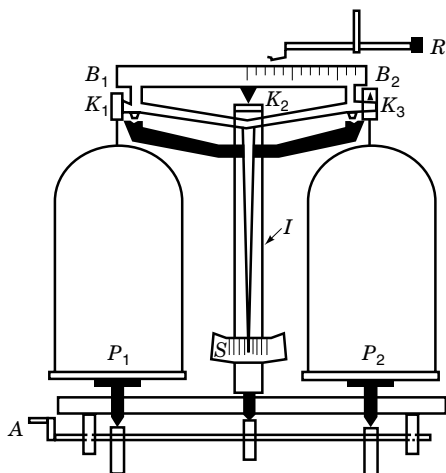


Figure 1. The beam balance. The unknown sample is placed in pan P_1 . Standard weights are added to pan P_2 to return the beam to the null position read from scale S . Masses smaller than 10 g are effectively added to beam B_2 through the rider R .

Taring is often accomplished with a counterweight mounted on one of the beams by an adjustable screw, so that the taring torque can be adjusted. Although a large amount of friction is undesirable in the knife-edge itself, damping the swinging motion of the scale indicator is desirable to reduce the time needed for a reliable measurement. Damping is often incorporated by placing small magnets near the end of one beam and using the eddy-current brake effect to damp the motion.

The sensitivity of the balance is fixed by the length of the indicator needle I , the unit measurement on the scale S , the size of the smallest mass placed in the rider system, and the increments on the rider scale. The balance is also more sensitive if the two beam arms are longer, but this increases the period of swinging caused by an increase in the moment of inertia of the balance. Practical beam balances, such as those manufactured by Mettler, measure from hundreds of grams to less than 0.0001 g.

Cantilever and Spring Balances

Both cantilever and spring balances use the elastic forces produced by a known material to balance an unknown weight. In a spring scale [Fig. 2(a)], loading an unknown mass stretches the spring beyond its unstretched length a distance Δx . The weight of the mass is related linearly to this displacement by Hooke's law:

$$F = k\Delta x$$

where k is the elastic spring constant. This relationship arises from the strain on the spring material, which is [Fig. 3(a)] related to a compressive or stretching force F applied perpendicularly to one end of a wire of cross-sectional area A via

$$\frac{F}{A} = Y \frac{\Delta x}{L}$$

where L is the length of the wire and Y is Young's modulus for the wire material. Spring balances are made of very stiff materials, such as steel, tungsten, or quartz, with elastic

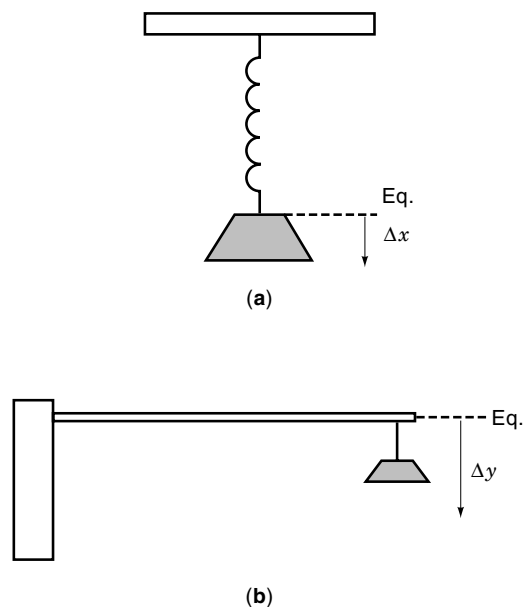


Figure 2. Spring (a) and cantilever (b) balances counter the weight with internal restoring forces. The displacement of the balance is proportional to the load over a wide range, where the constant of proportionality is called the effective spring constant k .

moduli close to 10^{11} N/m². The previous strain relationship indicates that one can make a more sensitive spring balance by increasing the length L of the spring; and by reducing the cross-sectional area A , one can make the deflection of the balance larger for a given applied force.

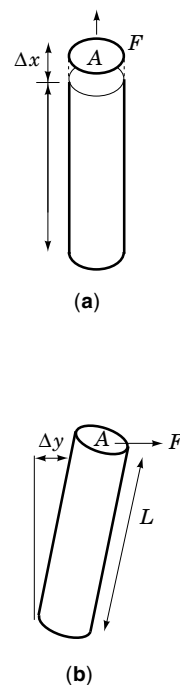


Figure 3. Two linear material deformations that may be exploited in balances. (a) Strain corresponds to an elongation in a fiber caused by a force applied perpendicularly to a cross section. (b) Stress corresponds to a sideways displacement of the fiber caused by a force applied parallel to a cross section.

In a cantilever balance, mass is loaded onto the end of a flexible rod or bar, mounted horizontally [Fig. 2(b)]. As in a spring scale, the deflection of the end of the rod is proportional to the weight added. This can be deduced by considering the shear force F acting on the fiber end [Fig. 3(b)] in the plane of the rod's cross-sectional area A :

$$\frac{F}{A} = G \frac{\Delta y}{L}$$

where Δy is the vertical deflection of the rod and G is the shear modulus of the material (typically the same order of magnitude as Y). As in spring balances, cantilever balances are made more sensitive by increasing the length of the rod and reducing the cross-sectional area.

Spring and cantilever balances are calibrated by loading standard masses onto the balance, while measuring the corresponding set of displacements. The data set allows determining k , the effective elastic constant of proportionality. When unknown masses are to be determined, they must be within the range of masses used for calibration, because both springs and cantilever beams are linear over only a limited range of loads.

Another method for calibrating spring and cantilever balances requires that they serve as high- Q (i.e., low-friction) mechanical oscillators. For a highly underdamped spring system, the resonant frequency of oscillation is given in Hertz by

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$

If frequency difference measurements of 1 ppm are reliably and repeatedly made on an oscillating mass, a mass change of 2 ppm can be detected.

Torsion Balances

Two types of balance are denoted by this title. In one type, similar to a beam balance, a taut wire is used as a fulcrum in place of a knife-edge [Fig. 4(a)]. Like other analytical balances, this balance measures a gravitational torque by a downward angular deflection $\Delta\theta$ of the beam.

A more commonly used torsion balance relies on the elastic properties of a fiber to determine small forces acting to rotate a beam. A beam of arm R is supported from its center of gravity by a wire or fiber w [Fig. 4(b)]. Forces apply a couple simultaneously to both ends of the beam, which rotates through an angle $\Delta\psi$. The total torque Γ and $\Delta\psi$ are related through $\Gamma = \kappa \Delta\psi$, where κ is an elastic constant determined by the shear modulus of the fiber G , the length of the fiber l , and the fiber radius r :

$$\kappa = \frac{r^4}{l} G$$

Thus, one can see that this form of torsion balance is most similar to spring and cantilever systems, in that known material properties of the fiber produce the balancing forces. Also, like these other balances, a torsion balance can be made more sensitive by increasing the fiber length and reducing the wire radius.

Torsion balances are unique in that they are designed to measure forces occurring in a horizontal plane. They are

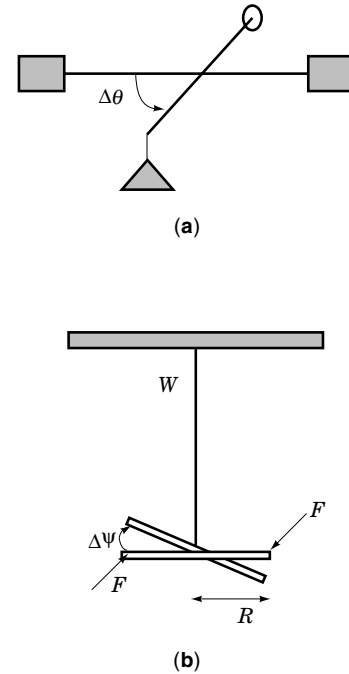


Figure 4. Two types of balances called torsion balances. (a) This is most similar to a beam balance because a weight added to one end creates a downward deflection of a beam. The torsion wire takes the place of a knife-edge. (b) This torsion balance is most similar to spring and cantilever balances. Material properties of the wire w balance the torque tending to twist the beam in a horizontal plane.

calibrated by using the frequency of oscillation of the beam about its unperturbed equilibrium position. (Practical torsion balances have periods that are several seconds to several minutes long.) The natural torsion frequency is given by

$$f = \frac{1}{2\pi} \sqrt{\frac{\kappa}{I}}$$

where I is the moment of inertia for the beam twisting in a horizontal plane about the fiber point of suspension. This I must be calculated or measured independently to determine κ .

Applications of the torsion balance have been numerous and include determining the Newtonian gravitation of constant G and the magnetic properties of superconductors. Such applications require sensitivities of $10^{-9} \text{ N} \cdot \text{m/rad}$, which are readily achieved in many laboratories.

INSTRUMENTATION

There are two main types of balance in terms of how measurements are acquired: null-detection and deflection type. In deflection-type balances, force or mass measurements are related to how far the balance moves from an equilibrium position. Cantilever and spring balances fall in this category. If such balances are to be automated, a position transducer is needed to convert a deflection into an electrical signal. In null-detection balances, the balance is to be maintained in equilibrium, and the force needed to restore the balance is read as the measurement. Beam balances are of this variety. Automated null-detection balances require a negative feedback loop.

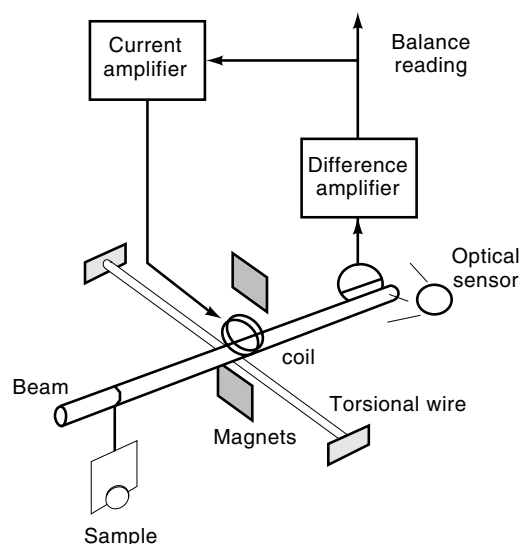


Figure 5. A null-detection microbalance. Position sensing is accomplished optically, and feedback is provided through the energized coil on the beam that attempts to align with the field created by the stationary magnets. Such balances are capable of detecting mass increments as small as 10^{-7} g.

A position transducer converts a displacement into an electrical signal, which is fed back to a force-producing mechanism that restores equilibrium. Figure 5 shows a Cahn beam microbalance, in which an electromagnetic coil energized by an optical position sensor tries to maintain alignment with the field created by static magnets. Because position sensing is central to both deflection and null-detection balances, various transducers are described below.

Optical Detection

One common method for sensing the position of a beam balance uses a light source and split photodiode arrangement [Fig. 6(a)]. The photodiode outputs a voltage signal proportional to the difference in source light intensity S reaching its two separate halves. The beam balance is set so that, at equilibrium, it covers both halves of the split photodiode symmetrically. As the beam moves upward, the lower half of the detector is more illuminated than the upper half, and the photodiode output changes accordingly. Many balance position-detecting systems use a similar split-transducer arrangement. Relatively simple, inexpensive circuits can be developed, such as the one shown, that output a voltage proportional to the difference between the two detector halves. Hence, the transducer output is null at equilibrium and produces a bipolar signal dependent on the direction of displacement.

Reflection is often used with balances that create angular deflections, such as cantilevers, beam balances, and torsion balances. In such systems, a well-collimated intense light source, such as a laser, is reflected from a mirror attached to the pivoting body or movement arm [Fig. 6(b)]. As the balance arm moves through an angle θ , the beam is deflected through 2θ . As the balance pivots, the light beam traverses an arc, the length of which is determined by how far the detector is placed from the balance. The length of this optical lever arm determines the sensitivity of the position transducer.

Two types of sensors are available for measuring the deflected light beam: linear charge-coupled device (CCD) arrays

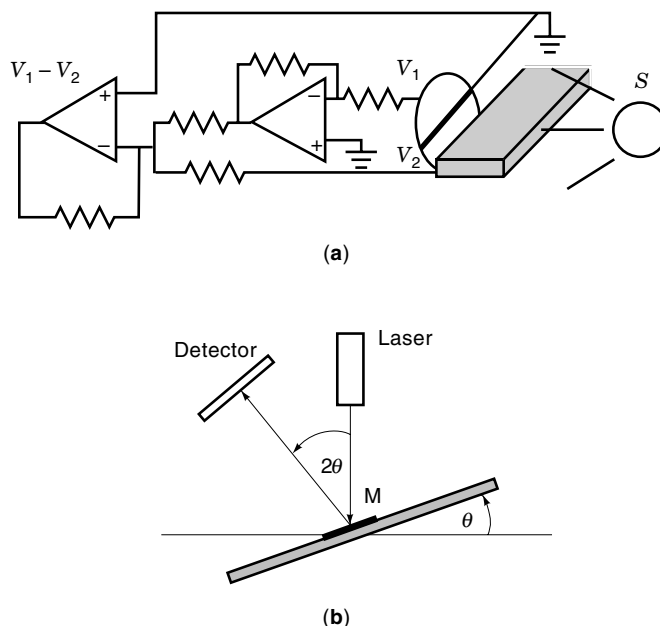


Figure 6. Optical position sensing. (a) Split photodiode. If all resistors shown have equal value, the output of the second operational amplifier is proportional to the difference in the detector outputs. (b) Reflection, or optical lever-arm device.

and position-sensing devices (PSDs). A CCD array consists of a set of photosensitive elements, typically a few microns in size. Arrays are commercially available with thousands of so-called pixels. The device produces a digital signal that provides information about the light intensity recorded as a function of discrete position along the one-dimensional array. This method is limited in resolution by the number of photosensitive elements in the array and also requires some computer processing to determine the centroid of the laser light spot. Thus, resolution in such devices is gained at the expense of speed. The second type of detector, the PSD, consists of a single photodiode several cm in length. A resistor network built into the device behaves as a voltage divider, so that the device output is related to the position of the laser spot on the photodiode active area. Unlike the CCD array, the signal is continuous, and no centroid determination is required. However, a specialized analog circuit is required to make the voltage output proportional to the beam-spot position on the device. PSDs are available that measure $0.1 \mu\text{m}$ of linear motion.

Capacitive Sensing

Capacitors store electric charge and energy. A capacitor consists of two conductors separated by an air gap or dielectric. How much charge the two conductors store is determined by the shape and size of the conductors and by the geometry of the gap separating them. These features can be used to convert motion into an electrical signal.

The simplest configuration is the parallel-plate capacitor, the capacitance of which is determined by the area of the plates A and gap size d :

$$C = \frac{\epsilon_0 A}{d}$$

where ϵ_0 is the permittivity of free space (8.85×10^{-12} F/m). The small value of ϵ_0 means that position transducers using

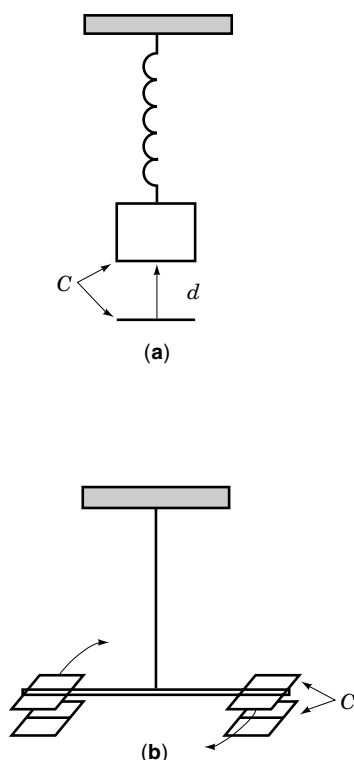


Figure 7. Two types of capacitive position sensors. (a) The capacitance decreases as the separation between the two plates increases with balance motion. (b) The area of overlap between the capacitor plate on the beam and the split capacitor sensors changes as the balance rotates.

capacitors must be very sensitive to minute changes in C of the order of nanofarads or picofarads.

Two types of capacitive position sensors can be developed, depending on whether the gap changes [Fig. 7(a)] or the area of the plates changes [Fig. 7(b)] as the balance moves. [Figure 7(b)] indicates that although the movable capacitor plate is single, the stationary plate consists of a split conductor. For the same reason that split photodiodes are used in optical sensors, split capacitor plates allow determining the direction of balance movement.

One also needs a circuit designed to detect changes in C . This is often done by placing the variable position-sensing capacitor in an ac-excited bridge [Fig. 8(a)]. The potential difference between points A and B in the figure remains null, as long as the four capacitors in the bridge have equal values. As the variable C s change because of motion of the balance, the amplitude and phase of the voltage difference ($V_A - V_B$) changes correspondingly. Another common arrangement is to place the capacitive position sensor in a resonant LC circuit [Fig. 8(b)], excited by an ac source driving the circuit near the resonant frequency of the circuit, given by $1/(2\pi\sqrt{LC})$. As the capacitance changes, the amplitude and phase of the voltage drop across C also change.

Strain Gauges

Strain gauges are devices that convert physical elongation or compression into an electrical signal. They are readily incorporated into deflection-type balances, such as the spring scale and beam balance. A strain gauge consists of four metal-film

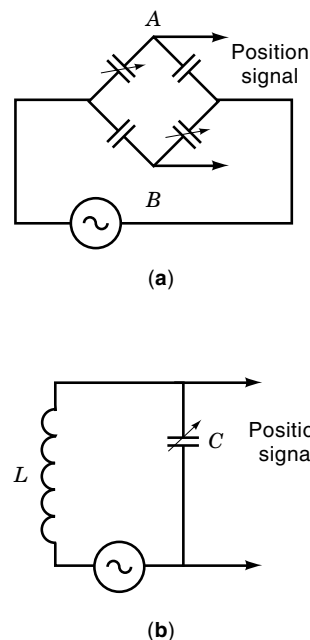


Figure 8. Two circuits used to change capacitance variations into an electrical signal. (a) Ac capacitance bridge, in which a moving capacitor plate deflects across a split capacitor plate (two variable C s). (b) Resonant LC circuit.

resistors placed in a Wheatstone bridge configuration. As the balance deflects, the resistors deform and destroy the bridge electrical balance. It is difficult to find strain gauges that measure small displacements reliably, so that these devices are more appropriately used in balances designed to measure large loads, hundreds of grams to thousands of kilograms.

Inductive Sensing

A common technique for measuring deflection in balances is a device known as a linear variable differential transformer (LVDT). As all transformers contain a pair of coils coupled with an iron core, so do LVDTs. However, the primary coil, the one to be excited directly with an ac source, is center-tapped. This allows two secondary coils to share a common ground at the center of the transducer (Fig. 9). This configuration is reminiscent of the split photodiode sensor

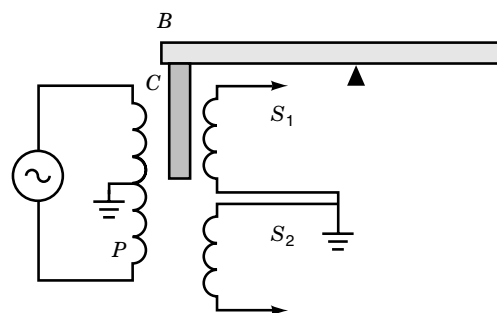


Figure 9. Linear variable differential transformer (LVDT), an inductive position sensor. The center-tapped primary coil P is flux-coupled to the secondary windings through an iron core C . As the beam B deflects up or down, either S_1 or S_2 becomes more strongly coupled to the primary.

described previously, so it is evident the LVDT is capable of determining the direction of balance motion. The iron core is attached to the balance beam and is centered at equilibrium. As the beam displaces upward, coil S_1 in the secondary is more strongly coupled to the primary than coil S_2 . The voltage signals from the two secondary windings are processed by an electronic circuit that produces a signal proportional to their difference.

MICROBALANCES AND VACUUM MICROBALANCES

A microbalance is a device that either determines a small mass (μg range) or force. Applications are numerous, including measurement of adsorbed chemicals on surfaces, film thicknesses in vacuum deposition systems, minute pressure changes, dehydration and hygroscopic processes, pyrolytic reactions, corrosion, etc. Measurement in vacuo is often necessary, to remove buoyancy or chemical reactions that cannot be controlled. Under these conditions the balance is termed a vacuum microbalance. The main difference between a vacuum microbalance and other such devices is that the balance parts must be capable of being heated to high temperature to circumvent outgassing and appropriate mechanisms for loading and reading must be placed within the vacuum chamber.

As was mentioned in the previous section, most microbalances before 1960 were similar in design to standard automated balances. Differences were required in the small, delicate parts, from which beams, springs, and cantilevers were machined, and the low levels of signals produced by the position transducers. In this section, relatively new forms of the microbalance are presented. Two of these, the quartz crystal microbalance (QCM, described later) and the surface acoustic wave (SAW, described later) microbalance, are based on the piezoelectric nature of quartz. When a quartz crystal is placed in a sufficiently strong electric field, it undergoes a characteristic deformation. Oscillating quartz crystals can be made sensitive to small material deposits on their surfaces.

Atomic-force microscopy is also an unfolding field, in which a miniature cantilever beam records forces on a microscopic scale. If such a device is used to create images of materials, it is a microscope. However, if such a device is used to measure forces absolutely, it is a microbalance.

Atomic Force Balances

An atomic force microscope (AFM) has at its heart a micromachined cantilever beam, which responds to microscopic forces as it is dragged over a surface or through a fluid. Although cantilever fibers and leaves have also been used to measure small forces, currently V-shaped structures of Si_3N_4 are popular because they resist lateral motion and the tendency to twist. These beams are only tens of micrometers thick and hundreds of micrometers long.

Regardless of the shape of the cantilever beam, a remaining challenge for such small devices is calibrating them to determine their effective spring constants. One technique involves calculating the k from the size of the cantilever and its bulk properties (namely, the Young's modulus of the material from which it is made). However, this calculation is unreliable because the thin films may not have the same Young's modulus as large blocks of material. Another problem concerns the fact that many cantilever beams have gold surfaces sputtered on them to make them easier to track via optical reflection. The

gold film also changes the Young's modulus in a way that is difficult to determine.

Another method of calibrating these cantilever beams involves loading the end of the beam with mass and measuring the beam displacement or the shift in the resonant frequency of the beam set into oscillation. Both methods are common to calibrating any spring-type balance. However, for atomic force balances this is a destructive method, as glue is often necessary to attach a μg mass to the beam. It has also been found that the effective spring constant measured by mass loading and frequency shifting is affected by the attachment location. New methods of calibrating these cantilevers to determine microscopic forces will continue to develop, as AFMs become more widely used for absolute force measurements.

Quartz Crystal Microbalances

Quartz crystal microbalances (QCMs) are extensively used in materials research and in chemistry laboratories. Their longest application has been for film-thickness monitoring during vacuum deposition. More recently, they have been adapted for investigating liquid solutions, in which chemical reactions at the liquid-quartz interface leave deposits only one atomic layer thick.

QCMs are an attempt to use frequency measurement as a mass indicator. A thin (several hundred micrometer) wafer is cut from a quartz crystal, and typically gold electrodes are deposited on both faces of the wafer [Fig. 10(a)]. Shear-mode oscillations are set up in the wafer when an oscillating electric field is applied to the electrodes [Fig. 10(b)]. If the thickness of the wafer is one-half the wavelength of the shear-mode fundamental, the antinodes of the oscillation occur at the faces of the wafer. In this case, very little energy loss occurs at the wafer faces, and the acoustic wave essentially becomes trapped inside the wafer. Constructed in such a way, a quartz crystal oscillator

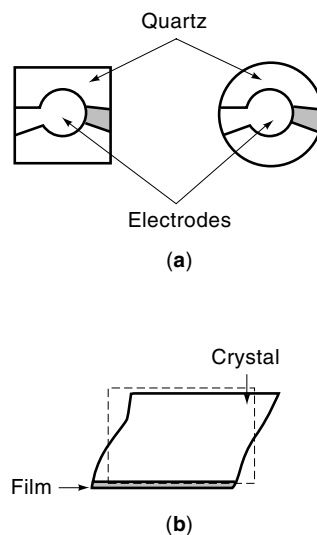


Figure 10. Operation of the quartz crystal microbalance. (a) Quartz wafers showing electrodes. The dark gray region indicates the orientation of the electrode on the opposite side of the wafer. (b) Shear oscillation in the bulk of the quartz crystal. As long as the film deposit is thin compared with the crystal width, the shear wave is damped very little by the film.

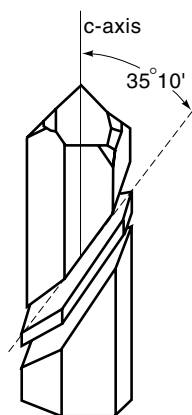


Figure 11. Orientation of an AT-cut wafer in a quartz crystal.

can have a very high Q , and thus a narrow resonance frequency of oscillation.

When quartz wafers are prepared, care is taken to minimize the dependence of this natural frequency of oscillation on temperature. Such a dependence causes errors in thickness monitors because mass deposition is usually measured as a change in the resonant frequency. QCMs are manufactured with an AT cut, a particular crystallographic orientation known to minimize temperature dependence of the resonant frequency at room temperature (see Fig. 11).

The resonant frequency of an AT-cut quartz crystal is given as $f = A/d$, where d is the thickness of the wafer and A is a constant, 1.67 MHz-mm. The initial thickness of the wafer is cut to give f near 10 MHz. Smaller cuts produce a higher frequency but reduce both the sensitivity of the crystal and the thickness of foreign material that can be practically deposited.

Mass deposits Δm on the QCM are monitored according to the Sauerbrey relationship:

$$\Delta f = -2.3 \times 10^6 \frac{\text{cm}^2}{\text{Hz} \cdot \text{g}} f_0^2 \frac{\Delta m}{A}$$

where A is the area over which the film has been deposited, and Δf is a measured shift in the quartz oscillation frequency. This relation shows explicitly that the QCM becomes more sensitive to deposits as the natural frequency f_0 of the crystal increases.

Several well-studied problems occur in using QCMs, both in vacuum deposition systems and in electrochemical environments. One difficulty arises from the sensitivity of the QCM to varying environmental conditions other than a change in thickness, especially to temperature changes and variations in liquid viscosity and density. The main difficulty in using QCMs as film-deposition monitors in vacuum is the fact that the evaporated substance conveys thermal energy to the crystal, thereby changing its temperature. Even if it is an AT-cut crystal, the crystal has a frequency dependent to a slight degree on temperature. The way around this problem is to place a shutter close to the crystal surface, so that the film does not cover the entire crystal surface. Also, the shutter is not opened immediately, so that the crystal has time to reach thermal equilibrium before thickness is monitored.

Another well-documented problem is that the crystal's sensitivity to thickness changes is not uniform across the wafer.

The sensitivity is highest in the center and lowest on the periphery. The profile may be made more uniform by polishing the crystal so that it exhibits a convex or plano-convex shape. However, most researchers take care to confine reactions and deposits to the area of the wafer covered by the electrode.

One final challenge to using QCMs concerns the approximation that the nature of the shear wave is not affected by the film deposit but that only the speed of the wave changes. In reality, the film changes the acoustic impedance at the surface, which in turn affects energy loss of the wave.

As a foreign substance is deposited on one face of the crystal, the thickness of the wafer increases, and the resonant frequency drops. As long as the foreign film is thin compared with the wafer thickness, the elastic properties of the foreign substance do not affect the behavior of the shear wave. Thus only the areal mass density affects the frequency. This approximation has been experimentally verified, as long as the foreign film remains thinner than a few percent of the wafer thickness. It is advisable to clean the wafer after a certain amount of deposit has accumulated, although a technique called Z-match, described in the next paragraph, allows larger deposits.

A more careful one-dimensional acoustic analysis of the shear wave propagating through the crystal recognizes the impedance mismatch at the boundary between the film and the quartz. In this instance, a transcendental equation relates the frequency shift of the oscillator to the mass deposited Δm on the crystal:

$$\Delta m/m_0 = -\frac{z_f f_0}{z_Q \pi f} \arctan \left(\frac{z_Q}{z_f} \tan \frac{\pi f}{f_0} \right)$$

where z_f and z_Q represent the acoustic impedances of the film and quartz materials, respectively, and m_0 is the initial mass of the crystal. Third-generation thickness monitors use this relationship, which is valid up to loads of 70% of the quartz mass. A disadvantage to using this Z-match technique is that the ratio of the two impedances must be determined by an independent measurement and entered in the deposition-monitor memory for each type of film deposited. Because this ratio may not accurately be known, the Z-match feature provided in such monitors may not necessarily be accurate up to high loads. Monitors using the Z-match technique must also keep track of which films have been preloaded before a particular deposition and the impedance of these preloaded samples.

For QCMs submerged in liquids, it has been found that the shear wave loses energy exponentially as it enters the liquid solution. The energy loss adds a term to the Sauerbrey equation, so that frequency shifts Δf do not depend solely on the mass added:

$$\Delta f = -f_0^{3/2} \sqrt{\frac{\eta_s \rho_s}{\pi \mu_Q \rho_Q}}$$

where η_s and ρ_s are the viscosity and density of the liquid solution, respectively, μ_Q is the shear modulus of quartz (2.947×10^{11} g/cm-s²), and ρ_Q is the density of quartz (2.648 g/cm³). A second crystal in contact with the fluid may compensate for frequency shifts due to viscosity and density of the solution.

As described above, the thickness of a deposited film can be measured as a change in the resonant frequency of a quartz crystal. Figure 12 shows a block diagram of a common method to extract this frequency change. The quartz crystal forms part

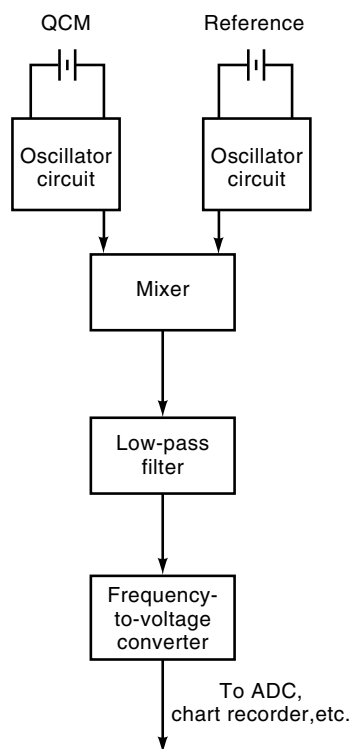


Figure 12. Typical block diagram showing the configuration of a QCM in a film-deposition monitor.

of a positive feedback oscillator circuit. The output of this circuit is mixed with a reference oscillator not subjected to the film deposition and thus having a constant resonant frequency. The mixer creates a signal in the audio frequency range, which may be counted, or may be input into a frequency-to-voltage converter. The f -to- V converter has the advantage that the output signal can be directly fed into a recorder or computer. Commercial QCM systems are available for less than \$10,000. Many detect mass increments less than 1 ng.

As an alternative to placing the QCM in an oscillator circuit, the QCM may be excited with a signal generator. When the generator is shut off, the crystal exhibits a decaying oscillating behavior characteristic of a high- Q mechanical oscillator. The resonant frequency can be determined from the oscillations within the decaying envelope.

The need for measuring z -values for better Z -matching has introduced more techniques in instrumentation. One technique involves placing the quartz/film resonator to be studied in a circuit where multiple resonances of the device are measured by using a vector voltmeter. (This is a specialized instrument that measures both the amplitude and phase of an oscillating electrical signal.)

Surface Acoustic-Wave Devices

At first glance, a SAW device appears related to the QCM [Fig. 13(a)]. Electrodes are deposited on a piezoelectric quartz wafer, and the crystal is set in oscillation electrically. However, the SAW device is excited only on one surface [Fig. 13(b)], and thus the excited waves occur at the surface, unlike the shear waves propagating in the bulk of a QCM.

However, just as in a QCM, changes in frequency of oscillation occur if mass is deposited on the surface. The frequency

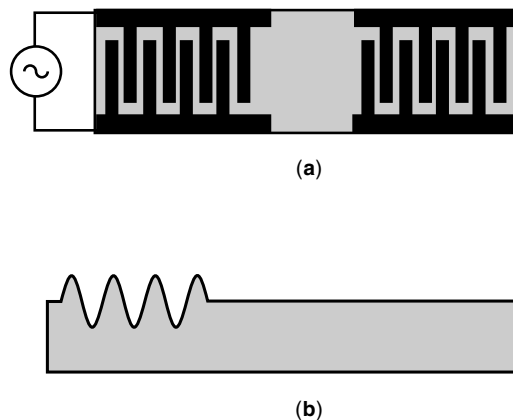


Figure 13. Operation of a surface acoustic wave (SAW) device. (a) Two sets of interdigital transducers are deposited on a quartz substrate. The configuration shown corresponds to an RF delay line. (b) Surface acoustic wave energized by the oscillating voltage source shown in (a).

shift created by a mass Δm deposited on a SAW crystal of area A is found from

$$\Delta f = -1.26 \times 10^6 \frac{\text{cm}^2}{\text{Hz} \cdot \text{g}} f_0^2 \frac{\Delta m}{A}$$

where f_0 is the unloaded resonant frequency. The SAW device has two advantages over the QCM. First of all, the SAW is uniformly sensitive to deposits over its entire surface. Secondly, there is no practical limit to the f_0 of the surface waves because the resonant frequency is not determined by crystal thickness. In an SAW device, f_0 is determined by the spacing in the interdigital transducer (IDT) electrodes, which must be smaller and closer together to produce a higher frequency. Thus, the SAW has greater inherent sensitivity than the QCM.

SAWs were initially constructed to serve as delay lines in radio-frequency (RF) circuits, which can be seen in the SAW configuration in [Fig. 13(a)]. One set of electrodes is excited with an RF generator, and the other set receives the signal a short time later. Because the wave speed of surface acoustic waves is far less than the speed of light waves, a short SAW device creates a lengthy delay in the RF signal without appreciable energy loss. SAW devices were soon found impractical for RF applications, however, because they are extremely sensitive to ambient conditions and hence display varying frequency.

SAW devices are incorporated into film-deposition microbalances similarly to QCMs. Typically two separate SAW devices are placed on the same quartz substrate and are excited by a matched pair of RF driver circuits. One of the SAW devices is exposed to the process under study, and the other is a reference. Both signal and reference are mixed, so that the frequency difference is measured in the kilohertz range.

BIBLIOGRAPHY

Balance Types

1. W. Sprinkl, E. Lavemann, and H. Ries, An easy-to-build balance for automatic measurement, *J. Phys. E: Sci. Instrum.*, **20**: 1452–1454, 1987.

2. G. L. Miller et al., A rocking beam electrostatic balance for the measurement of small forces, *Rev. Sci. Instrum.*, **62**: 705–709, 1992.
3. R. M. Schoonover, *Mass Comparator for In-situ Calibration of Large Mass Standards*, *J. Res. Nat. Bur. Stands.*, **90**: 289–294, 1985.
4. J. T. Buontempo and F. A. Novak, An inexpensive Wilhelmy balance for the study of Langmuir monolayers, *Rev. Sci. Instrum.*, **63**: 5707–5713, 1992.
5. G. T. Gillies and R. C. Ritter, Torsion balances and torsion pendulums, *Rev. Sci. Instrum.*, **64**: 283–309, 1993.
6. S. Gordon and C. Campbell, Automatic and Recording Balances, *Anal. Chem.*, **32**: 271R–289R, 1960.
- acoustic wave resonator mass microbalance, *Rev. Sci. Instrum.*, **62**: 1624–1629, 1991.
25. W. D. Bowers and R. L. Chuan, Surface acoustic-wave piezoelectric crystal aerosol mass microbalance, *Rev. Sci. Instrum.*, **60**: 1297–1302, 1989.

L. I. WINKLER
Appalachian State University

Non-Quartz Crystal Microbalances

7. L. Bruschi et al., Novel needle microbalance for adsorption studies, *Rev. Sci. Instrum.*, **68**: 4536–4541, 1997.
8. S. J. Suh and H. J. Eun, Small-mass measurement by optical glass-fibre elastic cantilever, *Meas. Sci. Technol.*, **1**: 556–560, 1990.
9. L. Bruschi and G. Torzo, A vibrating fibre microbalance for measuring layer-by-layer adsorption on graphite, *Rev. Sci. Instrum.*, **62**: 2772–2777, 1991.
10. N. Uetake, T. Asano, and K. Suzuki, Measurement of vaporized atom flux and velocity in a vacuum using a microbalance, *Rev. Sci. Instrum.*, **62**: 1942–1946, 1991.
11. D. M. Astill, P. L. Hall, and J. D. C. McConnell, An automated vacuum microbalance for measurement of adsorption isotherms, *J. Phys. E: Sci. Instrum.*, **20**: 19–21, 1987.
12. J. E. Sader et al., Method for the calibration of atomic force microscope cantilevers, *Rev. Sci. Instrum.*, **66**: 3789–3798, 1995.
13. A. Torii et al., A method for determining the spring constant of cantilevers for atomic force microscopy, *Meas. Sci. Technol.*, **7**: 179–184, 1996.
14. L. I. Maissel and R. Glang (eds.), *Handbook of Thin Film Technology*, New York: McGraw-Hill, 1970, pp. 1–103 ff.

Quartz Crystal Microbalances

15. D. S. Ballantine and H. Wohltjen, Surface acoustic wave devices for chemical analysis, *Anal. Chem.*, **61**: 704A–712A, 1989.
16. M. R. Deakin and D. A. Buttry, Electrochemical applications of QCM, *Anal. Chem.*, **61**: 1147A–1154A, 1989.
17. A. Wajid, Improving the accuracy of a quartz crystal microbalance with automatic determination of acoustic impedance ratio, *Rev. Sci. Instrum.*, **62**: 2026–2033, 1991.
18. E. Benes, Improved quartz crystal microbalance technique, *J. Appl. Phys.*, **56**: 608–626, 1984.
19. P. J. Cumpson and M. P. Seah, The quartz crystal microbalance; radial/polar dependence of mass sensitivity both on and off the electrodes, *Meas. Sci. Technol.*, **1**: 544–555, 1990.
20. M. D. Ward and E. J. Delawski, Radial mass sensitivity of the QCM in liquid media, *Anal. Chem.*, **63**: 886–890, 1991.
21. Product Review: Quartz crystal microbalances, *Anal. Chem.*, **68**: 625A–628A, 1996.
22. M. Rodahl and B. Kasemo, A simple setup to simultaneously measure the resonant frequency and absolute dissipation factor of a quartz crystal microbalance, *Rev. Sci. Instrum.*, **67**: 3238–3930, 1996.
23. S. Bruckenstein et al., Dual quartz crystal oscillator circuit. Minimizing effects due to liquid viscosity, density, and temperature, *Anal. Chem.*, **66**: 1847–1852, 1994.
24. W. D. Bowers, R. L. Chuan, and T. M. Duong, A 200 MHz surface

BRIDGE INSTRUMENTS

Bridge instruments have a broad field of applications in metrology and industrial laboratories, not only for the comparison or the determination of electrical quantities but also for the measurement of nonelectrical quantities by electrical methods. The aim of this article is to combine the need to provide an understanding of the principles used in some modern dc and ac bridge instruments without forgetting those involved in the most classical ones. This article is divided in three parts: bridge networks, dc bridge instruments, and ac bridge instruments.

BRIDGE NETWORKS

Elementary Dipoles

Bridge networks are realized by the connection of elementary electrical elements generally fitted with two terminals and called dipoles. An electrical dipole is an electrical element which could be externally connected by means of its two terminals. The current entering by one terminal is identical to the one leaving the other. One electrical dipole can supply or transform electrical energy. Two main categories of dipoles can be distinguished:

- Active dipoles like voltage sources and current sources
- Passive dipoles like resistors, capacitors and inductors

If one dipole is linear, its behavior is governed by a mathematically well-defined linear correspondence between the applied voltage and the flowing current. Then, we can write for each dipole:

- A relationship voltage-current $v = f(i)$
- A relationship current-voltage $i = g(v)$

Sources. The sources are dipoles which are able to modify the electrical state of networks. There are two types of ideal sources: voltage sources and current sources. The ideal voltage source is a hypothetical dipole which maintains a constant voltage at its terminals, independent of the current flowing through it:

$$\frac{dv}{di} \equiv 0 \quad (1)$$

The ideal current source is a dipole for which the current is independent of the voltage at its terminals:

$$\frac{di}{dv} \equiv 0 \quad (2)$$

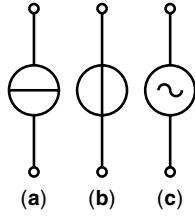


Figure 1. Representation of the used active dipoles: (a) dc current source. (b) dc voltage source. (c) ac generator.

The symbols used for the dc sources are those indicated in Recommendation No. 117 of the International Electrotechnical Commission (IEC) and represented in Fig. 1(a) and 1(b). To distinguish the ac sources and due to the fact that it is very common in the literature, the symbol used has been abstracted from IEEE Standard 315-1975 and is shown in Fig. 1(c).

Passive Dipoles

Resistor. A resistor of resistance R and of conductance G is a dipole which creates a voltage drop $v(t)$ at its terminals, when flown through by an instantaneous current $i(t)$:

$$v(t) = Ri(t) \quad (3)$$

The inverse relationship current-voltage is

$$i(t) = Gv(t) \quad (4)$$

Capacitor. A perfect capacitor of capacitance C is a dipole which possesses an electrical charge $q(t)$ when the instantaneous voltage at its terminals is $v(t)$:

$$q(t) = Cv(t) \quad (5)$$

The relation of current-voltage can be written

$$i(t) = \frac{dq(t)}{dt} = C \frac{dv(t)}{dt} \quad (6)$$

and the relationship of voltage-current:

$$v(t) = \frac{1}{C} \int_0^t i(t) dt + v(0) \quad (7)$$

where $v(0)$ is the value of v at $t = 0$.

Inductor. A perfect inductor of inductance L is a dipole which embraces a magnetic flux $\Phi(t)$ when flown through by an instantaneous current $i(t)$:

$$\Phi(t) = Li(t) \quad (8)$$

The instantaneous voltage at the terminals of this inductor is

$$v(t) = \frac{d\Phi(t)}{dt} = L \frac{di(t)}{dt} \quad (9)$$

The relationship of current-voltage is

$$i(t) = \frac{1}{L} \int_0^t v(t) dt + i(0) \quad (10)$$

where $i(0)$ is the value of i at $t = 0$.

Passive Dipoles in ac

If we apply an ac voltage to each of the previous passive dipoles, an ac current $i(t) = I_M \cos \omega t$ will flow through them. We will hereafter successively examine each dipole.

Resistor. The voltage difference at the terminals of the resistor is

$$v_R = RI_M \cos \omega t = V_M \cos \omega t \quad (11)$$

The impedance Z and the admittance Y are real numbers:

$$Z = R \quad \text{and} \quad Y = \frac{1}{R} \quad (12)$$

Inductor. The voltage difference at the terminals of the inductor is

$$v_L = L \frac{di}{dt} = -L\omega I_M \sin \omega t = L\omega I_M \cos \left(\omega t + \frac{\pi}{2} \right) \quad (13)$$

The impedance Z and the admittance Y are imaginary numbers:

$$Z = jL\omega \quad \text{and} \quad Y = \frac{1}{j\omega L} \quad (14)$$

Capacitor. The voltage difference at the terminals of the inductor is

$$v_C = \frac{1}{C} \int_0^t i dt + v_C(0) \quad (15)$$

if $v_C(0) = 0$ at $t = 0$
Then, we have

$$v_C = \frac{I_M}{C\omega} \sin \omega t = \frac{I_M}{C\omega} \cos \left(\omega t - \frac{\pi}{2} \right) \quad (16)$$

The impedance Z and admittance Y are also imaginary numbers:

$$Z = \frac{1}{j\omega C} \quad \text{and} \quad Y = j\omega C \quad (17)$$

Symbols. In this document, the symbols used to represent the different passive dipoles are coming from the same IEEE Standard or IEC Recommendation as already indicated previously and are shown in Fig. 2(a–e).

Lumped Networks in ac

A lumped network is realized by the connection of separate passive dipoles like a resistor, an inductor, and a capacitor. This

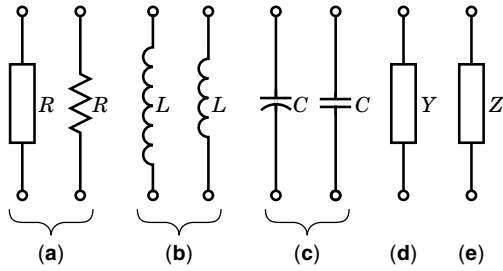


Figure 2. Representation of the used passive dipoles: (a) Resistor. (b) Inductor. (c) Capacitor. (d) Admittance. (e) Impedance.

concept is very familiar in ac bridges and is used for representation of imperfect dipoles. An imperfect inductor is represented in two ways in Fig. 3(a) and (b). The conversion relationships between the two representations are

$$L_s = L_p(1 + Q^{-2}) \quad \text{and} \quad R_s = \frac{R_p}{1 + Q^2} \quad (18)$$

where

$$Q = \frac{R_p}{\omega L_p} = \frac{\omega L_s}{R_s} = \frac{1}{D} = \frac{1}{\tan \delta} \quad (19)$$

Similarly, for an imperfect capacitor the relationships are

$$C_s = C_p(1 + D^2) \quad \text{and} \quad R_s = \frac{R_p}{1 + D^{-2}} \quad (20)$$

where

$$D = \frac{1}{\omega C_p R_p} = \omega C_s R_s = \tan \delta = \frac{1}{Q} \quad (21)$$

Generally, the series representation is preferred for inductors, and the parallel one is preferred for capacitors.

Kirchhoff's Laws

Two laws due to Kirchhoff form the basis of the electrical network theory.

Kirchhoff's Current Law. The current law is applicable to any node and may be expressed as

$$\sum i(t) = 0 \quad (22)$$

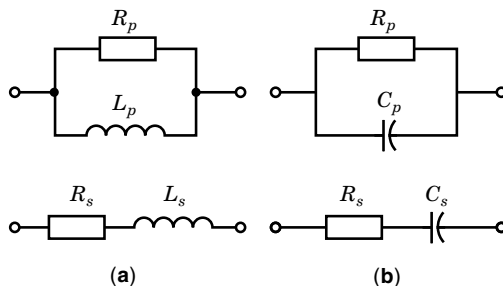


Figure 3. Lumped networks in ac: (a) Inductive (b) Capacitive.

This law may be regarded as a principle of conservation of charge.

Kirchhoff's Voltage Law. The voltage law is applicable to any loop, is made up of a set of electrical dipoles connected end to end in order to form a closed circuit, and may be expressed, in the case of purely resistive dipoles, as:

$$\sum e(t) = \sum Ri(t) \quad (23)$$

That is a generalization of Ohm's law.

General Principles Applicable to Linear Electrical Networks

An electrical network consists of a set of electrical dipoles connected together. If an electrical network is externally accessible by only two terminals, it constitutes by itself a dipole. An electrical network which is accessible from the external by means of four terminals, is considered like a quadripole. The general case is a multipole. We will only examine the principles which are useful for the understanding of the following sections. Indeed, complementary explanations about these principles can also be found in this encyclopedia or in other specialized books.

It is also assumed that the examined networks are linear. This means, in particular, that the current in any branch of a network caused by a source located in the same or any other branch is proportional to the electromotive force of that source. A branch is made up of one electrical dipole or of more electrical dipoles connected in series and flown through by the same current.

Superposition Principle (Helmholtz's Theorem). The current in any branch of a network is the sum of the currents which would be induced in the branch by each source of the network when all other sources are either short-circuited (voltage source) or open (current source).

Duality Principle. The current-voltage relationships for a given network can be deduced from the voltage-current relationships of the reciprocal network by application of the following bijections:

$$v \Leftrightarrow i$$

$$R \Leftrightarrow G$$

$$L \Leftrightarrow C$$

$$\phi \Leftrightarrow q$$

$$\text{series} \Leftrightarrow \text{parallel}$$

$$\text{loop} \Leftrightarrow \text{node}$$

$$\frac{dv}{di} \equiv 0 \Leftrightarrow \frac{di}{dv} \equiv 0$$

$$v \equiv 0 \Leftrightarrow i \equiv 0$$

(24)

This principle is useful when it is requested to solve a problem where the solution of the reciprocal problem is already known. So, in an electrical bridge, by application of this principle, the source and the detector may be interchanged without altering the balance condition. However, this change also modifies the impedance matching between the source and the detector.

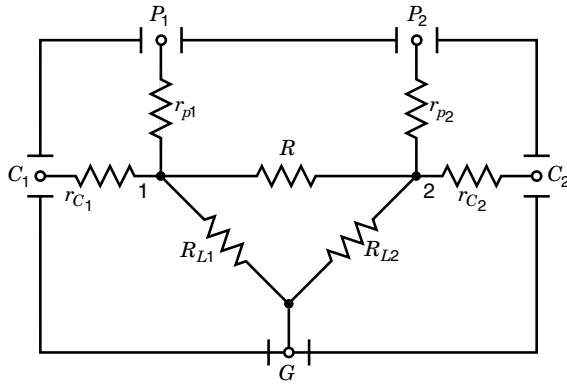


Figure 4. An equivalent circuit of a five-terminal resistor.

DC BRIDGE INSTRUMENTS

Definitions for a dc Resistor

The value of the combined standard measurement uncertainty, which accompanies the measurement result of a resistor, is often dependent of the characteristics of the device, itself. The factors contributing to this uncertainty are mainly

- Imperfect design of the resistor
- Connection terminals and the insulation materials
- Influence of the measurement parameters such as the temperature, the relative humidity, the pressure, and the measuring current.

At present, the most completely defined resistor is the one called five-terminal. The schematic diagram of a five-terminal resistor is shown in Fig. 4:

- P_1 and P_2 are the potential terminals
- C_1 and C_2 are the current terminals
- G is the guard terminal and is connected to the metal container surrounding the resistive element
- R is the resistor to be measured. It must respond to the accepted definition of a four-terminal resistor, which is the ratio of

$$R = \frac{U_{12}}{I_{12}} \quad (25)$$

where U_{12} is the open circuit voltage between nodes 1 and 2, and I_{12} is the current flowing through R .

- r_{c1} , r_{c2} , r_{p1} , and r_{p2} represent the lead and terminal resistances, respectively, at the terminals C_1 , C_2 , P_1 , and P_2 .
- R_{L1} and R_{L2} represent the leakage resistances from the same terminals C_1 , C_2 , P_1 , and P_2 .

This type of five-terminal resistor is now commercially available in the range from 1Ω to $10^9 \Omega$.

The four-terminal configuration is the same as in Fig. 4 without the Guard terminal G . The measured resistance differs

from the four terminal definition and is given by:

$$R_{\text{meas}} \cong R \left(1 - \frac{R}{R_{L1} + R_{L2}} \right) \quad (26)$$

From this formula, it can be seen that this design is valid for lower value dc resistors ($< 10^3 \Omega$). Although an old design, this is still used for the most stable dc standard resistors at the 1Ω level (1).

In the three-terminal configuration, there is no distinction made between the potential and the current terminals but the guard terminal is kept. The lead and terminal resistance r_{c1} and r_{c2} for this kind of resistor add to R . The measured resistance is now given by:

$$R_{\text{meas}} = R + r_{c1} + r_{c2} \quad (27)$$

Because r_{c1} and r_{c2} for this kind of resistor are below 1Ω , it is generally accepted that this design can be used for resistors above $10^6 \Omega$.

The two-terminal configuration is a simplified model of the three-terminal one. That was the first configuration used. It is still used in practice when it is compatible with the required measurement uncertainty.

Wheatstone Bridge

The bridge principle was first introduced by S. H. Christie in 1833, but was neglected until 1843, when Sir Charles Wheatstone applied this previously described arrangement of wires to the comparison of resistances (2–4).

In the original form of the Wheatstone bridge, shown in Fig. 5, X is the unknown resistor, A and B are equal resistors, and S is an adjustable resistor. S is varied until the detector deflection is zero. A battery is connected between nodes 1 and 3, and a detector joins the nodes 2 and 4. A node is a connecting point for one or more branches. S is adjusted until no potential difference exists between the junction points 2 and 4, so there is no current through the detector. The magnitude of X is to be

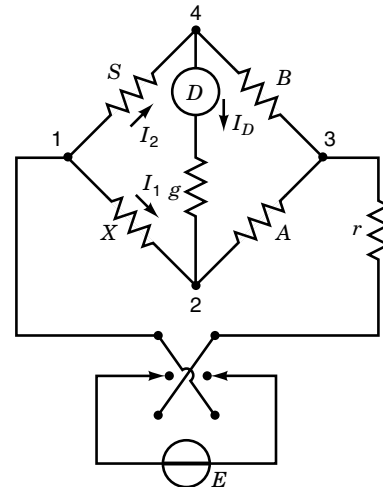


Figure 5. A Wheatstone bridge.

calculated from the equation:

$$X = S \frac{A}{B} \quad (28)$$

In this bridge, four of the six branches of the network containing the resistors A , B , S , and X are also called arms. Due to the fact that the resistors X and S are compared, Wheatstone called this arrangement a resistance balance. The resistors A and B are forming the arms of the balance by analogy with a weighing process.

In practice, the ratio A/B can be 1, 0.1, 0.01, or even lower. Inverse ratios 10, 100, or even higher can also be used. As also indicated in Fig. 5, the polarity of the voltage source is periodically reversed to reduce the effect of thermoelectric voltages. Thermoelectric voltages are mainly generated at the junctions of dissimilar metals at different temperatures. If we consider that the voltage source E has a negligible internal resistance r , and that the null detector has an input resistance g , we can calculate the current I_D in the detector by solving three loop equations in the networks 123E, 143E, and 124 as follows:

$$\begin{bmatrix} X+A & 0 & -X \\ 0 & S+B & S \\ -X & S & S+g+X \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \\ I_D \end{bmatrix} = \begin{bmatrix} E \\ E \\ 0 \end{bmatrix}$$

$$I_D = \frac{\begin{vmatrix} X+A & 0 & E \\ 0 & S+B & E \\ -X & S & 0 \end{vmatrix}}{\begin{vmatrix} X+A & 0 & -X \\ 0 & S+B & S \\ -X & S & S+g+X \end{vmatrix}}$$

$$I_D = E \frac{X(S+B) - S(X+A)}{(X+A)(S+B)(S+g+X) - X^2(S+B) - S^2(X+A)} \quad (29)$$

$$I_D = E \frac{XB - SA}{AX(B+S) + BS(A+X) + g(A+X)(B+S)} \quad (29)$$

When the bridge is balanced, $I_D = 0$, and we find the same relation as in the Eq. (28).

The first two products of the denominator may be replaced by

$$AB(S+X) + SX(A+B)$$

or by

$$AS(B+X) + BX(A+S)$$

The relative current sensitivity σ_i of the Wheatstone bridge near the balance condition can be expressed as follows:

$$|\sigma_i| = \left| \frac{\Delta I_D}{\frac{\Delta S}{S}} \right|_{\Delta S \rightarrow 0} \quad (30)$$

The sensitivity σ_i of a bridge can be optimized by selecting the elements of the bridge. If we slightly modify the balance of the

Wheatstone bridge given in Fig. 5 by changing the adjustable resistor S by a small amount ΔS , the corresponding variation ΔI_D indicated by the detector will be given according to Eq. (29):

$$|\Delta I_D| = \left| E \frac{A \Delta S}{AX(B+S) + BS(A+X) + g(A+X)(B+S)} \right| \quad (31)$$

knowing that at balance, we have also

$$\frac{A}{B} = \frac{X}{S} = \frac{A+X}{B+S} \quad (32)$$

Equation (31) can be rewritten:

$$|\Delta I_D| = \left| E \frac{A \Delta S}{AS(A+X) + BS(A+X) + gS(A+X) + gB(A+X)} \right| \quad (33)$$

$$|\Delta I_D| = \left| E \frac{A \Delta S}{(A+X)[AS + BS + gS + gB]} \right| \quad (34)$$

$$|\sigma_i| = \left| \frac{\Delta I_D}{\frac{\Delta S}{S}} \right| = \left| E \frac{A}{(A+X) \left[A + B + g + \frac{gB}{S} \right]} \right| \quad (35)$$

The parameters of the bridge must be chosen so that this current sensitivity is maximum, taking into account that the value of the source voltage E is limited by the admissible power dissipation in each resistor.

The usual Wheatstone bridge is designed for the measurement of two-terminal devices and so is not suitable for accurate resistance measurements below 1 M Ω .

At present, the Wheatstone bridge remains the basic bridge for many industrial applications with sensors like strain-gauge installations for force, torque, or pressure measurements.

Automated Guarded Wheatstone Bridge

At present in metrology laboratories, the main application of the Wheatstone bridge is reserved for the measurement of high value resistors.

In this case, a guarded Wheatstone bridge is used to compare three-terminal resistors. The availability of highly linear and fully computer-controlled dc sources or calibrators based on the pulse width modulation technique has led to the development of a new type of automated guarded Wheatstone bridge represented in Fig. 6 (5–7). In this bridge, the arm resistors A and B are advantageously replaced by two dc calibrators V_A and V_B having a very good linearity, a low temperature coefficient, and low output impedance. This bridge can be balanced automatically. At balance, the value of R_x is given by the following equation:

$$R_x = \frac{V_A}{V_B} R_S \quad (36)$$

This type of bridge can be used to calibrate resistors from 10 M Ω up to 1 T Ω or more in less time and with lower uncertainty than with the conventional guarded Wheatstone bridge. Combined

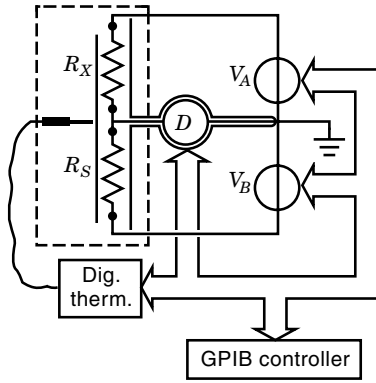


Figure 6. An automated guarded Wheatstone bridge.

measurement uncertainties of order of 1 part in 10^5 can be obtained at $1 \text{ G}\Omega$ with this bridge.

Kelvin Double Bridge

William Thomson (later Lord Kelvin) first described this bridge in a paper published in 1862. This network became known as the Thomson bridge, the Thomson double bridge, the Kelvin bridge, and the Kelvin double bridge. According to F. Wenner (2), Thomson seems to have been the first to attempt measurements of the highest precision attainable with the apparatus then available and to have an understanding of the factors limiting the precision of measurement. As shown in the schematic diagram of Fig. 7, the circuit contains a second set of ratio arms, labeled a and b . The guard circuit is not represented in Fig. 7. This Kelvin bridge may be transformed into an equivalent Wheatstone bridge. The Δ network formed by arms a and b and by link k can indeed be transformed in the Y network by using the theorem of Kennelly.

The detector indication will be zero when the voltage at node 2 equals the voltage at node 5.

$$U_2 = \frac{A}{A+B} I \left[S + X + \frac{(a+b)k}{a+b+k} \right] \quad (37)$$

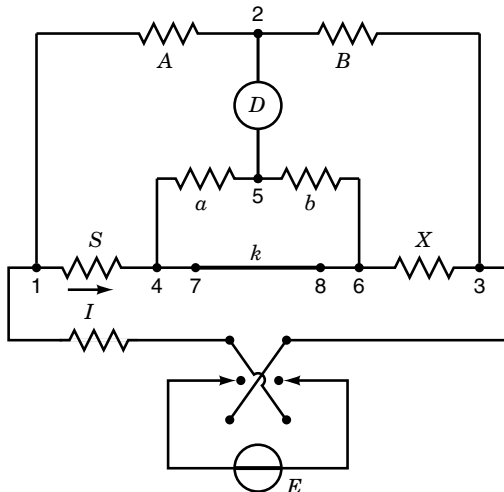


Figure 7. A Kelvin double bridge.

$$U_5 = I \left\{ S + \frac{a}{a+b} \left[\frac{(a+b)k}{a+b+k} \right] \right\} \quad (38)$$

We can solve for X by equating U_2 and U_5 and simplifying:

$$S + X + \frac{(a+b)k}{a+b+k} = \frac{A+B}{A} \left\{ S + \frac{a}{a+b} \left[\frac{(a+b)k}{a+b+k} \right] \right\} \quad (39)$$

$$X = S \frac{B}{A} + \frac{ak}{a+b+k} \left(\frac{B}{A} - \frac{b}{a} \right) \quad (40)$$

By making two successive balances using configurations with and without the link k opened, the condition $B/A = b/a$ is fulfilled. The well-known relationship is established:

$$X = S \frac{B}{A} \quad (41)$$

This bridge is suitable for comparing five-terminal and four-terminal resistors. Nevertheless, it does not entirely satisfy the theoretical definition of a four terminal resistor. Indeed, there exists a voltage difference between nodes 1 and 3. This means, in particular, that a current is flowing through one of the potential leads and terminals of each resistor X and S . This bridge is especially useful for the measurement of 4 Terminal resistors below 100Ω .

Warsawsky Bridge

I. Warsawsky described this bridge in 1955 (8). The schematic diagram is given in Fig. 8. The screens of the resistors are not represented on this diagram. This bridge is of the Wheatstone type and is fitted at each of its four corners with a combining network. Each combining network consists of two auxiliary resistors connecting terminals of adjacent main resistor and also consists of one low-resistance link joining the other terminals. At each corner, the ratio of the auxiliary resistors must be the same as that of the main resistors they join. The exact balance equation of this bridge can be calculated by transforming, at

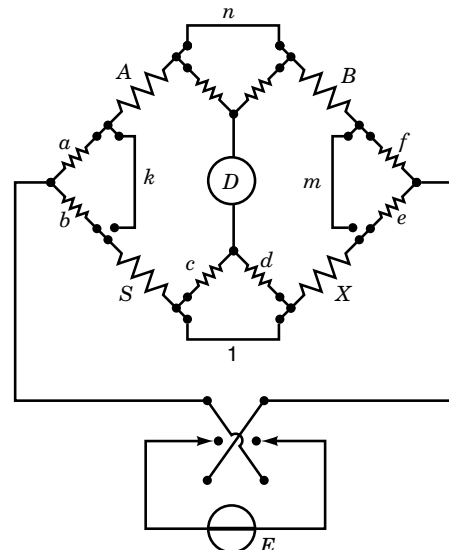


Figure 8. A Warsawsky bridge.

each corner, the Δ network in a T network. The ideal relationship for this bridge is similar to that of the Wheatstone bridge:

$$X = \frac{B}{A} S \quad (42)$$

Departure of less than 1 part in 10^8 from this ideal relationship has been reported if all the resistors of the bridge satisfy the following conditions (9):

- The four main resistors are within 1 part in 10^4 of their mean value.
- The auxiliary resistors are matched pairwise to better than 1 part in 10^4 .
- The low-resistance links are less than 1 part in 10^4 of the main and auxiliary resistances.

These conditions can be achieved in particular when main resistors of 10 k Ω are compared.

Dc Current Comparator Bridge

Originally, the basic design of this bridge comes from the work of N. L. Kusters et al. at the National Research Council of Canada (10,11). The essential feature of the dc current comparator bridge (CCB) is shown in Fig. 9. This CCB consists of

1. Two uniformly distributed ratio windings of N_x and N_s turns, wound on a toroidal magnetic core, are excited in a opposite way by two isolated and reversible direct current sources acting as Master and Slave
2. A zero flux detector and a double core magnetic modulator with a magnetic shield. This magnetic shield has the form of a hollow toroid, made of high permeability material
3. A null detector

The balancing of the bridge is completed when two balance conditions are fulfilled:

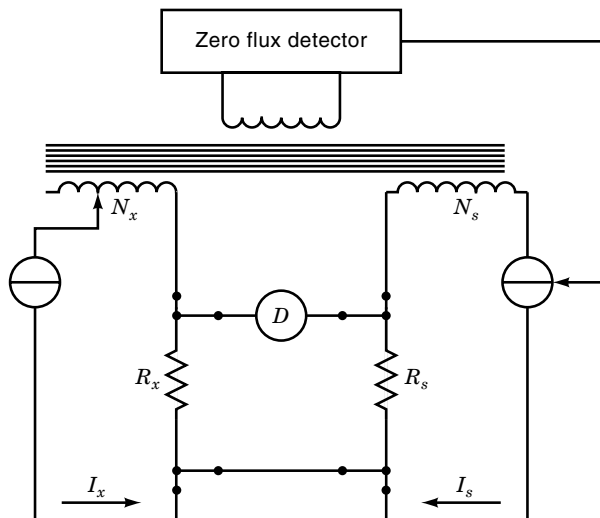


Figure 9. A dc current comparator bridge.

- The magnetomotive force balance is achieved automatically by means of the servo circuit operating from the output signal of the zero flux detector. This servo circuit adjusts the current I_s delivered by the slave power supply in order to maintain the difference of the magnetomotive forces of the two windings as close to zero as possible. This gives the first condition:

$$N_s I_s = N_x I_x \quad (43)$$

- The zero voltage condition which is obtained, in the manually operated version of this type of bridge, by changing the turns ratio N_x by rotating switches until the zero indication of the null detector is achieved. This gives the second condition:

$$I_x R_s = I_x R_x \quad (44)$$

From these above-mentioned relations, it can be deduced

$$R_x = R_s \frac{N_s}{N_x} \quad (45)$$

The dc current comparator has several advantages with respect to the previously described dc bridges:

1. The measured ratio depends for the most significant decades only on numbers of turns. This means that this type of bridge are not likely to drift with time or to change with ambient temperature, relative humidity, and atmospheric pressure.
2. The latest available versions of this kind of current comparator bridge is automatic and feature a linearity often better than 3 parts in 10^8 .
3. It is possible to compare resistors according to the four-terminal and five-terminal definitions.
4. The unknown and standard resistors are not connected in series. This means that the current flowing through them can be different.

In addition to the limitations common to other bridges due to the resolution of the null detector and thermal noise, we have two other sources of noise: Barkhausen noise in the magnetic core and electrical noise in the electronic servo circuit. Currently, commercial versions of this type of dc current comparator are automatic. They can be used to measure dc resistors and shunts between 1 m Ω and 1 M Ω . Their sensitivity is about 0.1 μ A for a winding of 1 turn.

Dc Cryogenic Current Comparator Bridge

Similar in its principle to the dc CCB, the dc cryogenic CCB has its ratio windings, toroidal overlapping lead shield, detection winding and flux detector made of superconductors or superconducting devices. As for the dc current comparator bridge, at balance, the ratio of currents I_s and I_x flowing through the windings is also the inverse ratio of the number of turns. The working principle is the Meissner effect. If the magnetomotive forces generated by the windings inside the superconducting shield are not fully equal and in opposition, a current, proportional to this imbalance, will circulate in the outer surface of this toroidal shield. This current will be detected by means of

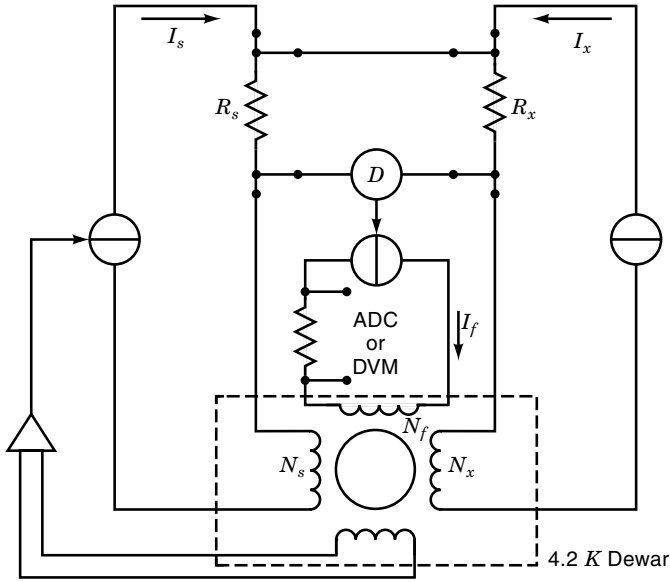


Figure 10. A dc cryogenic current comparator bridge.

a superconducting flux transformer made of a low inductance superconducting coil. The latter is connected in series with the superconducting input coil of a very sensitive SQUID (Superconducting Quantum Interference Device) magnetometer. This type of magnetometer is presently the most sensitive magnetic field sensor known. A schematic diagram of a cryogenic current comparator bridge (CCCB) is illustrated in Fig. 10. Two reversible isolated and fully floating current sources generate currents I_s and I_x in approximately the inverse ratio of the winding turns N_s and N_x . The SQUID provides the magnetomotive force balance by controlling the Slave current source through the SQUID output and an isolation amplifier. The null detector amplifies the potential difference across the two resistors R_s and R_x . The output of this detector is used to drive a third current source, which provides a feedback current I_f through a feedback winding N_f . This feedback current is known by measuring the voltage across a known resistor with a digital voltmeter or an analogue to digital converter (ADC). At balance, we have again two equations. At zero indication of the null detector, we have

$$R_s I_s = R_x I_x \quad (46)$$

and when the magnetomotive force balance is achieved,

$$N_x I_x + I_f N_f - N_s I_s = 0 \quad (47)$$

If we replace I_s in (46), we obtain

$$\frac{R_x}{R_s} = \frac{N_x}{N_s} \left(1 + \frac{I_f N_f}{I_x N_x} \right) \quad (48)$$

A typical 1σ combined uncertainty of a few parts in 10^9 can be obtained with a commercially available CCCB (12) in the measurement of 1:1 ratios for resistors of 100 Ω . This uncertainty depends mainly on

- Calibration of the bridge and accessories
- Leakage resistances

- Electrical noise in the null detector, the SQUID, and the resistors

A sensitivity of about 1 nA for a winding of one turn can be achieved.

Dc Sources

The balance conditions calculated above for most of the bridges are independent of the values of the voltage sources. To reduce the influence of thermal emf's, the polarity of the source is regularly and often automatically reversed. It is reasonable to think that after these reversal operations, the voltage of the source must be kept reasonably stable.

The advantages of computer controlled dc calibrators have also been demonstrated for the realization of automated guarded Wheatstone bridges. In the past, depending on the value of the measuring current, batteries or lead acid accumulators were generally used as sources in dc bridges. Mercury batteries were also often used in different bridges because of their low drift and low temperature coefficient. However, they are no longer available because of new legislation concerning the environment. Today, the efficiency of shielding and noise reduction techniques has also increased, so those sources operating from the electrical mains are now more often used. Also, thanks to design improvements and automatization, power supplies, sometimes connected in parallel, are now the most convenient solution for the calibration with dc current comparator bridges of large current-carrying resistors or resistive shunts for currents from 1 A to 2000 A or more. For applications involving low power consumption and when it is strictly necessary to avoid noise and leakage paths due to the electrical mains, batteries can provide fully isolated voltage or current sources.

Dc Detectors

Today, electronic null detectors have replaced nearly all the galvanometers and the photogalvanometer amplifiers. Some electronic null detectors, nanovoltmeters, and certainly picovoltmeters exceed the sensitivity of most of the older detectors. Picovoltmeters operating at 4.2 K and based on a SQUID have been developed (13). An increasing number of these electronic instruments are more and more often operated from controlling computers via a parallel or a serial bus despite the problem arising from the additional noise introduced by this type of connection.

Noise coming from the bridge influences the null detector. This includes

- Radiated and conducted electromagnetic interference
- Thermal noise or Johnson noise due to the bridge. This means, in particular, that in some applications, the resistance of the connecting leads must be taken into account
- Barkhausen noise, which can be present with magnetic circuits like those in dc current comparators bridges

In addition to external noise and even when they are operating on internal batteries, electrical noise is also generated inside these electronic null detectors (14):

- Thermal noise or Johnson noise which emanates from all resistive devices. The value of the open-circuit voltage is

given by the following equation:

$$V = \sqrt{4kTRB} \quad (49)$$

where k = Boltzmann's constant (about 1.38×10^{-10} J/K); T = Absolute temperature in Kelvin; R = Resistance in ohm; B = Equivalent noise bandwidth in hertz.

- Shot noise caused by the current flow across a potential barrier
- Popcorn noise mainly due to some manufacturing processes of semiconductors and integrated circuits
- Contact noise due to imperfect contacts in switches or relays

An equivalent circuit of a dc detector can be represented by an ideal noiseless detector of infinite input impedance, shunted by a ideal noise current generator i_n and having an ideal noise voltage source e_n in series with.

Reliable dc nanovoltmeters and picovoltmeters are now commercially available. The most sensitive of them, operating at room temperature, has an equivalent noise resistance of about 30 mΩ, a common mode rejection ratio above 180 dB, and can detect potential differences as small as a few tens of picovolts when connected to a low-impedance circuit.

Checking of dc Bridges Accuracy

To comply with international quality standards, laboratories send, at specified intervals, standard resistors and even bridges to a national metrology service or to a qualified accredited laboratory for calibration. Between such calibrations, it is good practice that each laboratory performs elementary checks on its bridges. Some checks of dc resistance bridges include

1. *Zero Voltage of the Bridge Source.* When possible, it is always useful to disconnect the power source from the bridge and to short the input leads to the bridge or to turn the output voltage down to zero. After doing this, any deflection of the null detector will be an indication of the remaining thermal emfs, external signals, and noise, which are not due to the source.
 2. *Zero Resistance.* Replace the device under test by a short circuit.
 3. *Infinite Resistance.* Replace the device under test by an open circuit.
 4. *Checks of the Bridge Linearity.* Linearity check can be made using individual resistors or resistor networks.
- *With Individual Resistors.* Let us assume that we have two similar resistors A and B to compare and that we make two successive measurements $Rd_1 = A/B$ and $Rd_2 = B/A$. If the bridge under test has no linearity error, the product of these readings must be equal to 1. It is interesting to point out that this check does not require knowing of the exact value of these resistors A and B . The only technical requirement for these resistors is a good short term stability.
 - *With Resistive Networks.* A network of n nominally equal four-terminal resistors connected end to end with ideal zero-resistance junctions. In this case, each resistor of the network is successively compared by means of the bridge

under test with a separate one of the same nominal value. A first approximation of the linearity of the bridge at one point of the string can be obtained if the reading is equal to the sum of the individual readings up to this point:

$$R_i = \sum_{n=1}^i R_n$$

- *A Hamon Resistor Network.* A Hamon network is realized by connecting first in series and then in parallel the same number of n nominally equal four terminal resistors. It can be demonstrated that if all of these resistors are within 1 in 10^4 of their mean value, the ratio of the series/parallel combination R_s/R_p is equal to n^2 within 1 part in 10^8 (15). Accurate 1/100 ratios and even 1/10 ratios can be made with these Hamon build-up resistors.
- Very recently a resistance network, similar to the Hamon's design, was especially built for checking dc resistance bridges. It consists of a network of four main resistors connected to a common carefully engineered four-terminal junction. This produces 35 different resistance values of the four-terminal type and provides a convenient way to verify the accuracy of resistance bridges (16).

AC BRIDGE INSTRUMENTS

Many of the ac bridge instruments described in this section are mainly used in the audio-frequency range. However, some of them are still correctly operating, with higher uncertainties, slightly above the 1 MHz range.

Ac Networks

Most of the existing circuit components and even most of the impedance standards to be measured with ac bridges are neither purely resistive nor purely reactive. This means that unwanted capacitance and inductance are present in resistors, unwanted resistance and inductance in capacitors as well as unwanted resistance and capacitance in inductors. The energy flow at sufficiently low frequencies in these real-world devices must be treated like a flow in a RLC network. When driven with an ac voltage, each RLC network will dissipate heat and store the energy of the generated magnetic and electric fields.

If no precaution is taken in the construction and the use of this RLC network, the magnetic and electric fields generated by this network will be modified by any change with respect to its immediate environment. The final measured value of this RLC network will be altered. As in ac bridges, the comparison is made in the energy flow through two or more sets of impedance standards; special care in order to limit the influence of these fields is the first criterion to be fulfilled in the design of these impedance standards. In theory, for a standard of capacitance, the only energy should be that associated with the electrical field E , and for a standard of inductance that associated with the magnetic field H . For a standard of resistance, the only energy should be dissipated in heat.

Ac Measured Devices

The definitions given for dc resistors are, even for the most sophisticated of them, not satisfactory in low-frequency ac measurements mainly due to the frequency dependence of the

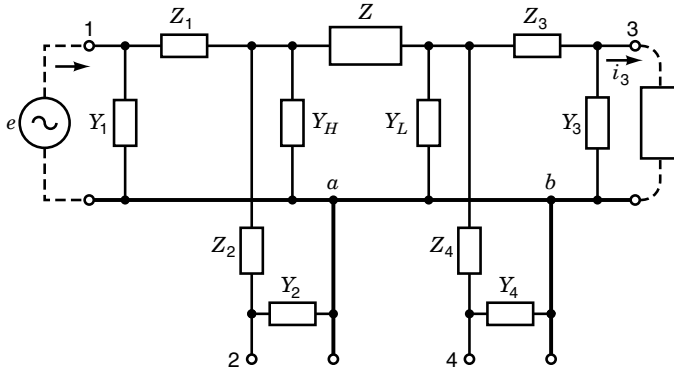


Figure 11. Four terminal-pair definition.

measured value, the series impedances, the shunt admittances, and the stray fields present in each device. Other definitions for the devices to be measured in ac have been developed in order to overcome these problems.

Four Terminal-Pair Definition. Today, it is generally accepted that the four terminal-pair definition is the most accurate. The interest of this definition has been firstly pointed out by R. D. Cutkosky (17) and more recently refined and intensively explained by B. P. Kibble and G. H. Rayner (18). An example of a simplified circuit of a four terminal-pair impedance is shown in Fig. 11. The numbering of the terminal-pairs or ports has been kept similar to that used by the first pioneers in this field. In order to fulfill the astatic condition, the coaxial geometry is intensively used. This means also that the impedance Z is contained in a shielded box, not represented in Fig. 11, and is connected to the measuring bridge by means of suitable coaxial sockets or plugs. In Fig. 11, again by analogy with the convention (9,18), the connections between the inner conductor of each coaxial connector are represented by a fine line and the connections between the outer conductor by a thicker line. This difference points out the low series impedance of this outer conductor per unit length. A generator may be connected to the current terminal-pair 1 to supply the necessary current and a second four terminal-pair impedance to be compared with the first one at the other current terminal-pair 3. The four terminal-pair impedance Z' is defined as

$$Z' = \frac{e_2}{i_3} \quad (50)$$

The required conditions for this definition are the following:

- At port 4: $e_4 = 0$ and $i_4 = 0$
- At port 3: $i_{3 \text{ OUT}} = i_{3 \text{ IN}}$
- At port 2: $i_2 = 0$

From Fig. 11, the impedance Z , including the low series impedance between the nodes a and b, can be expressed as

$$Z = Z'(1 + Z_2 Y_2)(1 + Z_3 Y_3) \quad (51)$$

These first order correcting terms $Z_i Y_i$ are due to the connections, the coaxial connectors, and the connecting cables. These terms may be forgotten if it is accepted that the impedance is

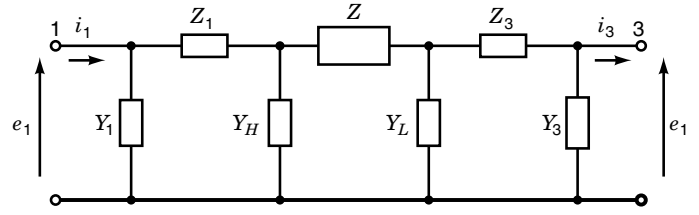


Figure 12. Two terminal-pair definition.

defined at the end of the connecting cables. It can also be calculated that these correcting terms $Z_i Y_i$ are lower than 1×10^{-8} at 1 kHz but nearly reach 1×10^{-3} at 1 MHz for 1 m length of 50 Ω ordinary PTFE isolated coaxial cable. It is also worth mentioning that these correcting terms do not involve the value of the impedance Z . This means that the four terminal-pair definition is valid for all values of the impedance Z and especially as long as the ratio between the length of the connecting leads or coaxial cables and the wavelength of the applied signal remains negligible. In practice, achievement of the definition condition at port 4 is limited by the sensitivity of the used null detector and usually requires the use of a combining network connected to both impedances to be compared. At port 3, the equality of the inward and outward currents at the terminals is obtained by means of a current equalizer. Port 2 is connected to a reference output voltage potential of the bridge, and auxiliary adjustments are necessary until no current flows at port 2. Despite the complexity of bridges designed for the calibration of four terminal-pair impedances, they constitute today the most widely accepted way to achieve the lowest uncertainty.

Two Terminal-Pair Definition. This definition is illustrated in Fig. 12. In comparison with Fig. 11, ports 2 and 4 have been deleted. The two terminal-pair impedance is defined as

$$Z' = \frac{e_1}{i_3} \quad (52)$$

The required condition is the following:

- At port 3: $e_3 = 0$ and $i_{3 \text{ OUT}} = i_{3 \text{ IN}}$

By analogy with the four terminal-pair definition, we have

$$Z' = kZ \quad (53)$$

where k is a correcting factor, the value of which can be found in the literature (17,18). From a practical point of view, it can be seen from Fig. 12 that the values of Y_H and Y_L do not affect this definition. However, Y_H and Y_L enter in the correcting factor k .

Four-Terminal Coaxial Definition. A simplified circuit illustrating this definition is shown in Fig. 13. The leakage resistance, the capacitance, the series resistance, and inductance of the connecting cables and of the input connectors, which are not represented in this figure, are also influencing the measured device. This type of definition is mainly suitable for resistive devices. Several ac resistance thermometer bridges are based on this definition (19).

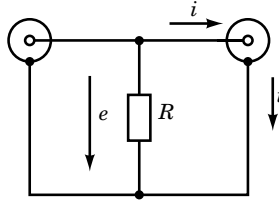


Figure 13. Four-terminal coaxial definition.

Three-Terminal Definition. Many capacitors manufactured in the past were based on this definition. In some of them, the outer conductor of the high potential connector is not connected to the outer conductor of the low potential one. This is not a fundamental problem when they are properly used.

Four-Arm ac Bridges

In ac, the conventional four-arm resistive bridges (also called Wheatstone bridges) and more complicated multi-arm bridges are used. Generally, these multi-arm bridges are modified four-arm bridges and can, in principle, be reduced to simple Wheatstone bridges through $Y - \Delta$ or $\Delta - Y$ transformations (theorem of Kennelly). The problem becomes more difficult when the arms of these bridges consist of complex impedances as it is the case for the measurement of inductors, and capacitors. A complex impedance can be arranged like a network in which the resistors, the inductors and capacitors can be connected in series or in parallel or both. This means that there exist a great variety of four-arm ac bridges. This multiplicity of possible network arrangements makes the classification of ac bridges somehow difficult taking also into account that some combinations of these networks are impractical and even impossible to balance.

The most general form of a four-arm ac impedance bridge is the generalization in ac of the dc Wheatstone bridge, represented in Fig. 14. The structure is similar but the resistances R_1, R_2, R_3, R_4 are replaced by impedances Z_1, Z_2, Z_3, Z_4 . The generator G and detector D are operating at least partially in a common frequency range. At balance, the general equation of this bridge will be

$$Z_1 Z_3 = Z_2 Z_4 \quad (54)$$

Using the complex notation and substituting in Eq. (54)

$$(R_1 + jX_1)(R_3 + jX_3) = (R_2 + jX_2)(R_4 + jX_4) \quad (55)$$

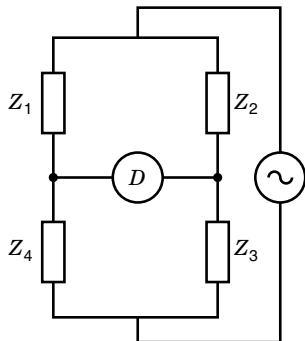


Figure 14. General form of a four-arm impedance bridge.

By solving the real and imaginary terms separately, we obtain

$$R_1 R_3 - X_1 X_3 = R_2 R_4 - X_2 X_4 \quad (56)$$

$$X_1 R_3 + X_3 R_1 = X_2 R_4 + X_4 R_2 \quad (57)$$

The main difference with the dc version is that the balance condition is given by a system of two equations which must be simultaneously fulfilled. This is rather normal because two unknown quantities, which are the resistive and reactive components of the same complex impedance, have to be measured. The balance calls for the simultaneous adjustment of at least one resistive and one reactive parameter in the network. This becomes clearer if we rewrite the general balance equation in the form:

$$|Z_1 Z_3| e^{j(\varphi_1 + \varphi_3)} = |Z_2 Z_4| e^{j(\varphi_2 + \varphi_4)} \quad (58)$$

The two conditions for the balance of an ac bridge in magnitude and in phase can then be expressed as follows:

$$|Z_1 Z_3| = |Z_2 Z_4| \quad (59)$$

$$\varphi_1 + \varphi_3 = \varphi_2 + \varphi_4 \quad (60)$$

Thus, the balance of an ac bridge is obtained when the products of the magnitudes and the sum of the phase angles of the impedances in the pairwise opposite arms are made equal.

We will only consider the most useful of these bridges. To simplify the study of the different bridges, the balance conditions of each bridge will be determined assuming that perfect two terminal components and connecting leads or connections are used, which is far from the truth in practice.

Wien Capacitance Bridge. This bridge is shown in Fig. 15. Two adjacent arms of this bridge have nonreactive resistors, and the other two arms have capacitors respectively in series and in parallel with resistors. It is used to determine capacitance in terms of resistance and frequency. The balance equation is

$$R_1 \left(\frac{R_x}{1 + j\omega C_x R_x} \right) = R_2 \left(R_3 + \frac{1}{j\omega C_3} \right) \quad (61)$$

By solving this equation, we obtain

$$C_x = \frac{R_1 C_3}{R_2 (1 + \omega^2 R_3^2 C_3^2)} \quad \text{and} \quad R_x = \frac{R_2 (1 + \omega^2 R_3^2 C_3^2)}{\omega^2 C_3^2 R_1 R_3} \quad (62)$$

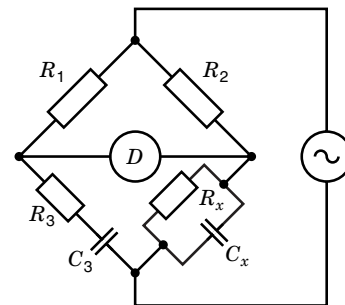


Figure 15. A Wien capacitance bridge.

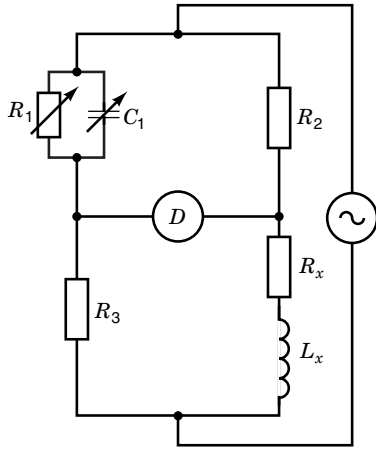


Figure 16. A Maxwell-Wien bridge.

This Wien bridge is also widely used in electronics, for example, to realize low distortion RC oscillators. In this case, the balance conditions may be rewritten

$$\frac{C_x}{C_3} = \frac{R_1}{R_2} - \frac{R_3}{R_x} \quad \text{and} \quad \omega^2 = \frac{1}{C_3 C_x R_3 R_x} \quad (63)$$

Maxwell-Wien Bridge. This bridge is normally used for the measurement of inductance in terms of resistance and capacitance. It was originally introduced by Maxwell as a ballistic method and at first used as an ac bridge by Wien (3,4). This bridge is shown in Fig. 16. The balance equation is

$$\frac{R_1(R_x + j\omega L_x)}{(1 + j\omega C_1 R_1)} = R_2 R_3 \quad (64)$$

or if we separate the real and imaginary parts:

$$L_x = C_1 R_2 R_3 \quad \text{and} \quad R_x = \frac{R_2 R_3}{R_1} \quad (65)$$

It can be deduced that the balance is independent of frequency.

Hay Bridge. This bridge is particularly suitable to measure inductances with large time-constants. It differs from the Maxwell-Wien bridge in that in the arm opposite to the inductor, the capacitor is in series with the resistor. The circuit arrangement is given in Fig. 17. At balance,

$$\left(R_1 + \frac{1}{j\omega C_1}\right)(R_x + j\omega L_x) = R_2 R_3$$

Solving for L_x and R_x ,

$$L_x = \frac{R_2 R_3 C_1}{(1 + \omega^2 R_1^2 C_1^2)} \quad \text{and} \quad R_x = \frac{\omega^2 R_1 R_2 R_3 C_1^2}{(1 + \omega^2 R_1^2 C_1^2)} \quad (66)$$

These balance conditions are frequency dependent. The frequency must be well known and very stable. This is obtained nowadays very easily by using frequency synthesizers.

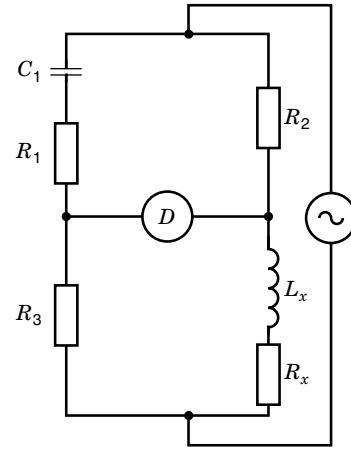


Figure 17. A Hay bridge.

Schering Bridge. Schering first suggested this bridge in 1920. This bridge is used for the measurement of capacitance and dissipation factor. The unknown capacitor C_x and the standard loss-free capacitor C_2 form two adjacent arms, while the opposite arm to the capacitor C_x consists of a resistor R_1 and a capacitor C_1 in parallel. The fourth arm is a nonreactive resistor R_3 . The balance may be written referring to Fig. 18:

$$\left(R_x + \frac{1}{j\omega C_x}\right)\left(\frac{R_1}{1 + j\omega C_1 R_1}\right) = \frac{R_2}{j\omega C_2} \quad (67)$$

By separating the real and the imaginary parts:

$$C_x = \frac{C_2 R_1}{R_2} \quad \text{and} \quad R_x = \frac{C_1 R_2}{C_2} \quad (68)$$

The balance conditions are independent of frequency.

Wagner Principle. Already in many dc bridges but still more in ac bridges, the problem of leakage currents, coming, for example, from the source directly to the detector due to the poor isolation from the conductor of the mains supply called ground or earth or due to stray capacitances, must be under control. In many ac bridges, this is achieved by adding an auxiliary circuit or branch called Wagner earth to the bridge. It can be deduced from Fig. 19 that the admittances of the bridge and

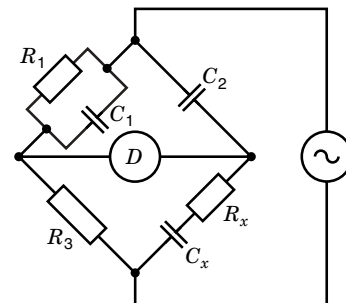


Figure 18. A Schering bridge.

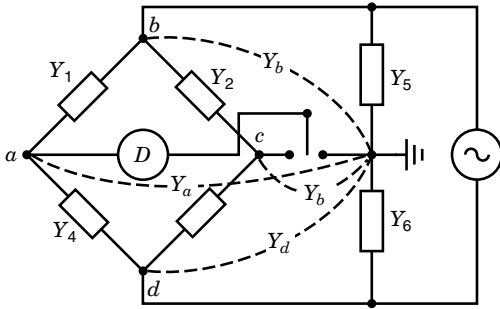


Figure 19. The Wagner principle.

the auxiliary circuit respect the following conditions:

$$\frac{Y_1}{Y_4} = \frac{Y_2}{Y_3} = \frac{(Y_5 + Y_b)}{(Y_6 + Y_d)} \quad (69)$$

then the detector terminals are at the same potential as the earth or ground potential without being connected to the earth or to the ground. The admittances Y_a , Y_c and Y_b , Y_d have, consequently, no effect because they are respectively connected between equipotential points and in parallel with the admittance Y_5 and Y_6 of the auxiliary arms. The condition (69) is fulfilled by balancing successively the bridge constituted by Y_1 , Y_2 , Y_3 and Y_4 and the bridge consisting of Y_1 , Y_4 , Y_5 and Y_6 or of Y_2 , Y_3 , Y_5 and Y_6 . The procedure is iterated until a consistent balance is obtained for the two possible connections of the detector.

Multiple-Arm Bridges

Other ac bridges having six arms and even more rarely seven arms have been described in the technical literature and also used successfully (3). However, it will be shown that several of these networks may be reduced to the basic four-arm configuration by suitable $Y - \Delta$ or $\Delta - Y$ transformations.

Anderson-Type Bridge. The interest of the modification introduced by Anderson has been demonstrated in the Maxwell-Wien bridge to measure inductance in terms of capacitance and resistance. If we apply the theorem of Kennelly to the triangle cde of the network shown in Fig. 20, we have the following equations:

$$Z'_3 = \frac{Z_5 Z_6}{Z_3 + Z_5 + Z_6} \quad (70)$$

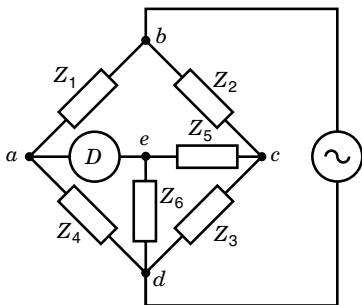


Figure 20. An Anderson bridge.

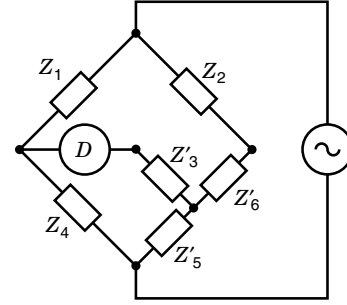


Figure 21. A modified Anderson's type bridge after a $Y - \Delta$ transformation.

$$Z'_5 = \frac{Z_3 Z_6}{Z_3 + Z_5 + Z_6} \quad (71)$$

$$Z'_6 = \frac{Z_3 Z_5}{Z_3 + Z_5 + Z_6} \quad (72)$$

The general balance equation of the network represented in Fig. 21 is

$$Z_2 Z_4 (Z_3 + Z_5 + Z_6) + Z_3 Z_5 Z_4 = Z_1 Z_3 Z_6 \quad (73)$$

Ac Kelvin Double Bridge. The need to characterize four terminal low value resistors and, in particular, to know their time-constants and their residual inductances has led to extend the use of the Kelvin double bridge from dc to low frequency ac. A general form of this bridge is given in Fig. 22. Let us apply again the theorem of Kennelly to the network 456:

$$Z'_a = \frac{Z_b Z_c}{Z_a + Z_b + Z_c} \quad (74)$$

$$Z'_b = \frac{Z_a Z_c}{Z_a + Z_b + Z_c} \quad (75)$$

$$Z'_c = \frac{Z_a Z_b}{Z_a + Z_b + Z_c} \quad (76)$$

The general equation of the network shown in Fig. 23 is

$$(Z_a + Z_b + Z_c)(Z_A Z_X - Z_B Z_S) = Z_c(Z_B Z_a - Z_A Z_b) \quad (77)$$

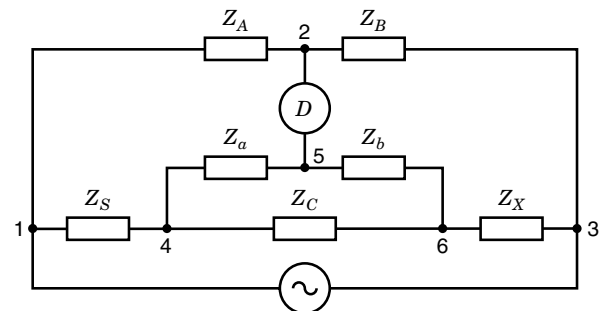


Figure 22. An ac Kelvin double bridge.

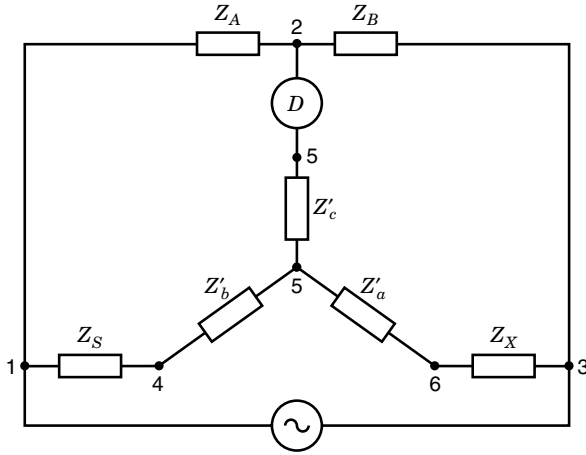


Figure 23. A modified ac Kelvin double bridge after a $Y - \Delta$ transformation.

Therefore, the double condition as follows permits the achievement of balance:

$$\frac{Z_S}{Z_X} = \frac{Z_A}{Z_B} \quad \text{and} \quad \frac{Z_A}{Z_B} = \frac{Z_a}{Z_b} \quad (78)$$

Twin-T Networks. Networks other than the four-arm bridges have been developed to compare impedances. An important one among them is the Twin-T network. This circuit illustrated in Fig. 24 may be regarded as a parallel-T network. The direct impedances Z'_6 and Z'_3 between nodes 1 and 2 can be calculated after $Y - \Delta$ transformations (Fig. 25). If these direct impedances Z'_6 and Z'_3 are equal in magnitude and have phase angles differing by π radians, the currents through these impedances cancel one another at the input of the detector. The other admittances obtained after these transformations do not affect the balance condition because they shunt either the source or the detector. The balance condition of this parallel-T network is given by the following equation:

$$Z_1 + Z_2 + Z_1 Z_2 Z_3 + Y_4 + Z_5 + Z_4 Z_5 Y_6 = 0 \quad (79)$$

Therefore the null at the detector is only possible if the impedances contain reactances. This type of bridge is particularly suitable for use at high frequencies. An interesting feature of this type of network is the fact that the source and the detector have a common point. Therefore, the bridge balance is not disturbed by stray capacitances.

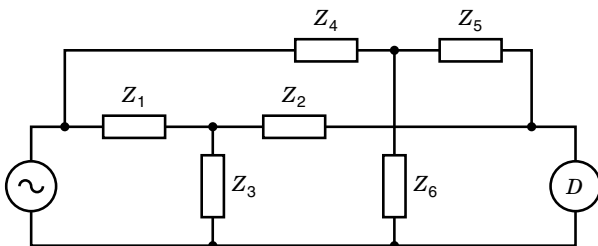


Figure 24. A Twin-T network.

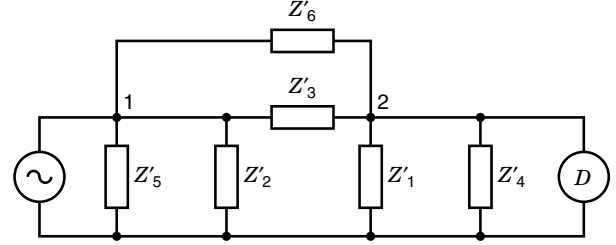


Figure 25. $Y - \Delta$ transformations applied to a Twin-T network.

Transformer-Ratio Arm Bridges

In the four-arm or the multi-arm ac bridges, which have been examined before, the ratio arms are generally realized by fixed nonreactive resistances in a definite numerical ratio. For forty years, ac bridges have been designed where the impedance ratios in adjacent arms are advantageously replaced by voltage ratios from inductive voltage dividers (IVDs). An IVD is an accurate autotransformer provided with a variable tapping point.

At present, 8- or 9-decade IVDs with a departure from the nominal ratio of the number of turns of the windings lower than one part in ten million in the audio-frequency range are commercially available. Some of them are also correctly operating up to 1 MHz. Applications using 30 bits binary voltage dividers have also been described (20). The automatization of the IVDs seems also to be a promising feature for these devices.

Ratios defined by IVDs are very stable as a function of time, temperature, humidity, and even frequency. At 1 kHz, uncertainties lower than 3 parts in 10^8 for the ratio and less than $1 \mu\text{rad}$ for the phase are currently achieved for IVDs manufactured by qualified specialists in the field.

In transformers especially designed for ratio measurements, the departure from the ideal transformer has mainly the following origins:

- Power losses due to the magnetization and eddy currents in the core
- Power losses due to Joule effect, skin effect, and eddy currents in the windings
- Flux leakages
- Leakage inductances and capacitances

Potential Transformer-Ratio Arm Bridge. A simple example of a bridge, which might be called potential transformer bridge, is shown in Fig. 26. The balancing condition is easily obtained. Indeed, at node 1,

$$I_1 = I_2 \quad (80)$$

$$\frac{V_1}{Z_1} = \frac{V_2}{Z_2} \quad (81)$$

If the potential transformer is ideal,

$$\frac{V_1}{V_2} = \frac{n_1}{n_2} \quad (82)$$

$$\frac{Z_1}{Z_2} = \frac{n_1}{n_2} \quad (83)$$

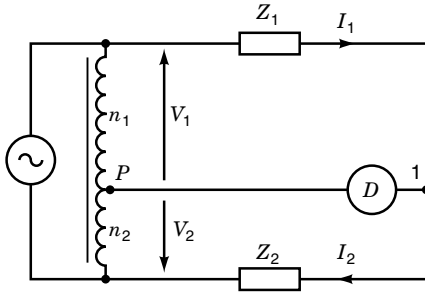


Figure 26. A potential transformer-ratio arm bridge.

In the potential transformer-ratio bridge, the ratio of the number of turns of the windings n_1 and n_2 determines the ratio of the voltage across the other arms. It can also be observed from the same diagram that, if the variable tapping point P is grounded, the stray capacitances associated with Z_1 and Z_2 are in parallel with the internal impedances of the two parts of the potential transformer. As these impedances are in general rather low, the admittances of these stray capacitances do not affect the balance of this simple bridge. If a lower uncertainty is requested, the generator must at least be isolated; a Wagner branch and an injection network have to be added to correctly balance this bridge.

Current Transformer-Ratio Bridge. This bridge, represented in Fig. 27, is the dual bridge of the potential transformer-ratio bridge. This bridge is a three-winding type where the extra winding detects the magnetomotive force balance. The magnetic and electrostatic screens between the toroidal windings and the detector winding are not represented. These screens are often necessary to improve the accuracy of this type of bridge. Contrary to the potential transformer-ratio bridge, which operates with a finite magnetic flux in the magnetic core, the flux in the magnetic core of a current transformer ratio bridge is theoretically zero.

When the bridge is balanced, the magnetomotive forces created by the currents flowing through the two impedances and the two windings nullify each other. At balance, we have

$$n_1 I_1 - n_2 I_2 = 0 \quad (84)$$

or

$$\frac{n_1}{Z_1} = \frac{n_2}{Z_2} \quad (85)$$

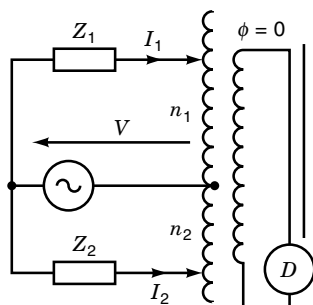


Figure 27. A current transformer-ratio bridge.

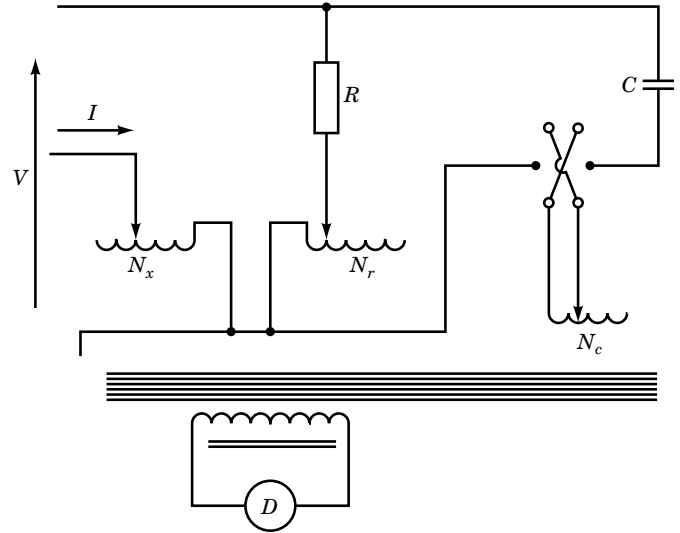


Figure 28. An ac current comparator bridge.

Ac Current Comparator Bridge

Many applications of the ac current comparator have been described for low as well as for high voltage and current applications (10). In low voltage applications, an ac current comparator is used in the development of many useful instruments like current transformers, active capacitor/quadrature current references, transconductance amplifiers, resistance-ratio ac bridges (21), and power comparators. We will describe this last one, which is also known as power bridge due to its use for the measurement of power and energy. The basic circuit, shown in Fig. 28, offers the possibility to provide a power standard (22). This bridge compares the magnetomotive forces created by the current I in the winding N_x with the magnetomotive forces in the winding N_r due to the in-phase current flowing through the resistance R and that in the winding N_c due to the quadrature current through the capacitor C . The balance condition is

$$IN_x = \frac{V}{R} N_r \pm j\omega C V N_c \quad (86)$$

By rewriting Eq. (86),

$$VI = \frac{1}{N_x} \left(\frac{V^2}{R} N_r \pm j\omega C V^2 N_c \right) \quad (87)$$

This bridge measures the apparent power VI that has one active component proportional to V^2/R and one reactive component proportional to $\omega C V^2$.

Coaxial ac Bridges

This is the latest refinement in the field of ac bridges. Measurement uncertainties as low as 1 part in 10^9 for impedance comparison can be achieved from about 200 Hz to 2 kHz with these kinds of coaxial ac bridges. However, the technique can be extended from 10 Hz to slightly above 1 MHz. The technique of the coaxial ac bridges is based on some of the following

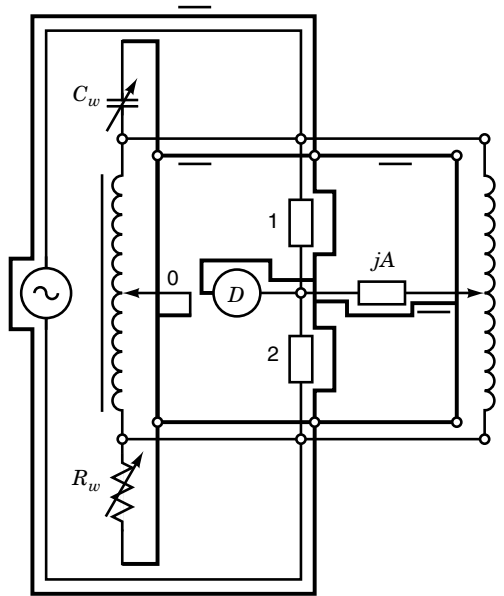


Figure 29. A two terminal-pair bridge based on IVDs [From (18) with the consent of Dr. B. P. Kibble and the permission of IOP].

features:

1. Complete equality in magnitude and opposition in sign between the currents flowing in the inner and the outer conductor of the coaxial cable. By this means, no external field is created, and no electromagnetic interference is influencing the bridge.
2. Low impedance of the outer conductor of the cable and of the screens of the measuring devices
3. Fully screened or shielded measuring devices provided with terminal-pairs

Coaxial ac bridges are used to realize the link between the farad and the ohm and vice versa. A simple example of a coaxial ac bridge is shown in Fig. 29, with the permission of one of the authors (18). This bridge built around a calibrated IVD can be used to compare two terminal-pair measuring devices of similar admittance with a measurement uncertainty lower than 1 part in 10^6 .

Automatic RLC Bridges

Many of the ac bridges described in the previous chapters are not very widespread in industrial laboratories and not very practical for component manufacturers. In many sectors of activities, the preference is mainly given to automatic *RLC* bridges, which are more easily available due to the fact that several manufacturers commercialize them. These *RLC* bridges offer several advantages:

- Automatic and quick measurement of the real and imaginary parts of the impedance vector and conversion of them into the desired parameters like $|Z|$, $|Y|$, Θ , R , X , G and B
- Computer control by means of an IEEE 488 interface
- Frequency range from a few hertz up to several megahertz

A schematic diagram of an HP 4284A Precision *LCR* Meter is represented in Fig. 30 with the permission of the manufacturer (23). As indicated in this figure, the structure of this bridge allows the measurement of 4TP devices. Coaxial cables are used to isolate the voltage sensing cables from the signal current path. Since the same current flows, but in an opposite direction through the outer conductor of the same coaxial cable, the magnetic field generated by the inner conductor is canceled. The outer conductor of the coaxial cable is also connected at the system ground.

This bridge uses the auto balancing method, where the current flowing through the device under test is balanced by the range resistor current in order to maintain a zero potential at the low terminal of the device under test. A null detector in amplitude and in phase measures the zero potential at this point. If the zero potential is not achieved, the null detector adjusts the amplitude and/or the phase of the second oscillator in order to maintain this balance condition. This bridge, operating up to 1 MHz, allows the measurement of devices with a basic accuracy of 0.05% at 1 kHz. Better performances could be achieved when using a substitution method and standards calibrated with lower uncertainties. Before using this type of bridge, it is also needed to proceed to an OPEN/SHORT/LOAD compensation technique. The length of the cables must be taken into account, especially if the frequency is above the kilohertz band.

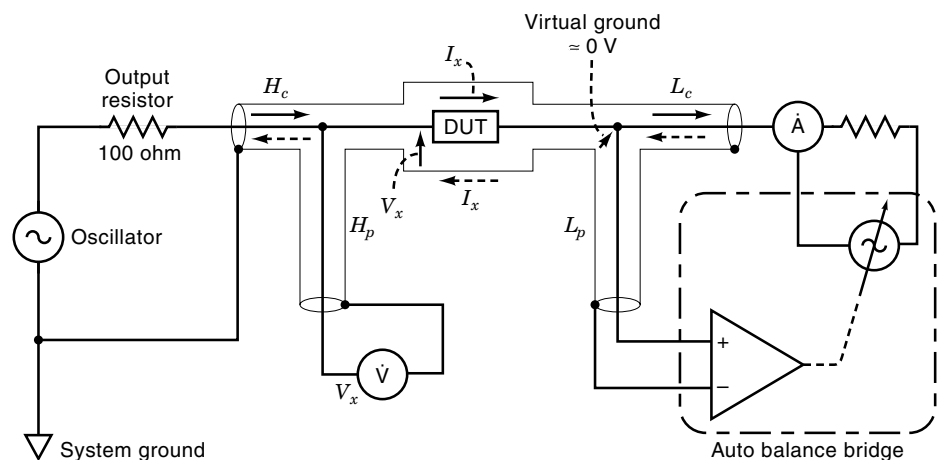


Figure 30. A schematic diagram of HP 4284A Precision *LCR* Meter (copyright Hewlett-Packard).

Ac Sources

From the balance conditions, it can be deduced that the first two requirements for an ac bridge source are the frequency stability and an output voltage sine waveform with very low distortion, especially if the balance condition is frequency dependent. The output power delivered by the ac bridge source may be insufficient, and the need for an additional power amplifier possibly associated with a matching transformer may perhaps be necessary. The addition of these devices will increase the total harmonic distortion of the source.

Modern *RC* audio-frequency sinewave generators usually present output voltages with a total harmonic distortion lower than 0.01% and generally offer the possibility to be phase-locked by a frequency synthesizer. Up to several kilohertz, as long as the total harmonic distortion of the output can be kept lower than 0.1%, it is also possible to work directly with a frequency synthesizer. Indeed, several of the commercially available frequency synthesizers in the audio frequency range have a frequency resolution higher than 1 part in 10^{10} . If necessary, the reference signal generated by the internal crystal oscillator may be replaced by the one produced by a higher stability oscillator like an atomic frequency standard. Digital signal generators may also be found useful due to their frequency and amplitude stability. However, the possible presence of glitches at the output of some of them may be a source of problems in the balancing process (24).

Ac Detectors

Passive detectors like the telephone receiver and the vibration galvanometer, which were very common at the beginning of ac bridge measurements, are no longer used. At present, nearly all detectors used in ac bridge measurements are fed from the mains supply or more rarely from batteries. The most important technical requirements for ac detectors are

- Amplitude sensitivity in current or in voltage
- Frequency selectivity and the ability to reject harmonics
- Phase sensitivity

It is to be noted that in ac, the experimenter can often use a matching amplifier between the network and the available detector.

The most common types of detectors are described below.

Wideband Detectors. Wideband detectors like oscilloscopes and ac millivoltmeters can be used. The wideband operation of these instruments can lead to the necessity of reducing the bandwidth by placing a band-pass filter or tuned preamplifier at the input of these detectors in order to achieve a better selectivity or signal-to-noise ratio.

Narrowband Detectors. Frequency-selective detectors, which are tuned to respond to the fundamental frequency of the in-phase and quadrature components of the out-of-balance signal, are also helpful. Their deflection will correspond to the sum-of-square of these components. Some commercially available detectors of this type also have the advantage of being battery operated. Nevertheless, their principle of operation can make the balancing of some bridges with poor convergence more difficult.

Phase-Sensitive Detectors. This type of detector is also known as a synchronous or lock-in detector. A phase-sensitive detector requires an auxiliary reference signal taken directly from the bridge source. It uses this signal to operate an electronic switch or a modulator. The most suitable detector of this type multiplies the incoming signal with the reference signal. The resulting signal is then passed through an integrating low-pass filter. Dual phase-sensitive detectors incorporated in the same instrument are also very convenient: one detector can be adjusted to respond to a signal with the same phase as the source, and the other to the quadrature component. In this way, the bridge convergence can be made more efficiently. This type of phase-sensitive detectors has the advantage of optimizing the recovery of the bridge out-of-balance signal from random noise. However, the direct connection from the source to the detector could be a source of problems if appropriate precautions, like the right use of an isolation transformer, are not taken.

Convergence of ac Bridges

The ability of a bridge to approach the balance is known as convergence. In ac bridges, this concept also includes the rate at which this balance is achieved. As seen previously, the balance condition is given by a system of two equations, which must be satisfied simultaneously. This indicates that in such bridges, at least two parameters, one reactive and one resistive, must be adjusted. The solution to this problem requires one to choose the adjusting devices and to determine the way to obtain the balance of the bridge in a minimum number of adjustments. As the best solution generally differs from one bridge to another, it is recommended to refer for this point to the bibliography (3,4).

Sensitivity of ac Bridges

As for the dc bridges we can define, at balance, the sensitivity of the detector as the ratio of the small deflection from zero resulting from a small change of the measured impedance to the value of the change. In this way, we can define the relative current sensitivity S_i as

$$S_i = \left| \frac{\frac{\Delta I}{\Delta Z}}{\frac{I}{Z}} \right|_{\Delta Z \rightarrow 0} \quad (88)$$

and the relative voltage sensitivity S_v as

$$S_v = \left| \frac{\frac{\Delta V}{\Delta Z}}{\frac{V}{Z}} \right|_{\Delta Z \rightarrow 0} \quad (89)$$

In theory, these values can be calculated from the parameters of the bridges including those of the detector (3,4). In practice, these parameters are easily measured by changing Z of a small amount from the balance and observing the small deflection ΔI or ΔV on the null detector.

Checking of ac Bridge Accuracy

Several principles explained for checking dc bridges could be applied for ac bridges if we take into account that we compare impedances and not pure resistors:

1. *Zero Voltage of the Bridge Source.* The same procedure as for dc bridges.

2. *Zero Impedance*. A suitable short circuit replaces also the device under test.
3. *Infinite Impedance*. An open circuit replaces the device under test. In this case, care must be taken in order to prevent the flow of capacitive currents.
4. *Checks of the Bridge Linearity*.

- *With Individual Resistors or Impedances*. Depending on the type of instrument, the substitution method or the transposition method can be used. These methods can be applied with all types of resistors or impedances. The used devices do not need to be calibrated. They must have only good short-term stability if the other influencing parameters can be kept constant.
- *With Resistance or Capacitance Networks*. Linearity checks of low frequency ac resistance bridges can be estimated using a string of equal value resistors connected in series. At present, a resistance network close to Hamon's design offers a convenient solution for the calibration of the ac resistance bridges from dc to 10^4 rad/s with an accuracy of about 1 part in 10^8 for 100 Ω (25).

Linearity checks of capacitance bridges can be executed by a set of capacitors successively connected in series and in parallel. Usually, the calibration of IVDs and voltage transformers can be made using an auxiliary transformer or set of capacitors (18).

ACKNOWLEDGMENTS

The author wishes to thank Mr. R. De Vré and Dr. T. J. Witt for the improvements brought to one part of this document. He is also grateful to his son Cedric and to Mr. G. Bairy for their patience in reviewing several drafts of the manuscript and to Mr. P. D'Hainaut for the realization of the figures.

BIBLIOGRAPHY

1. B. J. Pritchard and G. W. Small, Temperature coefficient variations in heat treated evanohm and their effect on the transient behavior of the NML 1 Ω resistors, *IEEE Trans. Instrum. Meas.*, **42**: 557–561, 1993.
2. F. Wenner, Methods, apparatus and procedures for the comparison of precision standard resistors, *J. Res. Nat. Bur. Stand.*, **25** (8): 230–294, 1940.
3. B. Hague, *Alternating Current Bridge Methods*, 6th ed., London: Pitman, 1971.
4. K. F. Harris, *Electrical Measurements*, New York: Wiley, 1952.
5. T. Sugiyama and K. Yamaguchi, Pulsewidth modulation dc potentiometer, *IEEE Trans. Instrum. Meas.*, **19**: 286–290, 1970.
6. L. C. A. Henderson, A new technique for the automated measurement of high value resistors, *J. Phys. E: Sci. Instrum.*, **20** (5): 492–495, 1987.
7. D. G. Jarret, Automated guarded bridge for calibration of multi-megohm standard resistors from 1 M Ω to 1 T Ω , *IEEE Trans. Instrum. Meas.*, **46**: 325–328, 1997.
8. I. Warshawsky, Multiple-bridge circuits for measurements of small changes in resistance, *Rev. Sci. Instrum.*, **26** (7): 711–715, 1955.
9. F. Delahaye, Dc and ac techniques for resistance and impedance measurements, *Metrologia*, **29** (2): 82–93, 1992.
10. N. L. Kusters, W. J. M. Moore, and P. N. Miljanic, A current comparator for the precision measurement of dc ratios, *IEEE Trans. Commun. Electron.*, **83**: 22–27, 1964.
11. W. J. M. Moore and P. N. Miljanic, *The Current Comparator*, London: Peter Peregrinus Ltd., 1988.
12. J. Williams, A. Matthews, and Y. Xu, Maintaining the primary standard of resistance using a cryogenic current comparator, *Cal Lab*, **3** (1): 12–16, 1996.
13. V. Polushkin, D. Drung, and H. Koch, A broadband picovoltmeter based on the direct current superconducting quantum interference device with additional positive feedback, *Rev. Sci. Instrum.*, **65** (9): 3005–3011, 1994.
14. H. W. Ott, *Noise Reduction Techniques in Electronic Systems*, 2nd ed., New York: Wiley, 1988.
15. A. Hamon, 1–100 Ω build-up resistor for the calibration of standard resistors, *J. Sci. Instrum.*, **31** (12): 450–453, 1954.
16. D. R. White and J. M. Williams, Resistance network for verifying the accuracy of resistance bridges, *IEEE Trans. Instrum. Meas.*, **46**: 329–332, 1997.
17. R. D. Cutkosky, Four-terminal-pair networks as precision admittance and impedance standards, *IEEE Trans. Commun. Electron.*, **83**: 19–22, 1964.
18. B. P. Kibble and G. H. Rayner, *Coaxial ac Bridges*, Bristol: Adam Hilger Ltd., 1984.
19. R. B. D. Knight, A precision bridge for resistance thermometry using single inductive current divider, *Euromas 77, IEE Conf. Pub.* **152**, 1977.
20. S. Avramov et al., Automatic inductive voltage divider bridge for operation from 10 Hz to 100 kHz, *IEEE Trans. Instrum. Meas.*, **42**: 131–135, 1993.
21. F. Delahaye and D. Bournaud, Accurate ac measurements of standard resistors, *IEEE Trans. Instrum. Meas.*, **42**: 287–291, 1993.
22. W. J. M. Moore and K. Ayukawa, A current comparator bridge for power measurement, *IEEE Trans. Instrum. Meas.*, **25**: 550–553, 1976.
23. HP 4284A Precision LCR Meter, Operation Manual: 6–9, Fig. 6-6.
24. D. Tarach and G. Trenkler, A noise-adaptive digital null detector, *IEEE Trans. Instrum. Meas.*, **46**: 435–438, 1997.
25. D. R. White et al., A simple resistance network for calibrating resistance bridges, *submitted for publication in IEEE Trans. Instrum. Meas.*

JACQUES NICOLAS
Belgian Ministry of Economic Affairs

OSCILLOSCOPES

The oscilloscope, or *scope*, is an important tool in engineering. The oscilloscope displays a graph of an electrical signal on a screen. In its basic operational mode, the oscilloscope graphs instantaneous signal values versus time. Displayed waveforms have three basic components: Y, the vertical or amplitude component; X, the horizontal or time component; Z, the signal brightness or intensity. Single-channel and multichannel oscilloscopes are available. The most popular is a two-channel instrument in which a single trace display with channel multiplexing is available. In this case not only instantaneous signal values versus time, but also one amplitude waveform component versus the other, can be observed.

Oscilloscopes are used in physics, electronics, instrumentation and measurements, communications, control, mechanical

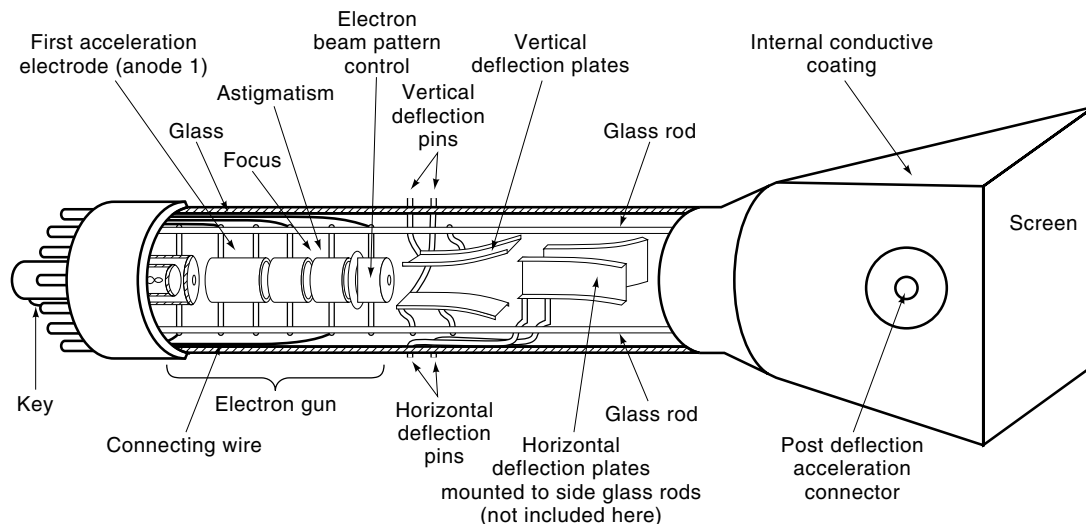


Figure 1. Internal structure of a CRT.

engineering, and many other areas where nonelectrical signals can be converted into electrical signals. The graphical display allows a user to determine many signal parameters, such as:

- Voltage and time values
- Frequency, period, and rise and fall times
- Phase shift
- Noise level, jitter
- Frequency ratio of two signals
- Calculated Fourier spectrum
- Calculated statistical parameters
- Calculated mathematical functions of signals, such as the integral, derivative, sum, difference, and product

Two major types of oscilloscopes have emerged over the last decade: analog and digital. Within these two categories, a number of operational modes can be distinguished. There are real-time data acquisition and display scopes, real-time acquisition and equivalent- or transformed-time display scopes, equivalent-time acquisition and display scopes, and random data acquisition and random display scopes. A typical analog scope directly acquires signals and displays them on the screen in real time. An analog sampling scope can operate in real time at low signal frequencies, and in equivalent time at high frequencies. A digital scope can operate in real-time or equivalent-time mode to acquire signals, but the display shows the reconstructed signals in the equivalent-time mode, which allows for great reduction of the display unit speed, which in turn achieves great reduction of the scope's cost. Currently, the major oscilloscope producers, such as Hewlett-Packard, Tektronix, and LeCroy, do not list analog scopes in their catalogs, but many analog oscilloscopes have been used and will be used in the near future. The digital scopes, also called digitizing or digital storage oscilloscopes, are considerably more powerful and flexible for signal inspection and analysis. Signal digitization has opened unlimited possibilities for signal storage and processing.

ANALOG SCOPES

Analog scopes without sampling acquire and display waves in real time. The most critical part of the analog oscilloscope is a cathode ray tube (CRT), whose maximum speed of wave tracing limits the speed of the entire scope. Figure 1 shows the simplified structure of a CRT. The beam of electrons generated by the cathode is directed towards the fluorescent screen by a set of electrodes. The beam makes a bright spot on the screen at a position controlled by the voltages applied to the vertical and horizontal deflection plates. The tube can have a single or a double beam system. The latter has two separate beam-forming and deflection structures. The tube shown in Fig. 1 employs a two-stage electron acceleration process. The electrons emitted from the cathode are formed into the beam, whose density is controlled by the cylindrical electrode. Later, the beam is focused and deflected, passing through the vertical and horizontal deflection plates. Final acceleration, called postdeflection acceleration (which is necessary to produce a visible spot on the screen), is achieved by applying very high voltage, at least several kilovolts, to a distributed electrode located inside the tube between the deflection plates and the screen. The inner surface of the tube screen is covered with a *phosphor*, which emits visual radiation on being bombarded by electrons. The length of time during which the phosphor emits radiation until its level decays to 10% of the initial value is called the *persistence*. Several types of phosphor inner coatings are applied in CRTs. The most popular phosphor, P 31, exhibiting high luminance and reasonable persistence of about 32 ms (1,2), is found in most CRTs.

In analog scopes, an image of a nonrepetitive signal fades quickly or slowly, depending upon the tube persistence. Fast single-shot events may not be registered by a classical CRT whose screen directly reacts to the electron current delivered to it. In order to enhance the quality of writing, microchannel plate CRTs are used in some analog scopes (3,4). Inside the tube, an array of hollow glass fibers is located. The fibers, internally coated with a semiconductor, multiply the number of electrons through secondary emission when activated by an adequately accelerated electron beam. The beam deflection can be

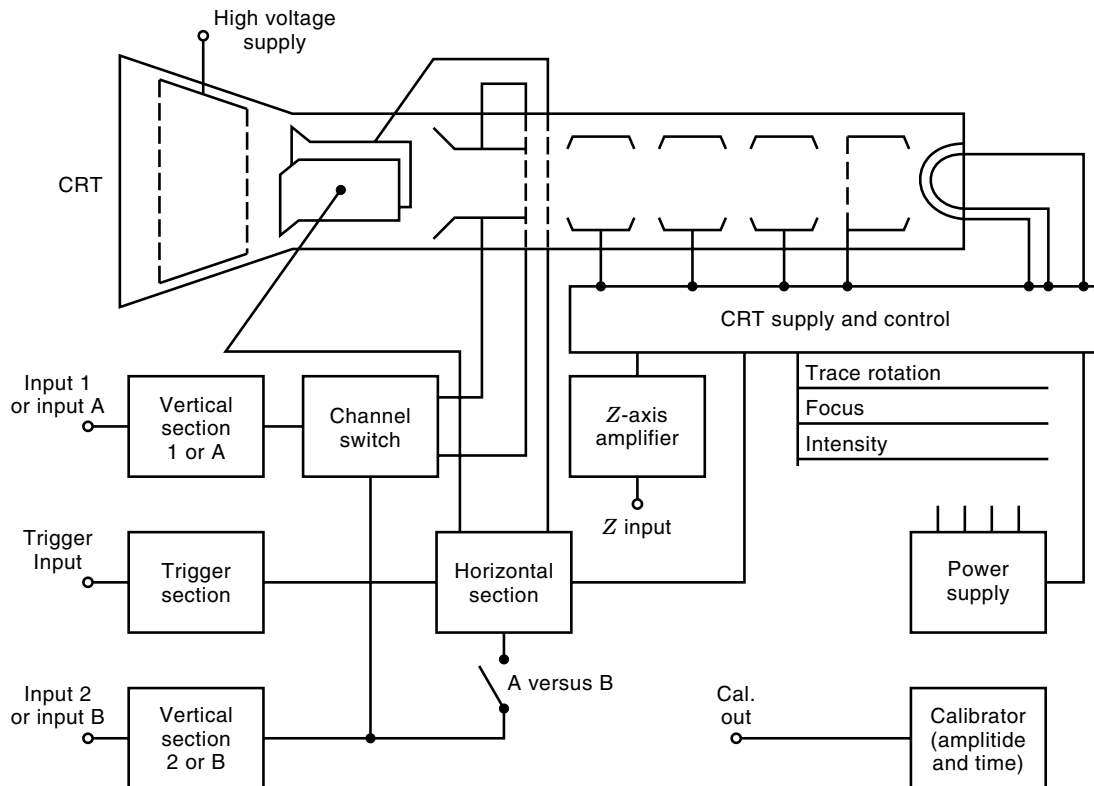


Figure 2. General diagram of an analog oscilloscope.

affected by interfering magnetic and electric signals produced within the oscilloscopes by the proximity of power supplies, especially transformers, chokes, and high voltage. To avoid interference, the tubes and potentially interfering components are well shielded.

The scheme of a typical analog scope is shown in Fig. 2. The vertical channels, A and B, include input coupling circuits, attenuators, preamplifiers, delay line, and differential output amplifiers driving the vertical plates of the CRT, as shown in Fig. 3. Quite often, voltage or current probes are connected to the inputs to provide better interfacing with the tested circuits. The horizontal section of the oscilloscope is composed of a triggering unit with its coupling, trigger shaping unit, trigger level control, and trigger holdoff units; a sweep generator, which defines the main time base of the scope; and a differential output amplifier driving the horizontal deflection plates of the tube (Fig. 4). The main sweep generator can be supported by a fast time-base unit triggered at a preselected level to expand desired parts of the tested waves. Figure 5 illustrates typical waveforms of analog scopes. Two time bases are activated to show distinguished sections of the signals under test with the help of the fast ramp, which is called the delayed time base. The waves observed on the tube screen are delayed in relation to the input signals due to the action of the delay lines forming parts of the vertical amplifier. In this way the trigger and time-base circuit delays and nonlinearities are compensated to create good conditions for tests of the wave leading edges without distortions. Figure 5 also shows internally generated triggering pulses initiated by the input signal of channel 1 or 2.

To stabilize the horizontal wave position on the screen, the trigger level is selected within the frame of the vertical span

of the incoming waves. At the same time an auxiliary system called the *holdoff* has to be adjusted to make the time-base generator insensitive to excessively frequent triggering. If more than one channel is used and a single-trace tube is applied, then a channel switch can direct only one channel to the screen at a time for the duration of the time base at high rates of the time base (alternating mode), or small segments of the individual channel signals can be displayed one by one at lower rates of the time base (chopped mode). The waves for both modes of the display are shown in Fig. 6. Multichannel data acquisition allows one to perform basic hardware arithmetic operations on signals, such as addition and subtraction. The process of subtraction, which can disclose a differential signal component, is corrupted by interchannel interference due to the common ground pathways and nonideal channel isolation of the channel switch. Many analog oscilloscopes have internal amplitude calibrators, which help check the scope amplitude calibration and verify or adjust the compensation of the voltage probes.

ANALOG SAMPLING SCOPES

Typical real-time analog oscilloscopes have operating bandwidths up to several hundred megahertz. The limits are mainly imposed by the CRT's writing speed. One of the means used to reduce the speed of observed wave has been time transformation by means of coherent sampling. The method can only be applied to repetitive waveforms. A basic set of waveforms for a sampling scope is shown in Fig. 7; the scope block diagram is depicted in Fig. 8. The signal under test applied to the input is delayed by a broadband transmission line, then sampled

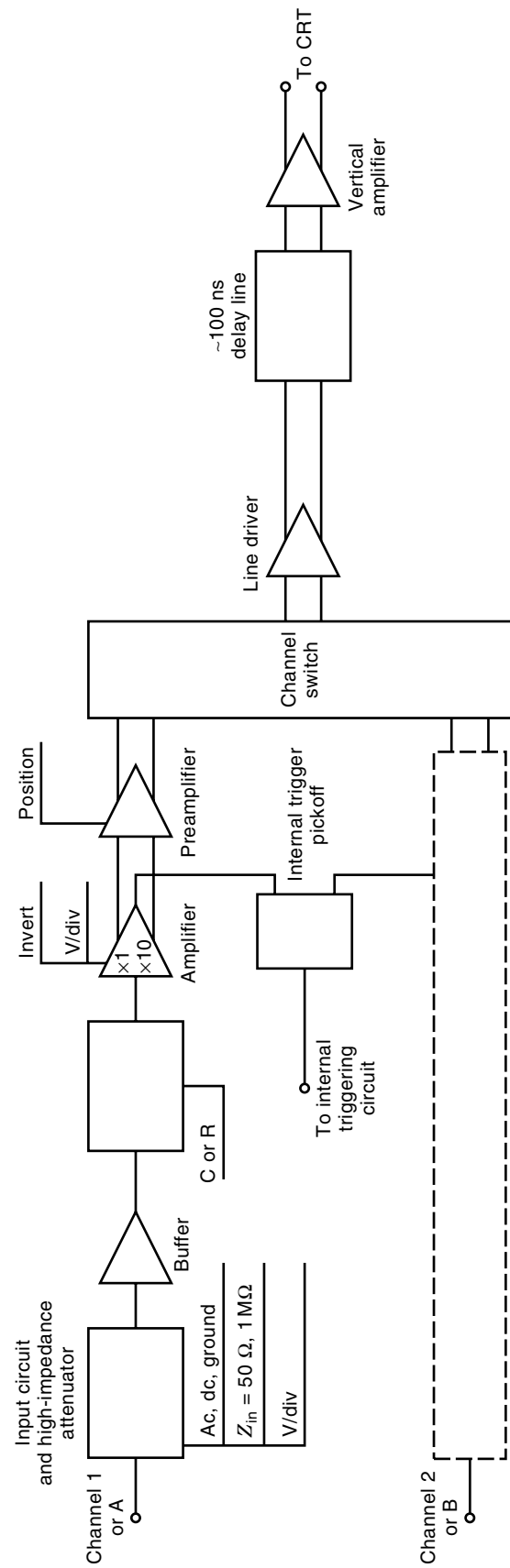


Figure 3. Vertical channel of an analog oscilloscope.

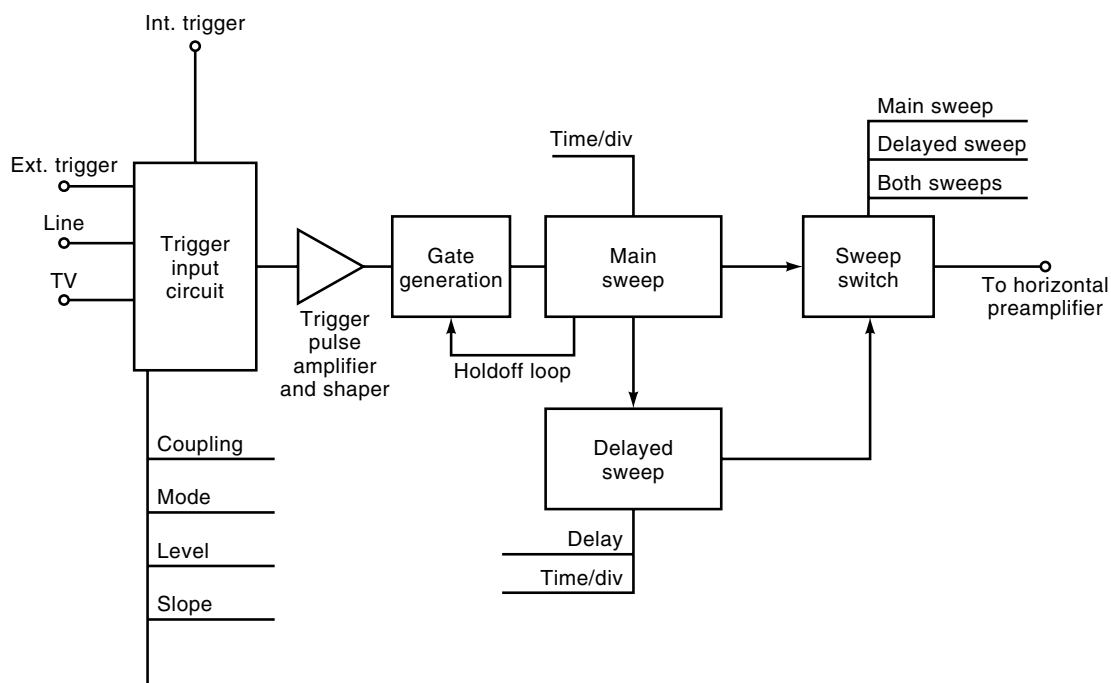


Figure 4. Horizontal channel of an analog oscilloscope.

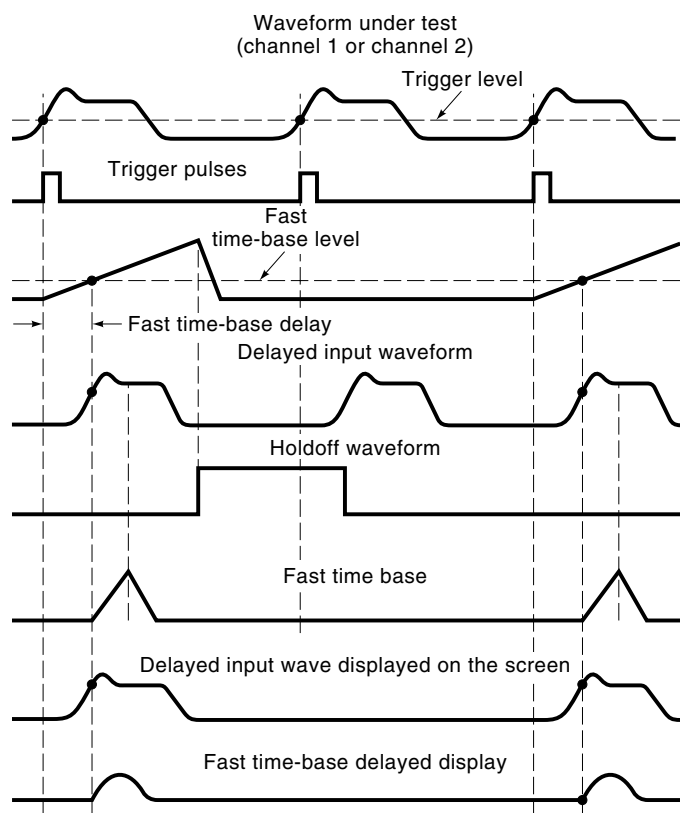


Figure 5. Waveforms in an analog oscilloscope.

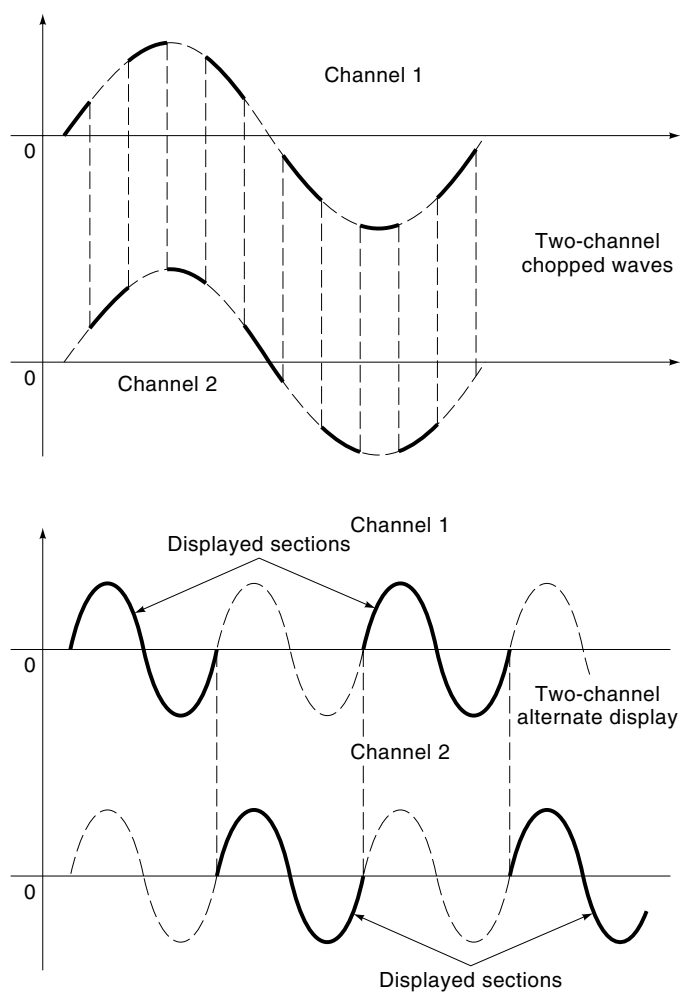


Figure 6. Modes of display of an analog oscilloscope.

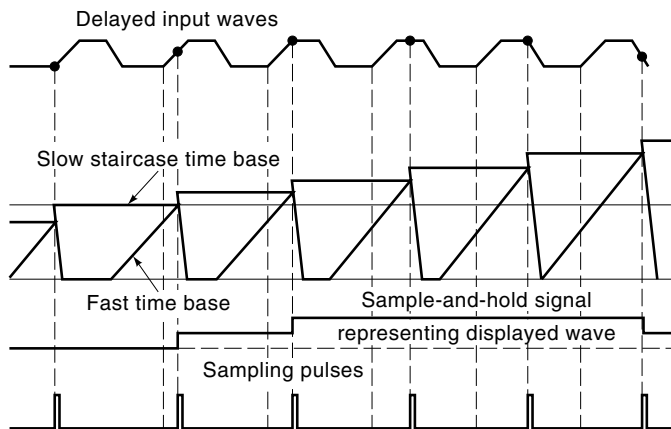


Figure 7. Time transformation in an analog sampling scope.

by very narrow pulses in a sampling gate. If the repetition frequency is less than several hundred kilohertz, which is the maximum repetition frequency of the trigger circuitry, a single sample is taken from each pulse of a frame defined by the time base. For higher repetition rates the trigger circuits divide the frequency so that every n th period one sample is taken from a single pulse of a repetitive wave. When the samples are taken, they are stretched, amplified, and memorized in a sample-and-hold circuit. Later, the samples are applied to low-frequency vertical amplifiers driving the tube. The sampling scope system uses a track-and-hold feature that enables the scope to process very small samples formed by the differences between the pairs of consecutive samples. Generation of subsequent sampling pulses is achieved by comparison of two fully synchronized time bases. One of them is a fast time base, and the other is a staircase signal. The horizontal channel provides the timing of all samples. It also delivers trigger signals to the sampling pulse generator and to the time bases. Analog sampling oscilloscopes are often applied in time-domain reflectometers (TDRs), and they can operate covering an equivalent

bandwidth up to 50 GHz, which corresponds to 7 ps for the rise-time without the delay line. The delay lines available for such scopes have their bandwidth limited to 5 GHz.

DIGITAL SCOPES

Digital oscilloscopes overcome many disadvantages of analog oscilloscopes, such as the necessity for delay lines and for fast and long-persistence CRTs. A digital scope display can be adjusted after wave acquisition. The input circuitry of a typical digital scope does not differ from that of analog scopes. After preliminary attenuation or amplification, analog signals are sampled and digitized. The digitized samples are stored in a memory. The tube receives the reconstructed analog samples of vertical and horizontal information through digital-to-analog converters. The process of data display is delayed and stretched in time, so the tube that displays the samples can operate much slower than the fast data-acquisition parts. Three major types of digitizers are applied in digital scopes:

- Real-time digitizers operating at high sampling and conversion rates can capture nonperiodic short transient processes.
- Coherent sampling digitizers work in the same way as the sampling oscilloscopes with time transformation described above. The analog-to-digital conversion is applied to much slower waves after the sampling and time transformation. In this approach, a pretrigger pulse is required; it is obtained either from the tested signal before the delay line and the sampler, or from a separate source whose output wave follows the trigger.
- Random interleaved sampling digitizers sample repetitive signals in a pseudorandom manner. The amplitude and timing data are digitized and memorized in corresponding memories.

All three digitizers can equally well acquire periodic signals. In the last two approaches, only a single sample can be taken

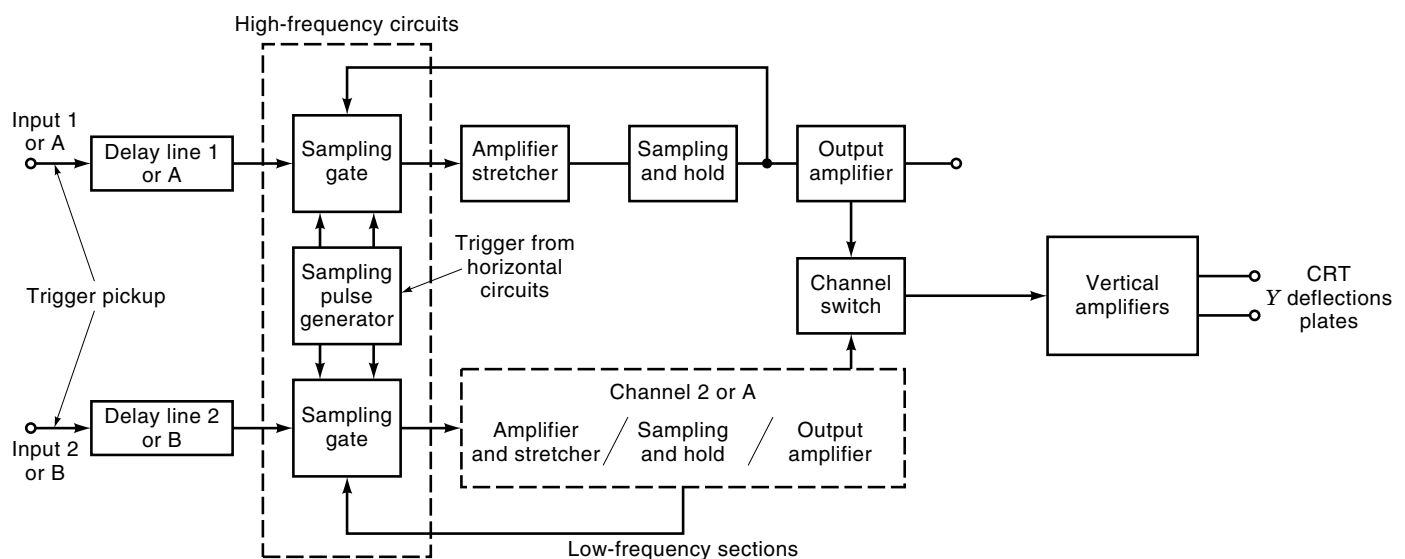


Figure 8. Vertical system of a coherent sampling oscilloscope.

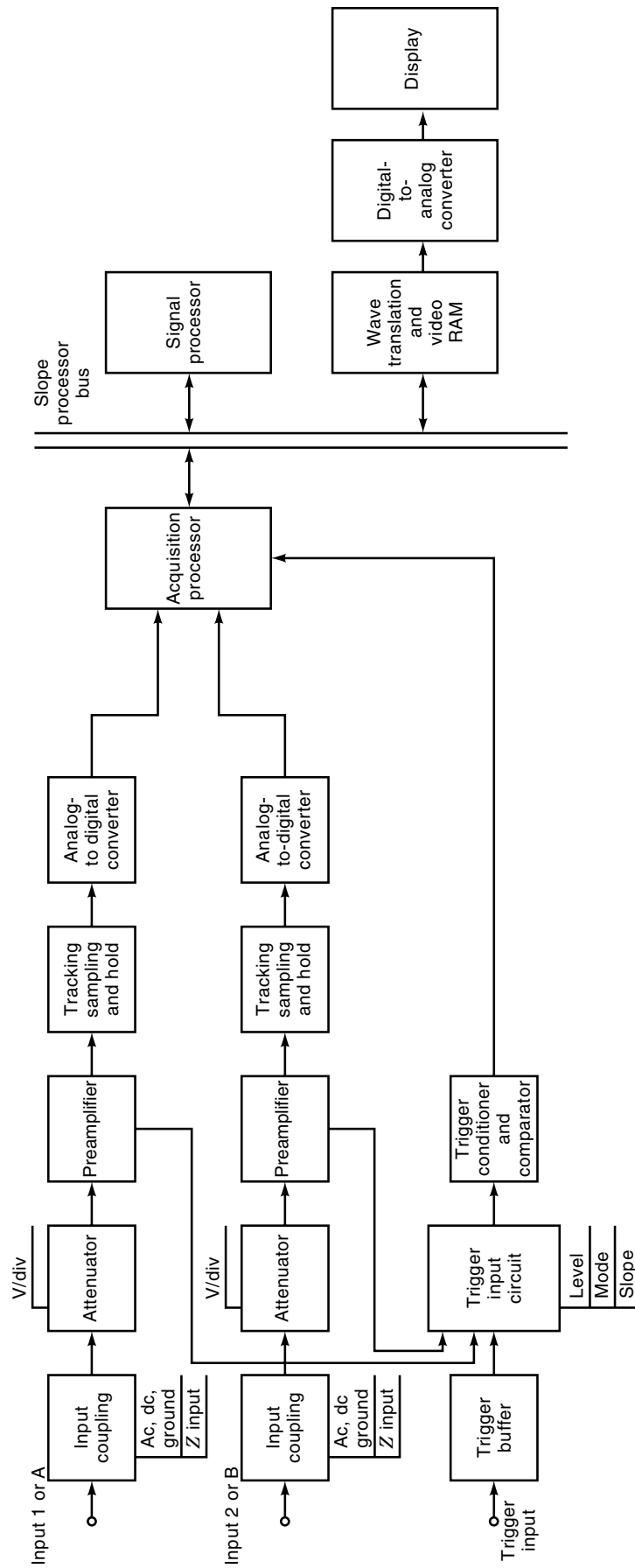


Figure 9. General block diagram of a digital oscilloscope.

during the input signal cycle, especially at low frequency of the input signal. Only real-time digitizers can acquire single short transient events. In some scopes, such as the HP 54600A, the random interleaved sampling digitizer can capture low-speed single events. The real-time signal capture capability of most digital scopes is limited by the maximum sampling rate, which is directly associated with the maximum speed of the analog-to-digital converters. The inherent storage capability of digital scopes, especially the possibility of storing long records of several channels, allows one to process the signals with complex mathematical operations, including statistical functions (5–7). When the oscilloscope calculates the fast Fourier transform (FFT) of the input signal, the frequency spectrum of the signal can be displayed versus frequency separately from the time-domain signal image. A typical block diagram of a two-channel digital storage scope (DSO) is shown in Fig. 9. The channels acquire and digitize the signals independently, and then the digital signals are accumulated and processed according to the chosen function. In the simplest case, the digitized signals are retrieved from the memory, converted back into their analog forms by the digital-to-analog converters, and displayed. The reduced speed of the signal display processes eliminates the need for high-frequency CRTs. Some scope displays utilize liquid crystal display (LCD) tubes, which are directly controlled by addressing the picture elements (pixels) of a display matrix. Analog scopes of this kind can update their pictures about 10^5 times per second. Time changes in the signals are then recorded immediately. The display of a typical digital storage scope is usually a memory reconstruction of the acquired samples taken from many periods, so the entire image can be acquired only 5 to 100 times per second. High quality of the picture display of analog scopes requires rather high frequency of image updates. Such update frequencies are not available when the signal is not repeated frequently. Some digital storage scopes with high-speed digitizers and processors, and with smart display algorithms, update their displays by creating high-quality *analog persistence* images over a broad range of signal frequencies (6). The number of samples displayed on the screen is fixed independently of the time-base settings, which leads to changes of the sampling rates for different time bases. In order to reproduce periodic waves, taking five to ten samples per period of the highest harmonic is regarded as the minimum for sufficient signal detail. The Nyquist sampling rate is too low to reconstruct periodic signals in all cases, especially when the signals are sampled near the zero crossings. Major features of the digital storage scopes can be summarized as follows:

- Large amounts of pretrigger information can be acquired without using delay lines.
- Single or repetitive waves can be stored permanently.
- Computer or printer interfacing to store or print more data can be easily achieved.
- One may reliably capture unpredictable events.
- One may reliably capture glitches shorter than the resolution time by means of peak detectors.
- Within the memory capacity, there is no dead time between acquired events.
- Measurements can be automated.

- Currently acquired waves can be compared with reference waves.
- Wave tolerance tests are possible, including go–no-go tests.
- Acquired waves can be processed mathematically, including statistical operations.
- Finite resolution and a small number of display sweeps per second can mask the time distribution of waveforms.
- The limited number of samples and the constant sample intensity commonly used in displaying simple repetitive waves can distort composite repetitive signals, including waves of various speeds.

TECHNICAL PARAMETERS AND LIMITATIONS

This section describes basic parameters of the oscilloscopes to help select the right one for an application. In the selection process, the scope parameters must be related to the waves under test. The major specifications of analog scopes are number of channels, bandwidth, voltage sensitivity, maximum time-base speed, amplitude noise, time jitter, and overall accuracy. From these parameters other scope features can be derived. For digital scopes, the basic parameters also include maximum sampling rate, vertical and horizontal resolution, and memory length.

Number of Channels

Common-purpose single-channel scopes are not very popular in measurement laboratories. If the number of channels is greater than one, the interchannel interference becomes important and a high common-mode rejection ratio (CMRR) is desired. For example, if $\text{CMRR} = 40 \text{ dB}$, then with one channel receiving a signal of 1 V and the other channel receiving a signal of 10 mV, the lower-amplitude channel will measure the amplitude with an error of about 10%, which is well above scope error specifications.

Bandwidth

The bandwidth (BW) is usually defined as the frequency range between dc and the frequency at which a sinusoidal signal amplitude observed on the screen is reduced 3 dB in relation to the value at dc or low frequencies. The input sinusoidal signal is supplied from a source having very low internal resistance, perhaps 50Ω . The bandwidth of an analog scope is primarily determined by the frequency response parameters of its vertical amplification channel, including attenuators and the tube. The magnitude response error introduced by the frequency limitations is shown in Table 1. The frequency response bandwidth relates to an equivalent time-domain parameter, namely, the rise-time, usually considered as the time the step response changes between 10% and 90% of the steady-state level. For the simplest frequency response approximation, the relationship between

Table 1.

Signal Freq. for -3 dB BW (MHz)	Error (%)
0.1	1
0.2	2
0.5	11
1.0	29

Table 2.

T_R^a (ns)	T_T (ns)	T_0 (ns)	Error (%)
3.5	1	3.6	260
3.5	2	4	100
3.5	5	6.1	22
3.5	10	10.6	6
3.5	20	20.3	1.5

^a BW 100 MHz.

the risetime and the bandwidth is as follows:

$$T_R = \frac{0.35}{BW} \quad (1)$$

For instance, for a 100 MHz scope the response time is 3.5 ns. This leads to time-domain measurement error, whose effects can be corrected for using the following expression:

$$T_T = \sqrt{T_0^2 - T_R^2} \quad (2)$$

where

T_T = actual risetime

T_0 = measured risetime

T_R = scope's own risetime

Table 2 shows the time-domain error calculated without correction.

Both tables suggest that in order to maintain reasonable measurement accuracy, the scope should be at least several times faster than the fastest tested wave. Frequency-domain or time-domain speed limitations of digital scopes are additionally imposed by the sampler sampling speed and the analog-to-digital (ADC) converter speed. The sampler includes the sample-and-hold unit, whose sampling pulse width determines the fastest change that can be captured and later converted in the ADC. Faster digital scopes using flash or parallel analog-to-digital converters have their sample and hold function performed by the sample clock after conversion. In this way,

the changes of the input signal continuously translated into the digital representation do not affect the sample position on the scope screen, so the sample is observed as a dot. With ideal sampling pulses, the sampling rate should be several times larger than the expected maximum frequency of the tested signals. The sampling rate also determines the scope horizontal resolution.

Vertical Sensitivity

This parameter expresses the scope's ability to accommodate small signals as well as large signals within the same space on the screen. Typical ranges of scope sensitivities are between several millivolts per division and 20 V per division. The sensitivity range is a compromise between the user's desired values and the technical limitations imposed by the noise level of the scope amplifiers and the maximum available attenuation of high-frequency compensated attenuators.

Amplitude Noise Level

The scope noise mainly includes the broadband noise generated by the input stages of the scope and by the source resistance. The observed noise levels are usually on the order of 1 mV. Bandwidth reduction decreases the noise level and allows one to increase the scope sensitivity. In digital scopes, the input analog amplifiers introduce their noise, thus degrading the resolution of the ADC conversion process. For example, 1 mV noise on a 1 V signal limits the resolution of the conversion to less than 12 bits.

Time Jitter

Time jitter is caused by threshold instabilities in triggering and time-base circuits. The jitter is observed on the screen as a spread of vertical parts of the tested signals. In digital scopes, the clock jitter is also added. The levels of jitter are on the order of picoseconds for fast sampling scopes, and on the order of nanoseconds for low-frequency scopes.

Accuracy

The inaccuracy of a scope comprises mainly vertical and horizontal calibration and reading errors. The reading errors of digital scopes, with built-in digital amplitude and time meters, are negligible in comparison with calibration errors of amplifiers, voltage attenuators, and input resistors. The total amplitude error can reach several percent, but the dc voltage error can be as low as one percent. The time-base errors in digital scopes are comparable with the errors of digital

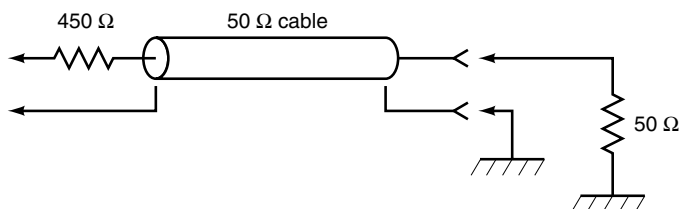


Figure 10. Low-impedance passive probe.

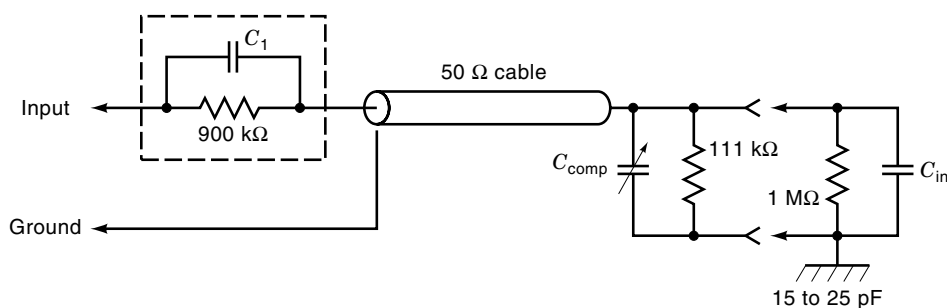


Figure 11. Passive probe, 1 MΩ input resistance.

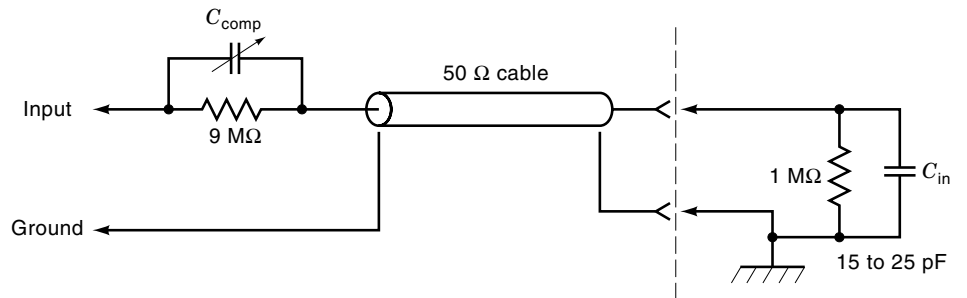


Figure 12. Passive probe, $10\text{ M}\Omega$ input resistance.

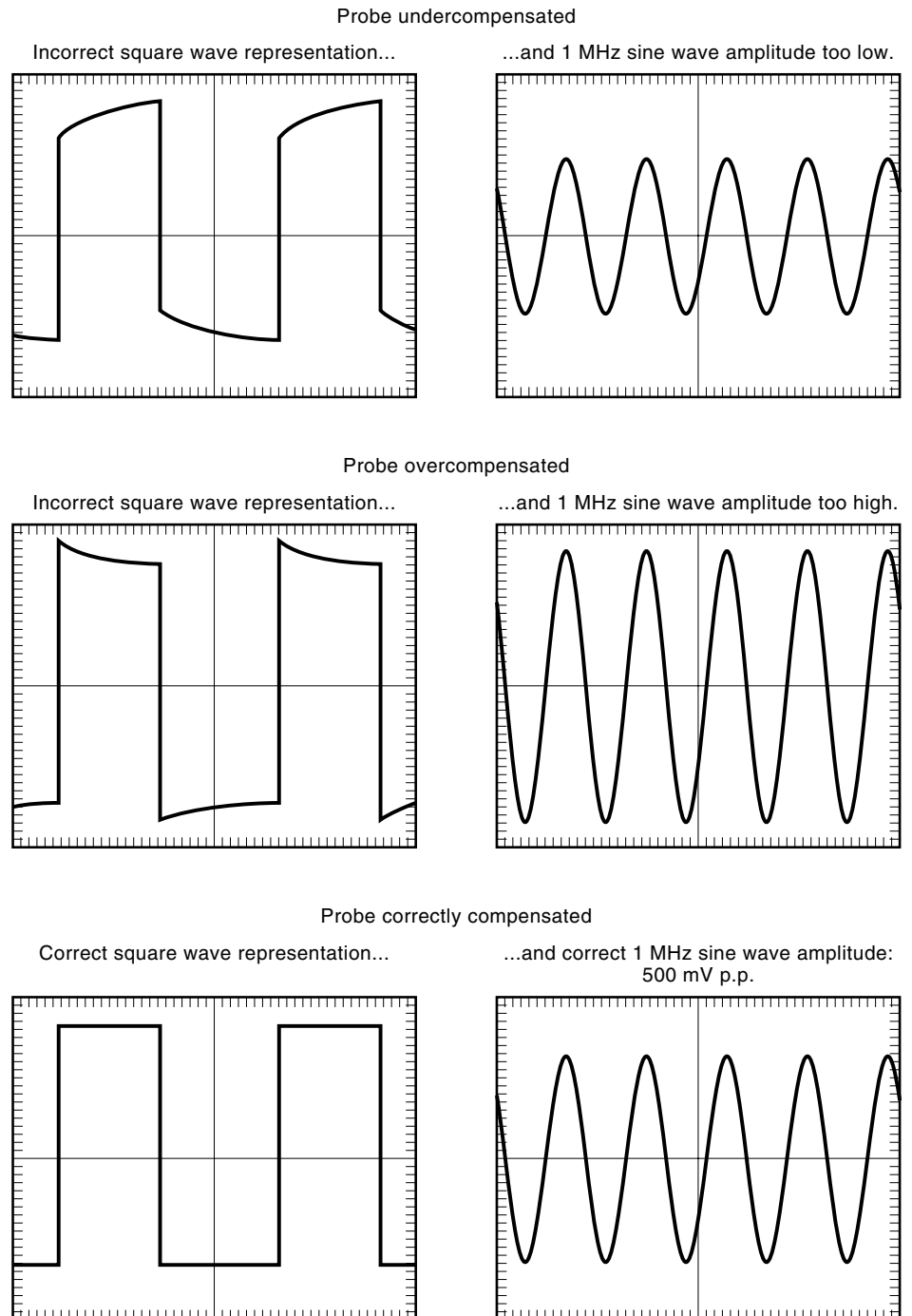


Figure 13. Compensated probe waveforms.

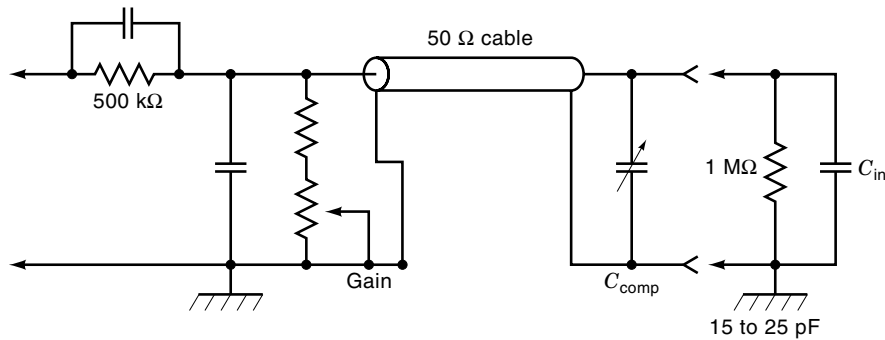


Figure 14. High-input-resistance high-voltage probe.

counters, which are below 0.1% (7), while the time base errors in analog scopes are usually 2% to 5%. The process of selecting a scope is primarily determined by information about the waves under test: their shapes, repetitivity, frequency and amplitude, drift, and speed. Additional conditions that should be considered are the length of the record, possibility of test automation, signal statistics, and wave storage. The scope's capabilities and parameters should be also chosen with regard to future applications and affordability.

OSCILLOSCOPE PROBES

Almost all standard oscilloscopes use different voltage probes to connect to a circuit or a device under test, chosen to minimize the effects of the scope input impedance and electromagnetic interference. Some probes convert a nonelectrical quantity or another electrical quantity, such as current, into a voltage, which is measured by the scope. Two major types of voltage probes are used. The first type is a passive probe, which yields high input resistance and low input capacitance but introduces significant signal attenuation, usually ten times. The second type is an active probe, which does not attenuate the signal but provides high input resistance and low capacitance. Unfortunately, the supply voltage required for active probe devices also introduces undesired noise and offset. Active probes usually employ field effect transistors and operational amplifiers (8). A typical voltage probe has two terminals: a signal terminal and a ground terminal. In many cases, tests of differential signals cannot be done by means of two-channel differential measurement, due to the limitations of the CMRR. For these tests, differential active probes, which have very high CMRR, are recommended.

Passive Probes

Passive probes include low-impedance resistive voltage-divider probes and compensated high-impedance voltage-divider probes. There are also high-impedance probes with

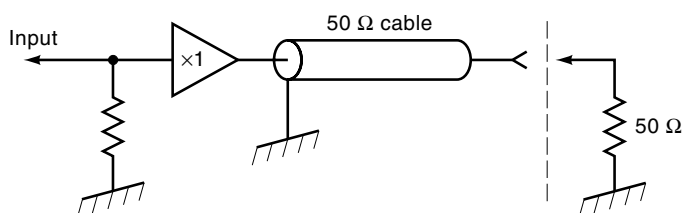


Figure 15. Diagram of an active probe.

a very large voltage division factor, which are used in high-voltage measurements. Figure 10 shows a low-impedance passive probe. This type of probe has very low capacitance loading and wide bandwidth. It can be used in testing low-impedance circuits. High-impedance passive probes are shown in Figs. 11 and 12. In both cases, the probes require compensation before the measurement processes start. The compensation is performed by observing a standard square wave signal generated by the scope calibrator. The effects of the pulse compensation on the square wave and the sine wave are shown in Fig. 13. High-voltage probes have similar structure to that in Fig. 13. The series resistor is of much higher value and larger physical size, to withstand voltages higher than 500 V. A high-voltage probe circuit is shown in Fig. 14.

The parameters of passive probes are

- Voltage attenuation (e.g., 10× or 1×)
- Bandwidth for 50 Ω source resistance
- Maximum input voltage (e.g., 500 V)
- Input resistance (e.g., 1 MΩ)
- Input capacitance (e.g., 2 pF)

Active Probes

Figure 15 is a diagram of an active probe with a single input. A differential probe is shown in Fig. 16. Both probes have very large input resistance, broad band, and no attenuation. The differential probe also has high CMRR, greater than 60 dB.

OSCILLOSCOPE MEASUREMENTS

Scopes can measure the following basic quantities:

- Voltage
- Current (with a current probe)
- Frequency and time, including period, time delay, pulse width, and duty factor
- Phase shift
- Rise and fall times
- Amplitude modulation index
- Jitter

Voltage

Voltage measurements are elementary tests done with oscilloscopes. Dc voltages are measured by observing the level shift

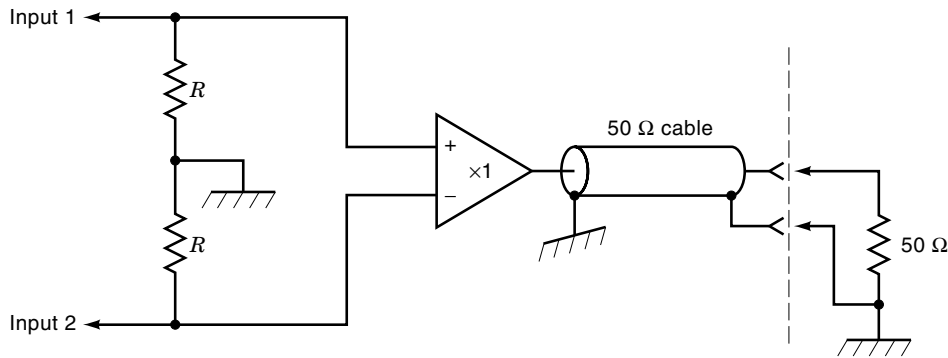


Figure 16. Differential probe.

of the horizontal line on the screen. The screen graticule and the vertical sensitivity setting indicate the amount of the shift. Some scopes have internal digital voltmeters indicating numerical values of the beam position and the positions of the different markers. Ac voltages, in older analog scopes, are measured in terms of the voltage differences, which are calculated with the aid of the graticule. In modern scopes, the marked levels are measured digitally and displayed numerically on the screen.

Current

Current measurements require the use of current-to-voltage converting devices, such as Hall effect sensors for dc currents, current transformer clamping sensors for ac currents, or combinations of both. The Hall effect current sensors are supplied from the scope or from a separate power supply, which also incorporates additional amplifiers.

Frequency and Time

With older scopes, the time-domain parameters can be measured using the screen reading in relation to the current time-base setting. Modern scopes, especially digital, have digital readout and display the time-domain parameters with reference markers. Rise and fall times are measured with the help of horizontal levels marked at 10% and 90%. Figure 17 illustrates the way the risetime is evaluated.

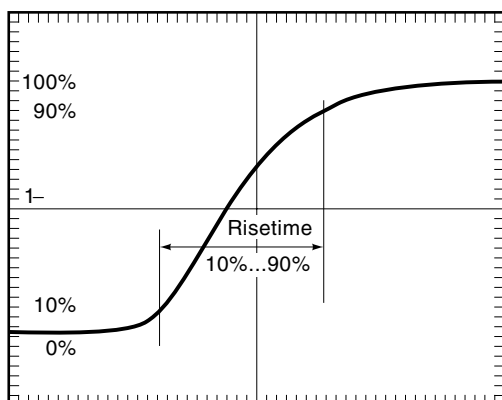


Figure 17. Measurement of the risetime.

Phase Shift

The phase shift between two waves is the amount of time, expressed in degrees, that passes between the beginnings or other selected points of the two waves. The phase shift is thus obtained from two time measurements in which the waves' relative position and period are found, followed by computations.

Another method, which can be used only for coherent sinusoidal signals, involves the XY display mode. One signal is applied to the horizontal channel, the other to the vertical channel. The XY figures displayed on the screen are called Lissajous patterns, and from the shape of the pattern the phase shift and frequency can be determined. Figure 18 shows different Lissajous patterns.

Amplitude Modulation Index

Figure 19 shows a time-domain AM wave displayed on the screen while the external triggering of the scope is synchronized by means of a modulated lower-frequency wave. The modulation index is calculated as a ratio of M to C . It is also possible to display an XY pattern in which the modulated wave is

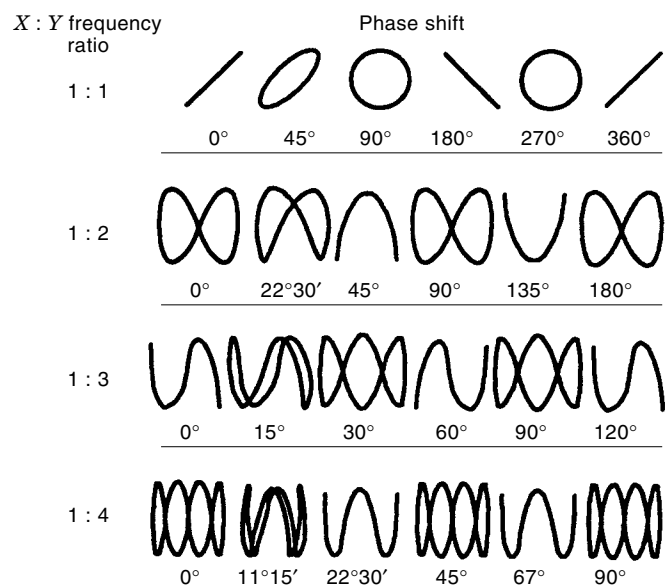


Figure 18. Phase measurement applying XY display.

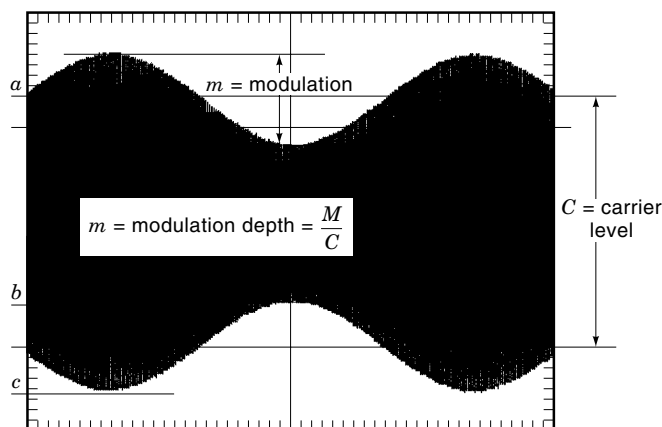


Figure 19. Time-domain measurement of the AM modulation depth (index).

applied to the scope vertical system, and the modulating signal is applied to the horizontal system. Figure 20 shows a typical AM wave pattern for this method.

Jitter

Jitter measurements are based on repeated readings of the time between the two repeated signal edges. The oscilloscope trigger jitter contributes to the jitter value observed on the screen. Analog scopes, even with long persistence, do not display repetitive waves with jitter in such a way that meaningful jitter evaluation is possible. Digital storage scopes with computer processors and statistical packages display waves with their intensity or colors related to the number of repetitions, thus revealing the frequency of occurrence of different cycles of the same wave. Some scopes (3) can display histograms computed from the samples. Statistical parameters, such as the

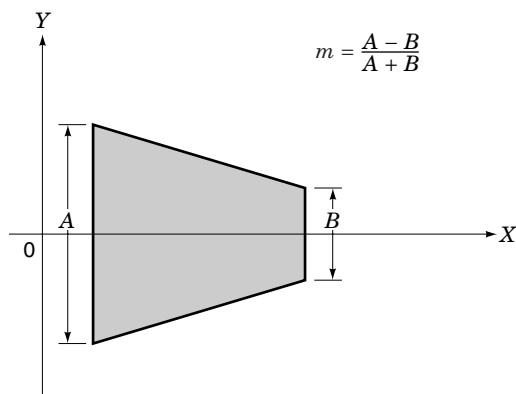


Figure 20. Measurement of the modulation depth (index) applying XY display.

standard deviation of the sample histogram or other time-domain and amplitude-domain statistical measures, can be computed.

PROGRAMMABILITY OF OSCILLOSCOPES

Most modern digital scopes can be controlled by computers. The computer provides more flexibility in storing data and programming experiments. Advanced interfacing techniques involve applications of GPIB (GPIB, IEEE-488) cards and connections controlled by the dedicated software, like HPVVEE of Hewlett Packard, or LAB VIEW of National Instruments (9,10).

BIBLIOGRAPHY

1. D. A. Helfrick and W. D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1990.
2. H. M. Berlin and F. C. Getz, Jr., *Principles of Electronic Instrumentation and Measurement*, Columbus, OH: Merrill, 1988.
3. *TDS 210 and TDS 220 Digital Real-Time Oscilloscopes*, 070-8483-02, Beaverton, OR: Tektronix, 1997.
4. R. Bristol, G. Fladstol, and C. Brannon, Oscilloscopes: How to find the one that fits, *IEEE Spectrum*, **24** (6): 34–39, June 1987.
5. *An Introduction to Digital Storage*, Beaverton, OR: Tektronix, 1987.
6. *Test & Measurement Product Catalog*, Chestnut Ridge, NY: LeCroy, 1997.
7. *Basic Instruments Tutorials*, Hewlett-Packard, [online]. Available [www: http://www.tmo.external.hp.com](http://www.tmo.external.hp.com)
8. K. Rush, W. H. Escovitz, and A. S. Berger, High-performance probe system for a 1-GHz digitizing oscilloscope, *Hewlett-Packard J.*, **37** (4): 11–19, April 1986.
9. R. Helsel, *Visual Programming with HPVVEE*, Upper Saddle River, NJ: Prentice-Hall, 1997.
10. *Instrumentation Catalog*, Austin, TX: National Instruments, 1998.
11. C. F. Coombs, Jr. (ed.), *Electronic Instruments Handbook*, 2nd ed., New York: McGraw-Hill, 1995.

Reading List

- ABC's of Oscilloscopes*, Everett, WA: Fluke.
- ABC's of Probes*, Beaverton, OR: Tektronix, 1995.
- HP 54600-Series Oscilloscopes, User and Service Guide*, Hewlett-Packard, 1992.
- XYZs of Analog and Digital Oscilloscopes*, Beaverton, OR: Tektronix, February 1993.
- 1740A Oscilloscope Operating and Service Manual*, Beaverton, OR: 1977.
- M. S. Holcomb and D. P. Timm, A high-throughput acquisition architecture for a 100-MHz digitizing oscilloscope, *Hewlett-Packard J.*, **43** (1): 11–20, February 1992.
- K. Rush and D. J. Oldfield, A data acquisition system for a 1-GHz digitizing oscilloscope, *Hewlett-Packard J.*, **37** (4): 4–11, April 1986.
- R. A. Witte, Low-Cost, 100-MHz digitizing oscilloscopes, *Hewlett-Packard J.*, **43** (1): 6–11, February 1992.

ANDREW RUSEK
Oakland University

ELECTRICAL VARIABLES

CHARGE MEASUREMENT

A charge is one of the fundamental quantities of electricity for which the unit is C (coulomb). One C is the total charge delivered by a current of one A for one second. Although it is generally difficult to measure the quantity and distribution of charges, they are measured by converting the charge quantity into voltage, current, light, or sound. The movement of a charge generates current and electromagnetic waves. This section, however, describes how to measure a static charge.

Here is a simple example of a parallel plate. When x is the position in the depth direction of the sample, the internal static charge $\rho(x)$, electric field $E(x)$, and potential $V(x)$ have the following relationships:

$$E(x) = \int \rho(x) dx / \epsilon_0 \epsilon_r \quad (1)$$

$$V(x) = - \int E(x) dx \quad (2)$$

Thus electric field and potential distributions are calculated by obtaining the charge distribution. For a capacitor whose capacitance is C (F), a simple example, the unknown charge quantity is obtained by measuring the potential difference as

$$Q = CV \quad (3)$$

TOTAL CHARGE MEASUREMENT

The Faraday cage is useful for measuring the total charge of an electrified material. As shown in Fig. 1, the cage has two baskets insulated from each other, and the sample is put in the internal basket. An integrated circuit with an operational amplifier is connected to both baskets. The capacitor of the integrated circuit stores the same amount of charge as the sample. The total charge quantity is measured by Eq. (3), where C is the capacitance and V is the output voltage of the integrated circuit.

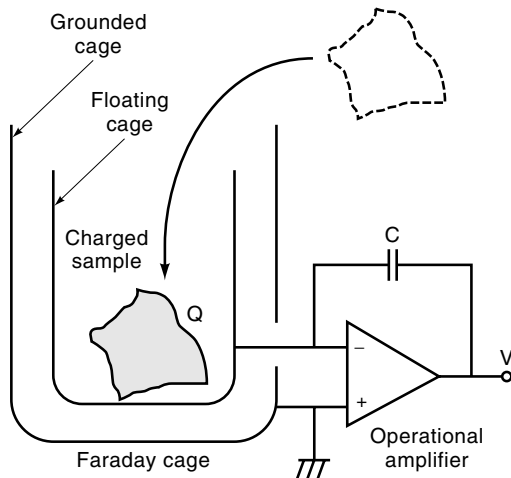


Figure 1. Total charge measurement by a Faraday cage.

SURFACE-CHARGE MEASUREMENT

Potential Probe Measurement

The surface charge of a triboelectrified sample is measured by detecting the surface potential. When the sample is a uniform plate, such as a plastic film, the capacitance per unit area should be constant, and the surface potential is proportional to the surface-charge, based on Eq. (3). The surface charge distribution is observed by detecting the surface potential. This is done by scanning the surface with a probe of a surface potential meter, as shown in Fig. 2.

Electro-Optical Surface-Charge Measurement

Figure 3 shows direct surface-charge measurement without scanning the surface. This optical measuring system uses a crystal in which an electro-optical effect occurs, for example, BSO:Bi₁₂SiO₂₀ or BGO:Bi₁₂GeO₂₀. When a charge accumulates on the surface of the crystal, an electric field is generated inside it. Because of the electric field, the crystal has birefringence, and so the linearly polarized light incident on the crystal becomes elliptically polarized light with retardation θ and permeates an analyzer. The intensity of the incident light I_0 and permeated light ΔI have the following relationship:

$$\theta = K\sigma \quad (4)$$

$$\Delta I / I_0 = \sin^2(\theta/2) \quad (5)$$

where K is the Pockels constant of the crystal, which is determined by the thickness, dielectric constant, and wavelength of the light and σ is the surface-charge density on the crystal. Thus, surface charge distribution is obtained by measuring the distribution of ΔI with a photodetector. When θ is small, however, it is difficult to measure the real surface charge because

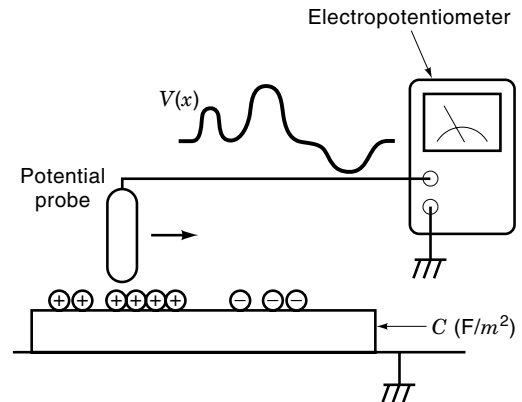


Figure 2. Surface-charge measurement. The charge distribution can be observed by detecting the surface potential with a potential meter probe.

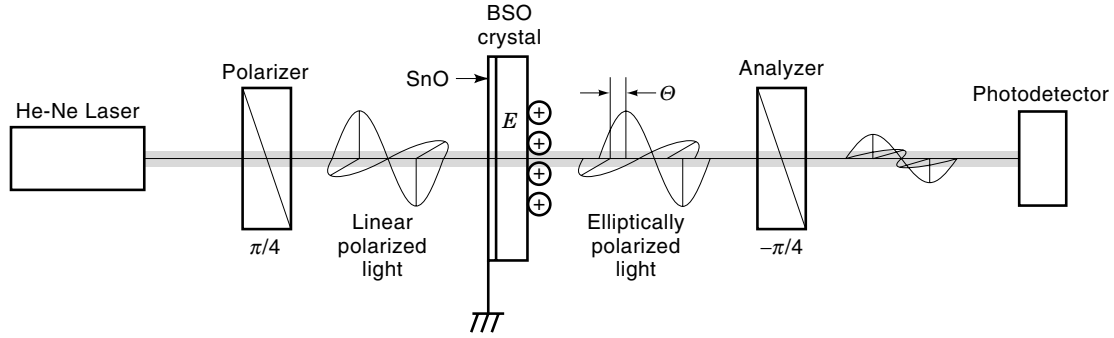


Figure 3. Direct surface-charge measurement without scanning the surface. This optical measuring system uses a crystal which produces an electro-optical effect.

insignificant leakage light generates an output signal even if there is no surface charge. In this case, a two-dimensional lock-in amplifier is used to measure the real surface charge (1). The lock-in amplifier modulates the signal to be measured and detects the synchronized component with the modulation period. Because this equipment detects very small signals, it is widely used in scientific research. The two-dimensional lock-in amplifier is an example of its application for detecting two-dimensional signals, such as images. Figure 4 shows a measurement system. The incident light is phase-modulated by applying a square voltage to an optical phase modulator made of BSO or other crystals that produce an electro-optical effect. Then, retardation θ becomes one of the following two values, alternatively:

$$\theta = K\sigma + \Delta\theta \text{ and } K\sigma - \Delta\theta \quad (6)$$

Then we calculate the difference between light intensity $\Delta I'$ in the case of $\theta = K\sigma + \Delta\theta$ and $K\sigma - \Delta\theta$. Because the same amount of leakage light exists in each case, the difference is independent of the leakage light and is given by

$$\Delta I' / I_0 = \sin(K\sigma) \sin(\Delta\theta) \quad (7)$$

Assuming that $\Delta\theta$ equals $\pi/2$, then $\Delta I'$ indicates the surface-charge distribution directly when the charge density σ is low. A two-dimensional lock-in amplifier is used to store the difference of two synchronized images in a computer. Averaging a lot of images results in a high signal-to-noise (S/N) measurement.

Figure 5 shows the surface-charge distribution on a BSO crystal. It displays the surface discharge obtained by applying an impulse voltage to a needle electrode on the surface of the crystal. This figure shows that a treelike charge pattern remains on the surface in the case of a positive discharge. In the case of a negative discharge, on the other hand, the charge pattern is round. In this way, the electro-optical measurement system with a two-dimensional lock-in amplifier directly observes the accumulation of charge changing with time.

SPACE-CHARGE MEASUREMENT

Pressure-Wave Propagation Method

In general, there are two ways to obtain a pressure wave: a piezoelectric device and a pulse laser. This section introduces the laser-induced, pressure-wave propagation method,

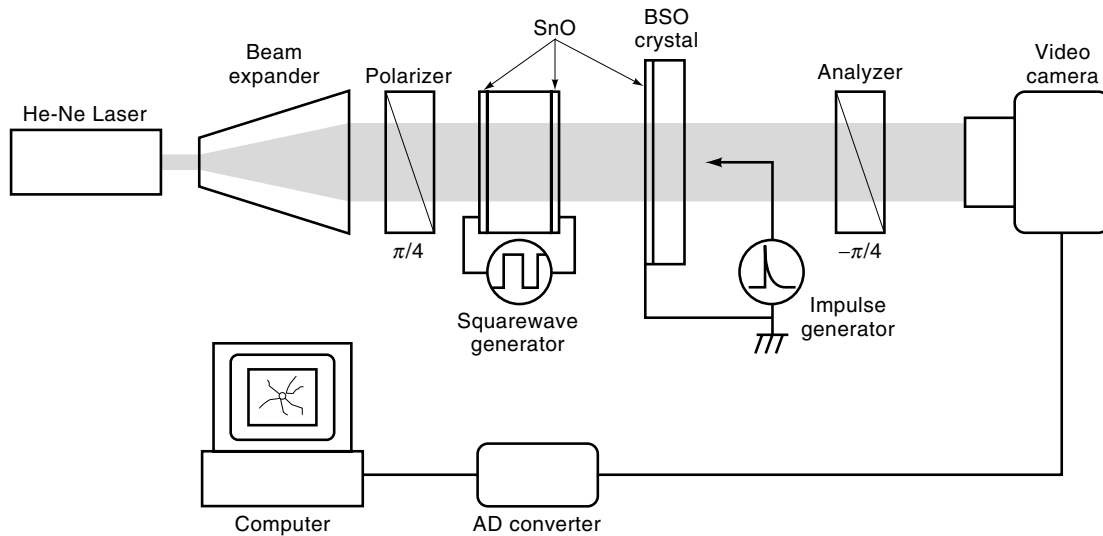


Figure 4. Two-dimensional lock-in amplifier system for surface-charge measurement.

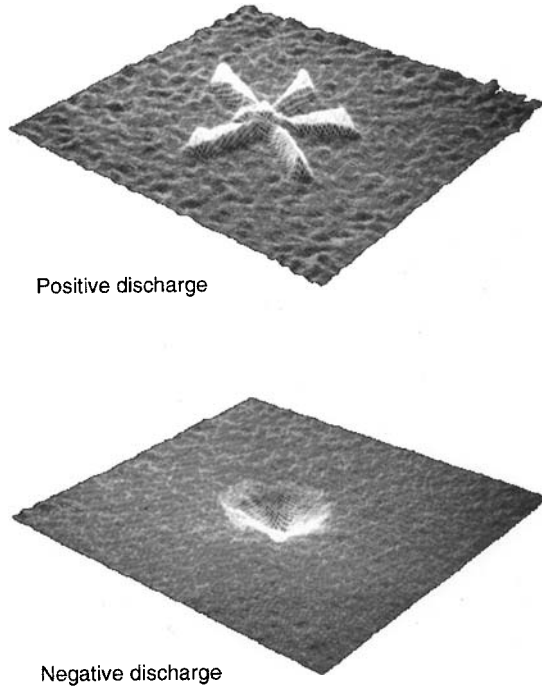


Figure 5. The surface-charge distribution on a BSO crystal. This figure shows that a treelike charge pattern remains on the surface after a positive discharge. With a negative discharge, on the other hand, the charge pattern becomes round.

which has better spatial resolution of the charge distribution and signal-to-noise ratio (2). A plate sample is placed between two electrodes, as shown in Fig. 6. When an extremely narrow-pulse laser (<1 ns) is used to irradiate the grounded electrode on the left, the electrode surface rapidly expands by being heated and then generates a pressure wave. The pressure wave passes through the left electrode and propagates in the sample at the velocity of sound v (typical polymers have sound velocities of 2000 to 3000 m/s). The pressure wave compresses a limited part of the sample. When this pressure wave passes through the position where an internal charge exists, the charge is forced to move right and left, resulting in a current pulse flowing through the conductor connected to both electrodes. The magnitude and direction of the current pulse represent the charge quantity and polarity, respectively. The

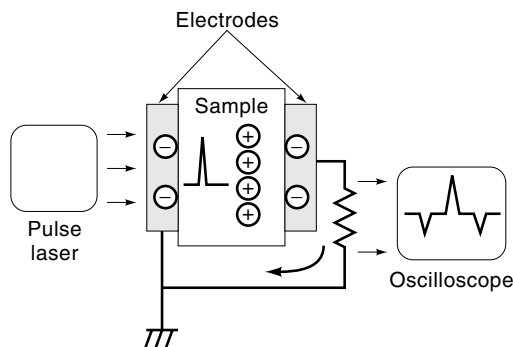


Figure 6. Laser-induced pressure-wave propagation method for space-charge measurement.

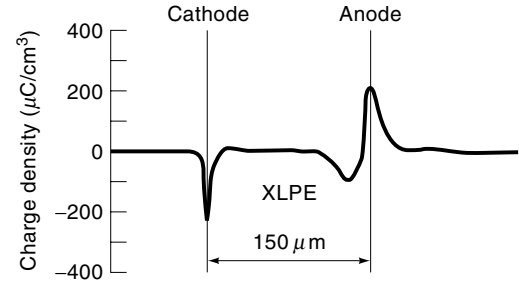


Figure 7. Space-charge distributions in an XLPE sheet measured by the laser-induced, pressure-wave propagation method.

delay from the application of the pressure wave to the current pulse generation indicates the position of the internal charge. These relationships are shown in convolution as the following equation:

$$i(t) = A \int \rho(\tau) p(t - \tau) d\tau \quad (8)$$

where $p(t)$ is the pressure wave generated by the pulse laser, $\rho(t)$ is the space-charge distribution, and $i(t)$ is the current between both electrodes. Here, $p(t)$ is an extremely narrow independent pulse, so that the pulse is regarded as the delta function, and Eq. (8) is simplified as

$$i(t) = A' \rho(t) \quad (9)$$

where A and A' are sensitivity constants depending on the strength of the laser-induced pulse. Thus the waveform of the current shows the charge distribution directly. When the width of the laser-induced pulse is 100 ps, the spatial resolution of the internal space charge becomes less than 1 μm . The disadvantage of the pressure-wave propagation method is that the amplifier in the measurement system is destroyed when surface discharge on the sample or sample breakdown occurs during radiative exposure or high voltage applications.

Figure 7 shows the space-charge distribution in cross-linked polyethylene (XLPE) measured by the laser-induced, pressure-wave propagation method (3). Because XLPE has higher thermal endurance than polyethylene, it is widely used for insulating power cables. The space charge was measured in the 150- μm XLPE film at 60°C under a 120 kV/mm dc electric field. Negative charge appeared near the anode just after the voltage was applied.

Pulsed Electroacoustic Method

The pulsed electroacoustic (PEA) method detects the space-charge distribution by a process opposite to that of the pressure-wave propagation method mentioned previously. A schematic diagram of the PEA method is shown in Fig. 8. When a narrow-pulse voltage is applied to the sample between the two electrodes, a pulse electric field is generated in the sample. This electric field $e_p(t)$ makes the space charge $\rho(x)$ generate a pressure wave because of its electrostatic force. The pressure wave propagates in the sample and the grounded electrode, and then the piezoelectric device under the grounded electrode changes it into an electric signal $v(t)$. The magnitude and direction of the pressure wave determine the quantity and the polarity of the charge. Because the delay appears when the pressure wave

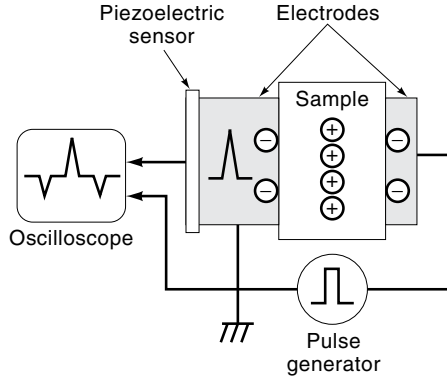


Figure 8. Pulsed electroacoustic (PEA) method for space-charge measurement.

propagates from the position of the charge to the piezoelectric device, as in the pressure-wave propagation method, the space-charge distribution $\rho(x)$ can be observed by using the sound velocity of the sample as follows.

The pressure wave $p(t)$ generated in the sample due to the applied electric field is given by

$$p(t) = A \int \rho(\tau) e_p(t - \tau) d\tau \quad (10)$$

where $\rho(t)$ is the space-charge distribution, $e_p(t)$ is the pulse electric field applied to the sample, and A is a constant depending on the dielectric constant, the sound velocity, and the density of the sample. Then the output voltage of the piezoelectric device $v(t)$ is given by

$$v(t) = \int k(\tau) p(t - \tau) d\tau \quad (11)$$

where $k(t)$ is a specific function of the piezoelectric device and is a square function whose height and width depend on the sensitivity and thickness of the device. When the piezoelectric device is so thin that the width of $e_p(t)$ is narrow enough, both $e_p(t)$ and $k(t)$ are regarded as individual pulses similar to delta functions. In this case, $p(t)$ equals $A\rho(t)$, and Eq. (11) is simplified as

$$v(t) = A'\rho(t) \quad (12)$$

Thus, the waveform of the output signal indicates the space-charge distribution. Although this method has slightly poor spatial resolution, it has the advantage that the measurement system is never destroyed when a discharge occurs at any place because the detection circuit is electrically insulated from the other parts of the electrode system, such as the high-voltage bias circuit. In particular, a system with an optical fiber for the output signal cable can carry out the space-charge measurement even during a dc breakdown test at 500 kV.

Figure 9 shows the time dependence of the space-charge distribution of XLPE (100 μm) under 10 kV. A packet charge is injected from the anode just after the voltage application, and it moves toward the cathode. Because the charge injection decreased the interface electric field at the anode, the next injection did not occur.

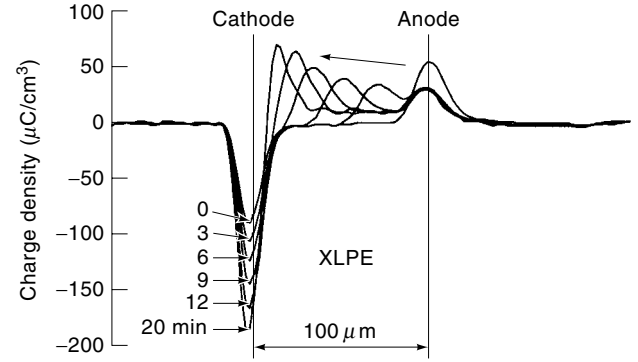


Figure 9. The time dependence of the space-charge distribution of an XLPE (0.1 mm) sheet under 10 kV. A packet charge is injected from the anode just after the voltage application, and it moves toward the cathode.

Deconvolution Signal Processing

As shown in Eqs. (8) and (11), the space-charge measuring methods mentioned previously detect a signal that includes a lot of factors, such as the waveform of the laser pulse, the waveform of the applied pulsive voltage, the frequency characteristics of the piezoelectric device and amplifiers, and so on. These unknown factors should be compensated for to avoid errors with a deconvolution technique by using a reference signal (5). In general, when the measurement system keeps its linearity, the input and output signals have the following relationship of convolution,

$$Y(f) = H(f) * X(f) \quad (13)$$

where $Y(f)$ and $X(f)$ are the Fourier-transformed signals of $y(t)$ and $x(t)$. In the measurements described, $y(t)$ is the output voltage of the amplifier observed by the oscilloscope, and $x(t)$ is the space-charge distribution $\rho(x)$. The sign $*$ is the function to multiply those components whose frequencies are the same. $H(f)$ is the transfer function of the measurement system obtained by Fourier transformation from the impulse response of the system. Here the impulse response is obtained as follows.

When a dc voltage is applied to a sample without internal charge, such as a PET film, the charge distribution becomes the same as the sheet charge on both electrodes of a parallel plate capacitance. The signal due to the accumulated charge on the one side can be regarded as an impulse (delta function) with no width in the depth direction, resulting in the impulse response itself. Because the accumulated charge q_0 is calculated easily, the Fourier-transformed Q_0 becomes a constant, independent of the frequency. When the first peak included in the output signal is described as $v_0(t)$, the Fourier-transformed $V_0(f)$ becomes

$$V_0(f) = H(f) * Q_0 \quad (14)$$

$$V(f) = V_0(f)/Q_0 * R(f) \quad (15)$$

Then, in general,

$$R(f) = Q_0 V(f)/V_0(f) \quad (16)$$

This equation, a division at the frequency region that is deconvolution, represents the Fourier-transformed, space-charge distribution. Thus, the real charge distribution is calculated by inverse Fourier transformation of $R(f)$. In practical analysis, a low-pass filter is needed to reduce the noise and division by zero in the high frequency region.

BIBLIOGRAPHY

1. T. Kawasaki et al., Highly sensitive measurement of surface charge distribution using the Pockels effect and an image lock-in amplifier, *J. Phys. D: Appl. Phys.*, **27**: 1646–1652, 1994.
2. C. Alquié, G. Dreyfus, and J. Lewiner, Stress-wave probing of electric field distributions in dielectrics, *Phys. Rev. Lett.*, **47** (20): 1483–1487, 1981.
3. N. Yoshifuji et al., Study of space-charge characteristics in polyethylene for power cable insulation by laser-induced, pressure-pulse technique, *Trans. Inst. Electr. Eng. Jpn.* **111A** (12): 1091–1098, 1991.
4. T. Maeno and H. Fukunaga, High-resolution PEA charge distribution measurement system, *IEEE Trans. Dielectr. Electr. Insul.*, **3**: 754–757, 1996.
5. T. Maeno et al., Signal processing of the high resolution PEA measurement system, *Trans. Inst. Electr. Eng. Jpn.* **115A** (5): 405–409, 1991.

TAKASHI MAENO
Communications Research
Laboratory

ELECTRIC CURRENT MEASUREMENT

Current measurements are used to monitor, control, and protect relaying. Current measurements are often the basis for estimating many other physical quantities. Electric current is measured in amperes. It is a measure of the rate of flow of electrons. Traditional direct measurement methods are based on passing the current through a measuring device. Although the device is not supposed to alter the current, direct measurement methods are generally intrusive by their very nature. The indirect measurement methods are based on measuring an effect of the current flow.

The galvanometer is an electromechanical device invented around A.D. 1794. It is based on measuring a magnetic field caused by a current flow. A sensitive magnetic needle deflects in response to the current through a coil. The galvanometer is the basis for many modern display meters.

Electric current measurement techniques have evolved through the years. The developments have addressed the need for better current transducers, for measurement of very large, very small, very fast, and very high voltage currents, and for measurement of current in unapproachable areas, like the human body and inside motors.

Maintenance of a current standard is not as straightforward as that for a resistance, mass, length, and the like. This is due to the fact that current measurement is intrusive by the very nature of the quantity. The earlier current standard was based on the Kelvin current balance (1). The balance is based on four fixed coils and two suspended coils all carrying the same current. The torque produced on the moving coils is

balanced by precise weights. An indirect current standard based on the voltage standard and the resistance standard has better acceptance.

CLASSICAL MEASUREMENT METHODS

Techniques (1,2) based on magnetic field, electric field, and heat have been in existence almost since electricity was discovered. They were basically designed for steady currents. The devices could also measure current variations slowly enough to be visible to the human eye. The dc (direct current) measurement was the standard. When ac (alternating current) was introduced, many of the dc instruments were adapted to measure equivalent dc currents.

Magnetic Field

Current flow in a conductor creates a magnetic field around it. This is called the Hall effect. The magnetic field is concentrated by allowing the current to pass through a coil. A ferromagnetic material experiences a mechanical force when placed in a magnetic field. This force is converted to a torque, and an indicator is allowed to move. This movement is used to measure the current flowing through the coil. This principle is the basis of a galvanometer. The galvanometer is still a common method for current measurement.

The wire resistance, wire inductance, and induced currents in other metal parts cause a burden on the measuring circuit. The burden can alter the current being measured. Several refinements have been made to increase sensitivity and reduce the current being diverted through the measuring circuits. Refinements include delicate balancing, use of springs for damping, and use of permanent magnets to amplify the response. If the current is alternating or varying very fast, the indication is the mean value. This method is suited for measuring the steady or average direct current. A common way of measuring alternating current is to use a rectifying bridge to measure the mean value of each half cycle. The mean value of half cycle of a pure sine wave is about 0.9 times the root mean square of the full cycle.

Electric Field

An electric potential creates an electric field around it. At normal voltages, the electric field is too small to be measured accurately by an electromechanical method. Instruments to measure the electric field have usually been very costly. Traditional voltage-measuring techniques have relied on measuring a current proportional to the voltage. At very high voltages, the electric field can be high enough to permit reliable measurements. The principle used is that there is a force of attraction between surfaces charged at a high voltage and surfaces at ground potential. This force is used to actuate sensitively balanced needles or mirrors. The resulting motion gives a measure of the voltage. This is an expensive method of very low intrusion and suitable for very high voltages.

Heat

The flow of current through a resistance results in part of the electric energy being converted to heat. The amount of energy

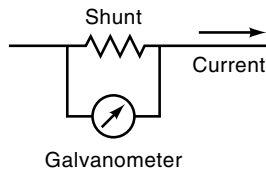


Figure 1. Current shunt.

converted is proportional to the square of current. If the current to be measured is made to flow through a thin wire, the heat expands the wire. The expansion can be made to actuate a needle or some other mechanical position indicator. The indication is a measure of the mean square of the current. Instruments based on heat is immune to ambient electromagnetic fields. However, their accuracy is affected by the ambient temperature and will need recalibration to account for this.

CURRENT TRANSDUCERS

It is often not practical to allow the current to be measured to flow through the measurement circuit. Two common methods to divert a small current proportional to the current to be measured are resistive shunts and current transformers. Rogowski coils and optical and wireless methods are some of the novel transducing techniques. A combination of these techniques is often proposed as the optimized solution offering advantages in special situations.

Resistive Shunts

A shunt is used to divert a small portion of the current into a measuring circuit. See Fig. 1. The diverted current is proportional to the unknown current. The effect on the unknown current is thus minimized. The resistive shunts are simple in concept and are widely used for steady direct current and steady alternating current as well as transient currents. The shunts should have very low resistance to minimize dissipated power. Their inductance should be negligible while measuring ac and transients. Typical values of shunt resistors are from 1.0 Ω to 0.01 Ω .

Current Transformers

Current transformers (CT) are the most versatile transducers for use in alternating current measurement. See Fig. 2. They are based on the existence of a magnetic field around any current-carrying conductor, which is called the Hall effect. They are used with currents up to 5000 A and 1000 Hz. frequency in power systems. These provide acceptable performance up to 1000 A. At higher levels, they suffer from drift problems and core saturation. They may need compensation methods while measuring harmonic currents much higher than 1000 Hz. An acceptable level of electrical insulation is required between the

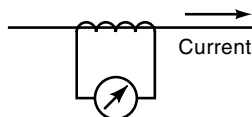


Figure 2. Current transformer.

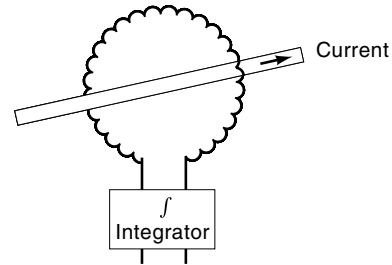


Figure 3. Rogowski coil.

CT and the high-voltage connection. Any measuring device connected to the secondary of the CT causes a burden on the main current flow. The burden should be as small as possible to avoid altering the current. For a given primary current, the burden is proportional to the core flux. Large lead distances increase the burden. High voltages require thick insulation and a large distance. Thus the use of conventional CT becomes very difficult at high voltages.

A current transformer cannot be used in direct current measurement. However, a proposed method (3) uses the effect of direct current and the magnetization curve of a transformer on its inductance characteristics. An *RL* multivibrator circuit with a nonlinear transformer is used to produce a voltage representing the current to be measured.

Rogowski Coils

The principle of Rogowski coils (4) has been well known since 1912. The Rogowski coil is a solenoidal air core winding of a small cross section looped around a conductor carrying the large transient current. The voltage induced across the terminals of the coil is proportional to the derivative of the current. The induced voltage is also proportional to the number of turns. See Fig. 3. The principle of using a Rogowski coil for current measurement is illustrated in Ref. 5. If the Rogowski coil is connected to a passive *RC* network (6), integrated effect of the voltage induced can be measured. That gives a measure of the current in the main conductor.

Rogowski coils can respond to rise times of a few nanoseconds. They have a decay time of several microseconds. They usually have been used for the measurement of very large currents such as hundreds of kiloamperes, which last for a very short duration, such as less than a microsecond. Such currents occur in electron beam accelerators and transient plasmas. They have the advantages of having a very fast response time and being free from nonlinearity and saturation effects.

Rogowski coils are not preferred for small currents because the voltage induced is small. Another disadvantage is the loss of an initial value caused by the integration process. The cost of Rogowski-coil-based current measurement is very high. A Rogowski coil with an operational amplifier integrator (7) has been shown to measure power system currents of the order of 500 kA.

Optical Transducers

The optical measurement of current (8,9) is recognized as a viable option in power systems. It has the following advantages:

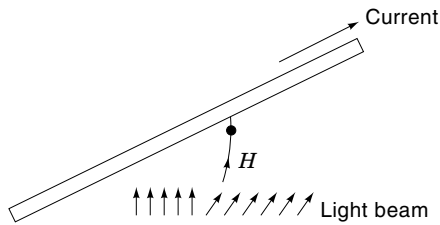


Figure 4. Pockels effect.

- Better immunity to ambient electromagnetic and electrostatic fields. They may be the only option in high electromagnetic field areas.
- Better linearity than iron core transformers. The accuracy can be substantially better.
- A high degree of isolation for safety of instruments and personnel. They are suitable for very high voltages.
- Suitable for very frequency and fast current changes.

There are various ways of implementing the optical transducer–decoder functions.

- A polarized light beam changes its plane of polarization when it passes through a magnetic field. This is called the Pockels effect. See Fig. 4. Several measurement methods based on Pockels effect have been developed (10). Optical elements like crystals, glass blocks, and fibers (11) have been used for sensing the polarizing effect of the magnetic field surrounding the electric current flow. A fiber-optic cable transmits the optical signal from the transducer to a decoder. The detection at the decoder end requires a polarization decoder.
- A polarized light beam changes its plane of polarization when it passes through an electric field. This effect was reported by Michael Faraday in 1845. This has generally been used for measuring high voltages (12,13). A method for measuring small dc currents on the order of nanoamperes to milliamperes at high voltages is presented in Ref. 14.
- In magnetostriction, the sizes of certain materials change as a result of magnetic field. A light beam reflected by or passing through such a material can produce light interference patterns caused by the phase difference. The interference will depend on the magnetic field, which in turn depends on the current. The interference pattern will give a measure of the current. A fiber-optic cable transmits the optical signal from the transducer to a decoder, which analyzes the interference pattern of the light received.
- Another method modulates the frequency of an inexpensive light emitting diode (15) and transmits that signal via an optical fiber. This approach avoids sensitive and expensive polarization coding and decoding process. An active electronic circuit, placed at the site of the high-voltage conductor, generates a voltage signal of several hundred kilohertz. The frequency deviation of the signal is proportional to the amplitude of the current. This voltage is used to activate a light emitting diode (LED). The optical fibers carry the signal to a ground station. A

photodiode responds to the optical signal and reproduces the signal at the same frequency. Another active electronic circuit at the ground station produces a voltage proportional to the frequency deviation. This voltage gives a measure of the current in the high-voltage conductor.

- Chromatic modulation (16) produced using a liquid crystal is the basis for another method. A current transformer placed on the high-voltage line and a current-to-voltage converter provides a voltage proportional to the line current. A broad band light source illuminates the liquid crystal elements of a modulator. The liquid crystal modulator attenuates a portion of the light source spectrum by an amount depending on the voltage. This light is picked up an optical fiber and brought to a ground station. The decoder consists of a double layer photodiode, with different but overlapping spectral sensitivities. The ratio of the currents produced by the photodiodes is measured with an analog-to-digital converter. This gives a measure of the current in the power line.

Transducers with No Physical Connection

Measurement of current with absolutely no physical connection has enormous advantages in cost and safety. The following methods are still being studied.

- A novel optical method eliminates (17) the use of the fiber optic connection itself. It proposes the use of a laser beam and the Faraday/Pockels effect. The reflected laser beam is used to estimate the current.
- Another method is to locate an active current-measuring device at the high-voltage level and transmit the measurement via radio frequency to a base station several meters away. There are several variations of the transmitted signal and the method of transmission. Recent developments in digital communications area have created new opportunities for investigating this approach.

CURRENT MEASUREMENTS IN SPECIAL SITUATIONS

Different measurement techniques offer different advantages and disadvantages under special circumstances. The costs can vary enormously. Some of the special situations and the preferred techniques for current measurement follow.

Currents Measurement Inside the Human Body

Human sensation of current is the combined effect of the duration, intensity, and frequency. Guidelines for limiting currents induced in the body for the frequency range of 3 kHz to 300 GHz are available (18). Electromagnetic pulses of 10 A may cause no sensation, and 100 A produces only a tingling sensation. However, steady current threshold of human body is much lower. The effect of induced currents at power frequencies is a matter of ongoing study. The steady currents induced in human body are on the order of milliamperes. The measurement techniques are less disturbing to the person and cause minimal change in the value being measured. Such techniques (19) have been developed for measurement of current in the 1 MHz to 200 MHz range.

Harmonic Current Measurement

The harmonics give a measure of the deviation from a pure sinusoidal waveform. Harmonic voltage measurements have been used as a measure of the quality of voltage in power systems. Harmonic current measurements are essential for apportioning the responsibility (20) for harmonic distortion. Conventional current transformers may not offer a sufficiently wide frequency range. An electronic technique (21) for reducing the flux in the magnetic core, thereby increasing the frequency response, has been developed. Another technique (22) reduces the flux by using two toroidal coils with opposing fluxes to have a similar increase in performance.

High-frequency currents tend to flow near the surface of conductors. Thus the resistance of the conductor varies with the frequency. Methods to compensate for the effect by analytical means (23) as well as by mechanical design (24) have been proposed.

Post Arc Current Measurement

There are arcs during circuit breaker operation, and the currents are usually high. Yet there is a current activity of milliamperes range around the zero crossing immediately following an arc. These currents provide valuable insight regarding the condition of the switchgear. The measurement sensor under these conditions must measure milliamperes with a megahertz bandwidth, without being damaged during high currents. A technique (25) automatically inserts and disconnects a measurement shunt very close to the zero crossing and measures such currents.

Motor Windings Current Measurement

There is no direct means of measuring the current in windings with no brushes. These currents are useful for control and monitoring purposes. They are estimated (26,27) using other measurable currents.

Lightning Current Measurement

Lightning-induced currents are a source of damage in electrical power lines and other conducting structures. Measuring the wave crest value and wave front duration is of interest in order to assess the damaging effects. The amplitude ranges up to thousands of amperes and the duration is on the order of a few microseconds. Some measurement results and techniques are reported in Ref. 28.

Measurement of Currents at High Voltage

The power line voltage necessitates isolation for safety purposes. Optical couplers are available for voltages up to 1000 V. These can be integrated in a circuit board along with other current-conditioning interfaces. High-voltage lines up to 1000 kV need additional levels of safety isolation interfaces. Highly insulated current transformers are commonly used in substations. They are bulky and expensive. Copper winding, iron core CT has been in successful use for 100 years. During that time, several refinements to the design have been made to improve safety and reduce failures (29). The optical (12) and wireless methods offer a great advantage at very high voltages.

Measurement of Large Currents

It is sometimes inconvenient or impossible to dismantle a bus bar for insertion of a current transformer. For these situations there are clamp-on current transformers in which a clip opens the winding for manual insertion. After insertion, the clamp firmly closes the gaps in the winding. The reclosed gap affects the permeability of the magnetic core as well as the resistance of the copper winding. However, the error introduced might not be acceptable at high currents. An electronic technique to increase the permeability of the gap and thus reduce the error has been demonstrated (30). Direct currents up to 25,000 A are measured using permalloy magnetic modulation comparators (31,32). Another design to measure currents up to 100,000 A is discussed in Ref. 33.

Measurement of Small Currents

The currents in the femtoampere range occur often as leakage currents in microelectronics. A method for implementing an in situ amplifier to amplify the small current to measurable levels is illustrated in Ref. 34.

Measurement of Fast Varying Currents

Fast varying currents occur in electron beam accelerators and transient plasmas. Rogowski coils have usually been used for measuring very large currents such as hundreds of kiloamperes, which last for a very short duration, such as less than a microsecond.

BIBLIOGRAPHY

1. British Standard Specifications for Indicating Ammeters, Voltmeters, Wattmeters, Frequency and Power factor meters, No. 89, 1929.
2. E. W. Golding, *Electrical Measurements and Measuring Instruments*, London: Pitman, 1960.
3. I. M. Filanovsky and V. A. Piskarev, Sensing and measurement of DC current using transformer and RL-multivibrator, *IEEE Trans. Circuits Syst.*, **38**: November, 1991.
4. W. Rogowski and W. Steinhaus, Die Messung der Magnetischen Spannung, *Arch. Elektrotech. (Berlin)*, **1**: 141–150, 1912.
5. D. A. Ward and J. L. T. Exon, Using Rogowski coils for transient current measurements, *IEE Eng. Sci. Educ. J.*, **2** (3): 105–113, June, 1993.
6. D. G. Pellinen et al., Rogowski coil for measuring fast high level pulsed currents, *Rev. Sci. Instrum.*, **51**: 1535–1540, 1980.
7. J. A. J. Pettings and J. Siersema, A Polyphase 500kA current measuring system with Rogowski coils, *IEE Proc., Part B*, **130** (5): 360–363, 1982.
8. T. Sawa et al., Development of optical instrument transformers, *IEEE Trans. Power Deliv.*, **5**: 884–891, 1990.
9. E. A. Ulmer Jr., A high accuracy optical current transducer for electric power systems, *IEEE Trans. Power Deliv.*, **5**: 884–891, 1990.
10. G. W. Day and A. R. Rose, Faraday effect sensors: The State of the Art, *Proc. SPIE*, **985**: 1–13, 1988.
11. T. W. MacDougall, D. R. Lutz, and R. A. Wandmacher, Development of a fiber optic current sensor for power systems, *IEEE Trans. Power Deliv.*, **7**: 848–852, 1992.

12. R. E. Hebner, R. A. Malewski, and E. C. Cassidy, Optical methods for electrical measurements at high voltage levels, *Proc. IEEE*, **65**: November, 1977.
13. B. E. Jones, Optical fiber sensors and systems for industry, *J. Phys. E*, **18**: 1985.
14. O. Tonneson, N. Beatty, and O. Skilbreid, Electrooptic methods for measurement of small DC currents at high voltage level, *IEEE Trans. Power Deliv.*, **4**: 1568–1572, 1989.
15. N. A. Pilling, R. Holmes, and G. R. Jones, Low power optical current measurement system, employing a hybrid current transmitter, *IEE Proc. Sci. Meas. Technol.*, **141** (2): 129–134, 1994.
16. N. A. Pilling, R. Holmes, and G. R. Jones, Optical fiber current measurement system using liquid crystals and chromatic modulation, *IEE Proc., Part C*, **140** (5): 351–356, 1993.
17. M. Abdel-Salam, D. O. Wiitanen, and M. Abd-Elsalem, Magnetic fields in a Faraday rotor underneath a high voltage transmission line conductor, *IEEE 22nd North Amer. Power Symp.*, Los Alamitos, CA, 1990, pp. 232–241.
18. *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields 3 kHz. to 300 GHz*, New York: IEEE, 1992.
19. M. J. Hagmann and T. M. Babij, Non invasive measurement of current in the human body for electromagnetic dosimetry, *IEEE Trans. Biomed. Eng.*, **40**: 418–423, 1993.
20. K. Srinivasan and R. Jutras, Conforming and non-conforming current for attributing steady state power quality problems, *IEEE Trans. Power Deliv.*, **13**: 212–217, 1998.
21. Y. Suzuki, A. Hirabayashi, and K. Yamasawa, Analysis of a zero flux type current sensor, *IEEE Trans. Magn.*, **29**: 3183–3185, 1993.
22. T. Sonoda and R. Ueda, A current sensor of high response and high sensitivity, *IEEE Ind. Appl. Soc. Annu. Meet.*, Seattle, WA, 1990.
23. R. Malewski et al., Elimination of skin effect error in heavy current shunts, *IEEE Trans. Power Appar. Syst.*, **100**: 1333–1340, 1981.
24. C. M. Johnson and P. R. Palmer, Current measurement using compensated co-axial shunts, *IEE Proc., Sci. Meas. Technol.*, **141** (6): November, 1994.
25. M. Barrault et al., Post arc measurement down to the ten milliamperes range, *IEEE Trans. Power Deliv.*, **8**: 1782–1788, 1993.
26. J. T. Boys, Novel current sensor for PWM AC drives, *Proc. IEE, Part B*, **135**: 27–32, 1988.
27. P. P. Acarnley, Current measurements in brushless DC drives, *Proc. IEE, Part B*, **140**: 71–79, 1993.
28. H. Haruki et al., Development of a lightning current waveform measuring system for 500kV overhead transmission lines, *IEEE Trans. Power Deliv.*, **4**: 1891–1896, 1989.
29. J. R. Boyle and H. Cummings, The Tennessee Valley Authority's experience and action plans with freestanding oil filled current transformers, *IEEE Trans. Power Deliv.*, **3**: 1769–1775, 1988.
30. E. So, S. Ren, and D. A. Bennet, High current high precision openable-core AC and AC/DC current transformers, *IEEE Trans. Instrum. Meas.*, **42**: 571–576, 1993.
31. N. L. Kusters, W. J. M. Moore, and P. N. Miljanic, A current comparator for the precise measurement of direct current ratio, *IEEE Trans. Commun. Electron.*, **CD-70**: 22–27, 1964.
32. H. Heping, Analysis and synthesis of the current comparator system of type DB-3A, *IEE Elect. Meas. Instrum.*, **12**: 3–9, 1981.
33. S. Ren, A 100000 ampere high precision on-site measurement calibration device for heavy direct current, *IEEE Trans. Instrum. Meas.*, **39**: 19–22, 1990.
34. P. Girard and P. Nouet, Evaluation of currents in the fA range, *Electron. Lett.*, **26** (13): 844–846, 1990.

K. SRINIVASAN

INDUCTANCE MEASUREMENT

Along with resistors and capacitors, inductors which are basically components with the property of inductance, are commonly used for building circuits for operating a wide range of electrical and electronic apparatus and instruments. The inductance parameter also inevitably enters the equivalent circuits of various electromagnetic devices and machines. Consequently, the need for inductance measurement arises in the characterization, design evaluation, and production tests of the relevant components and in the circuit modeling of electrical apparatus. Inductance measurement is also important in instrumentation systems where a physical quantity is sensed through its effect on the inductance of a transducer.

BASIC CONCEPTS, DEFINITIONS, AND UNITS

The Inductance Parameter

Inductance is the property whereby an electrical device sets up magnetic flux when current passes through it. In principle, any conductor of current has this property. However it is more pronounced in a coil of wire or any electrical device whose construction is akin to a coil. Taking the solenoid of N turns shown in Fig. 1(a) as representing a general situation, let a current of i amperes through it produce a magnetic flux of ϕ webers. For simplicity, if we assume that the closed path of every flux line is interlinked with all N turns, then the coil has flux linkages of $N\phi$ weber-turns. The inductance L of the coil, defined as the flux linkages produced in it per unit current, is measured in *henrys*. One henry corresponds to one weber-turn per ampere.

$$L = \frac{N\phi}{i} \quad (1)$$

Any physical device, such as the coil considered, which has the property of inductance, is called an *inductor*. It may also have other properties such as resistance and capacitance to a less

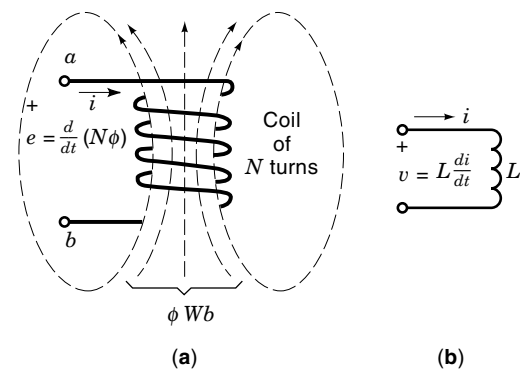


Figure 1. An inductor. (a) Physical configuration of a typical inductor. (b) Representation of a pure inductor as a circuit element.

consequential degree. If these other parasitic properties are absent, the device is referred to as an *ideal* or *pure inductor*. If the current i in an inductor and the flux ϕ produced by it are proportionally related, then the value of inductance as defined by Eq. (1) is constant and the inductor is *linear*. A *non-linear* inductor does not have a linear $N\phi$ versus i relationship and its inductance L , defined by Eq. (1), is a function of the current i .

An inductor is built in the form of a coil of good conducting material, usually copper, wound around a core of a nonmagnetic medium like air or a magnetic material such as iron. For the same core geometry, the inductance of a linear inductor is proportional to the square of the number of turns in the coil. Use of iron core permits larger values of inductance for a given inductor size but makes the inductor nonlinear because of the nonlinear magnetization characteristic of iron.

When the current i in an inductor varies, there is an induced voltage e in the inductor, which, according to Faraday's law of induction, is equal to the rate of change of its flux linkages, that is, $d(N\phi)/dt = [d(N\phi)/di](di/dt)$. This voltage acts in a direction opposing the change in i . For the coil in Fig. 1(a), if di/dt were positive, the induced voltage makes the coil behave like a voltage source driving a current in the local circuit in a direction opposite to that of i , that is, in a sense so as to arrest the growth in i . Terminal a then acquires positive polarity with respect to terminal b . For a decreasing i and hence a negative di/dt , the polarity is reversed. In either event, with the reference polarity shown in Fig. 1(a), the induced voltage in the inductor is given by

$$e = \frac{d(N\phi)}{di} \frac{di}{dt} = L \frac{di}{dt} \quad (2)$$

For a linear inductor, the value of L in the foregoing equation is the constant value of inductance given by Eq. (1). If the inductor is nonlinear, the value of L in Eq. (2) is the incremental inductance, that is, the slope of the $N\phi$ versus i characteristic at the particular operating current.

Inductor as a Circuit Element

Figure 1(b) is a representation of a pure inductor as a circuit element. In a pure inductor free from resistive and capacitive parasitic effects, the terminal voltage v equals the induced voltage e and the same current i passes through every section of the conductor. As a consequence, its terminal v - i relationship is given by

$$v = L \frac{di}{dt} \quad (3)$$

where L is the incremental inductance if the inductor is nonlinear. In the sequel, we take all inductors as linear, unless specifically stated otherwise.

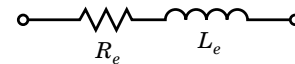
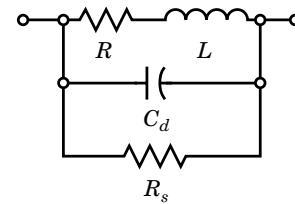
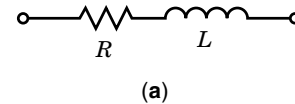
As seen from Eq. (3), the inductance parameter quantifies the opposition offered by an inductive device to a change in its current. The higher the rate of change of current, the more the applied voltage needed to support this change. The relation between v and i in Eq. (3) is a statement of the cardinal property of an ideal inductor from the circuit viewpoint and is applicable to general time-varying current and voltage. In ac circuits at steady state, where all currents and voltages are sinusoids of the same frequency, the phasors of v and i are related by the

algebraic expression,

$$\mathbf{V} = j(2\pi fL)\mathbf{I} = j\omega L\mathbf{I} = jX_L\mathbf{I} = \mathbf{Z}_L\mathbf{I} \quad (4)$$

where f is the frequency in Hz and $\omega = 2\pi f$ is the angular frequency in radians/sec, $X_L = \omega L$ and $\mathbf{Z}_L = jX_L$ are the reactance and complex impedance of the inductor, both expressed in ohms. Equation (4) embodies two important properties of an ideal inductor in an ac circuit. First, the impedance (opposition) offered to a sinusoidal current varies proportionately with frequency. Secondly, the voltage has a phase angle of $\pi/2$ radians more than the current.

Practical inductors do not occur in pure form. They include incidental resistive and capacitive properties to a degree which depends on their design. These impurities arise from the resistance of the conductor used to wind the inductor, the distributed capacitance between turns and between layers of the winding, and the imperfection of the insulation used. With dc and low frequencies, the effects of the latter two factors are negligible, and an inductor can be represented by the simple series circuit of Fig. 2(a), where L is the inductance and R the dc resistance of its winding. A more accurate representation valid over a wide range of frequencies is the circuit of Fig. 2(b). Here, C_d is the lumped-element approximation of the distributed capacitance of the winding. The shunt resistance R_s accounts for the dielectric power losses in the insulating material of the winding and the eddy-current power losses caused in the neighbouring conductors and the shield if any, by an ac current in the inductor. For iron-cored inductors, R_s also includes the effects of eddy-current and hysteresis power losses in the core and consequently has a lower value than the corresponding air-cored inductor. Furthermore, with magnetic cores, R_s and L are



$$R_e \approx R \left(1 + 2\omega^2 L C_d + \frac{\omega^2 L^2}{R R_s} \right)$$

$$L_e \approx L (1 + \omega^2 L C_d)$$

(c)

Figure 2. Circuit representations of a practical inductor. (a) Equivalent circuit suitable for an air-cored inductor at dc and low frequencies. (b) Equivalent circuit valid for a wide range of frequencies. (c) Simplified equivalent circuit valid for $R < \omega L < R_s$ and $\omega^2 L C_d < 1$.

nonlinear. Their values depend on the applied voltage at any given frequency.

In ac circuit applications of an inductor, the parasitic elements R , C_d , R_s restrict the useful frequency range over which the device has a predominantly inductive character. This range extends over frequencies which are well below the self-resonant frequency of the inductor ($\omega^2 LC_d < 1$) and at which the inductive reactance ωL is much larger than the series resistance R and much smaller than the shunt resistance R_s . Under these conditions, the equivalent circuit of Fig. 2(b) is approximated by the simpler series equivalent circuit of Fig. 2(c), where the effective inductance L_e and series resistance R_e are functions of frequency given by

$$L_e = L(1 + \omega^2 LC_d) \quad (5)$$

and

$$R_e = R \left(1 + 2\omega^2 LC_d + \frac{\omega^2 L^2}{RR_s} \right) \quad (6)$$

In the frequency range under consideration, the effect of self-capacitance C_d , thus, is increased effective inductance and series resistance over their dc values. A larger C_d leads to larger increases. The dielectric and other losses represented by R_s contribute to an increased R_e with negligible effect on L_e . A further contributing factor to the dependence of R_e on frequency is that the winding resistance R is a function of frequency which increases above its dc value at higher frequencies because of skin and proximity effects. The resulting modified current distribution over the cross section of the conductors also lowers the value of L at higher frequencies.

The quality factor Q is commonly used as a figure of merit for a practical inductor in ac applications. A measure of the effectiveness of an energy storage device in performing its role with small power dissipation, it has the general definition

$$Q = 2\pi \frac{\text{peak value of energy stored}}{\text{energy dissipated per cycle}} \quad (7)$$

For the series equivalent circuit of Fig. 2(c), its value is equal to the ratio of its reactance to resistance, that is,

$$Q_e = \frac{\omega L_e}{R_e} \quad (8)$$

The Q -factor of a practical inductor exhibits a characteristic variation. Increasing linearly with frequency at first, then more gradually, it attains a peak in middle range of frequencies before it tapers off to low values at higher frequencies well below the self-resonant frequency. The self-capacitance of the winding, the skin and proximity effects on the conductor resistance, and the various power losses represented by R_s all contribute to reducing Q at high frequencies. The midrange of frequencies over which Q is fairly large and nearly constant is important in applications of the inductor as a circuit element in filter circuits. Inductors with iron cores have larger values of Q but a smaller usable frequency range. The maximum attainable value of Q with practical inductors is limited to a few hundred.

The important parameters of an inductor in circuit design are L_e and Q_e . The latter is preferred to R_e because it shows a smaller variation over the frequency band of interest. Because

the theoretical deduction of the values of these parameters at any specified frequency has limited scope, it becomes necessary to measure them experimentally. Measurements on inductors imply the measurement of L_e and Q_e (or alternatively R_e) at different frequencies over a specified range.

The Mutual Inductance Parameter

The inductance of a single isolated coil discussed so far is called its self-inductance. The property of mutual inductance arises when two inductors are in proximity so that some or all the flux lines set up by the current in one inductor also link with the other inductor. Because of this magnetic coupling, when the current in the first inductor varies, its self-flux linkages and also the flux linkages created by it in the other inductor (called mutual flux linkages) vary, thereby inducing voltage in the latter.

To illustrate the concept of magnetic coupling, we consider two ideal linear inductors constituted by coils of N_1 and N_2 turns, as shown in Fig. 3(a), and with self-inductances of L_1 and L_2 . As a simplified description, let current i_1 passing through coil 1 establish a self-flux of ϕ_{11} webers which links with all of

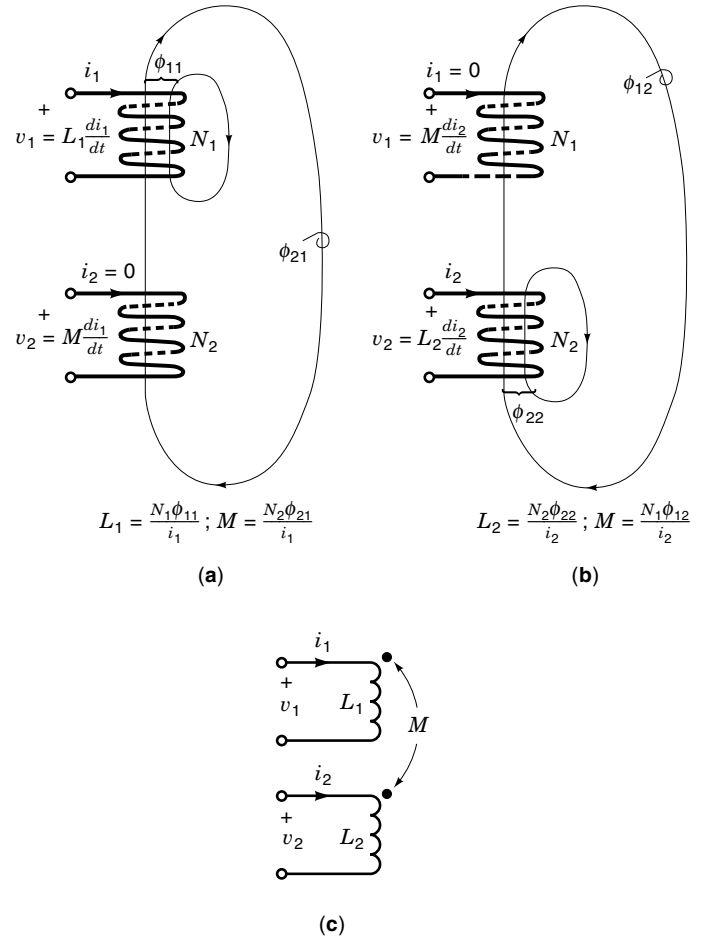


Figure 3. Mutual inductance between two coils. (a) Voltages induced when coil 1 alone carries a current. ϕ_{11} and ϕ_{21} represent the self and mutual fluxes. (b) Voltages induced when coil 2 alone carries a current. ϕ_{22} and ϕ_{12} represent the self and mutual fluxes. (c) Circuit representation of two coupled coils.

its N_1 turns and let ϕ_{21} , a part of this flux, link with all the N_2 turns of coil 2. The mutual inductance M between the coils is defined as the flux linkages in coil 2 per unit current in coil 1, that is, $M = N_2\phi_{21}/i_1$ and is measured in henrys, the same units adopted for self-inductances. It is also defined as the induced voltage in coil 2 per unit rate of change of current in coil 1. The self- and mutually induced voltages in the two coils, when only one of them carries a current, are specified in Figs. 3(a) and (b). The equality of $\phi_{21}N_2/i_1$ in Fig. 3(a) and $\phi_{12}N_1/i_2$ in Fig. 3(b) is established from energy considerations and is termed the reciprocity property of mutual inductances. The circuit representation of two coupled ideal inductors is given in Fig. 3(c). When they carry currents i_1 and i_2 simultaneously, their terminal voltages are given by Eqs. 9(a) and 9(b):

$$v_1 = L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} \quad (9a)$$

$$v_2 = M \frac{di_1}{dt} + L_2 \frac{di_2}{dt} \quad (9b)$$

One terminal of each inductor is specially marked, often with a *dot*, as in Fig. 3(c). If the two currents, in their reference directions, both enter (or both leave) their respective inductors through the dotted terminals, then the self- and mutual fluxes are in the same direction in each coil. In this event, the signs of both the self- and mutually induced voltages in each terminal equation are the same. They are opposite otherwise. In Fig. 3(c), the reference directions for i_1 and i_2 both enter the inductors through dotted terminals and hence Eqs. 9(a) and (b) have the same signs for both self- and mutually induced voltages. The dot notation avoids the necessity of indicating explicitly the winding sense of each coil as done in Figs. 3(a) and (b).

In steady-state ac circuits with sinusoidal currents and voltages of angular frequency ω , the terminal equations of the coupled inductors in terms of the respective phasors take the form

$$\bar{V}_1 = (j\omega L_1)\bar{I}_1 + (j\omega M)\bar{I}_2 \quad (10a)$$

and

$$\bar{V}_2 = (j\omega M)\bar{I}_1 + (j\omega L_2)\bar{I}_2 \quad (10b)$$

The quantity ωM measured in ohms is the mutual reactance between the inductors.

STANDARD INDUCTORS

The term “standard” refers to an object built to realize a physical quantity or parameter very accurately. Inductance is the parameter of concern in this section, and we deal with two groups of standards, namely, standards of self-inductance and those of mutual inductance. Either type is further classified as fixed or variable, depending on whether the standard has a single definite value of the inductance parameter or whether it permits a continuous variation of the value over a certain range through an appropriate adjustment.

Another classification of standards is absolute or secondary. An absolute standard of inductance is one whose inductance is computed from accurately measured dimensions of the device. Such standards serve as the basis for calibrating secondary standards. The latter are either reference standards (which are

directly compared with absolute standards) or laboratory working standards (which are calibrated with respect to the reference standards). Both the reference and working standards of inductance have essentially the same constructional features. These are discussed after the following section on absolute standards.

Absolute Standards

Attaining the highest possible accuracy rather than economy is the chief concern in designing absolute standards. These standards of inductance incorporate coil configurations that permit applying the relevant formula for inductance with negligible error. The choice of materials and details of construction ensure precise and permanent positioning of the conductor on the coil former. The coil configuration generally adopted is a single layer solenoid of copper wire uniformly wound on a truly cylindrical former made of glass, ceramic, or marble and with a precisely machined helical groove to accommodate the conductor. The cost involved in their construction and maintenance and the sophisticated optical methods used to measure their dimensions render these standards unsuitable for use in a general laboratory. The tasks of maintaining absolute standards and of employing them to calibrate secondary standards are undertaken only at the national standards laboratories in each country. Even here, absolute standards of mutual inductance are preferred to absolute standards of self-inductance because formulas of higher theoretical accuracy are available for mutual inductance than for self-inductance and because skin effect and self-capacitance of windings are of smaller consequence. Details of representative absolute standards of mutual inductance are available in Refs. 1,2,3.

The cylindrical cross-capacitor, whose capacitance is determined by a single length measurement has become a fundamental electrical standard of choice in national laboratories, because of its extremely high accuracy. As an analog, a calculable mutual inductance standard in the form of a cage of five parallel wires is proposed by Page (4) for calibrating mutual inductor reference standards. The mutual inductance of this calculable standard is given by

$$\frac{1}{10} \ln \frac{3 + \sqrt{5}}{2} \mu\text{H/m}$$

Fixed Inductance Standards

As a rule, standard inductors are made without iron cores to obtain an inductance value independent of current. Marble, glass, porcelain or wood impregnated with paraffin are some of the materials used to construct the coil former. They are nonmagnetic and provide good insulation and adequate dimensional stability. Litz wire is employed for the coil to reduce the consequences of skin effect in the conductor.

There are two basic coil forms, illustrated in Fig. 4, in common use for self-inductance standards. The flat cylindrical coil of square winding cross section shown in Fig. 4(a) provides larger inductance values for a given length and diameter of the conductor than the toroidal configuration of Fig. 4(b). The maximum inductance is obtained when $d = 3a$. However, the arrangement is significantly affected by external magnetic fields from other sources and produces a large interfering field of its own. The uniformly wound toroidal inductor, on the other hand,

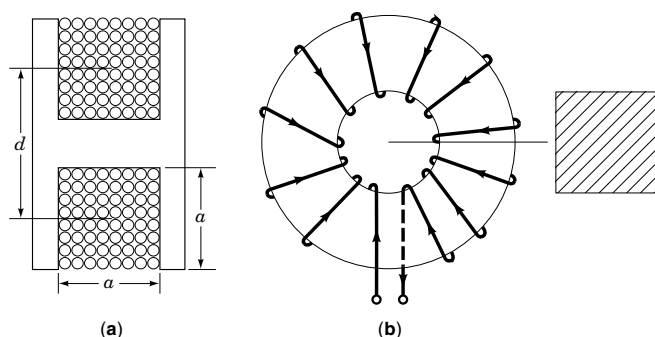


Figure 4. Windings for standard self-inductors. (a) Solenoidal coil provides larger Q values. Optimum design requires $d = 3a$. (b) Toroidal coil provides a nearly astatic arrangement.

has smaller Q values but provides a nearly astatic arrangement. As far as generation of and coupling with external fields is concerned, the toroidal winding is equivalent to a single turn with the mean radius of the torus. Figure 5 illustrates two ways by which this “single-turn effect” is effectively overcome. An auxiliary circular turn carrying a current in the opposite direction is provided on the surface of the toroid in the arrangement of Fig. 5(a). In the duplex winding of Fig. 5(b), the current from a lead does not take a path around the complete toroid, but splits equally into two parallel paths, each around half the toroid, and recombines at the diametrically opposite point. The currents in the two parallel paths have a cumulative action in establishing the magnetic flux around the toroid but act in opposite directions in setting up an interfering field in the axial direction.

Due attention is necessary in designing and constructing standard inductors to the size and placement of terminals and to the effects of temperature on the inductance value. See Refs. 1,2, and 5. The manufacturer of a standard inductor certifies its effective inductance at a particular frequency and suitable corrections are needed for use at other frequencies. In any

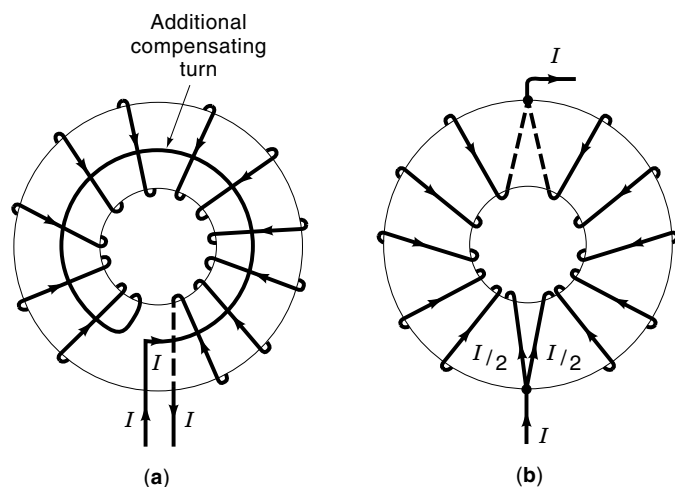


Figure 5. Methods of compensation for “single-turn effect.” (a) An additional circular turn carries current in the opposite direction. (b) In the duplex winding, current flows clockwise in one-half of the effective single turn and anticlockwise in the other half.

event, the operating frequency must be well below the self-resonant frequency of the inductor. The typical accuracy of a standard inductor is 0.1%.

Fixed mutual-inductance standards have design features similar to those of self-inductance standards except that they consist of two windings instead of one. Two circular coils of the same radius are fixed on the same bobbin axially displaced from one another, or they are concentric coils of different radii.

Variable Inductance Standards

Decade units of self- and mutual inductances are realized by an appropriate series connection of uncoupled fixed self-inductors or of sections of the secondary winding of a mutual inductor. The stepwise variation available with such units may not be sufficiently fine in several applications. Furthermore, in bridge measurements, the need exists for an inductor whose inductance value can be continuously changed without changing its resistance. Such a variable inductor is called a variometer or an inductometer. As a rule, these devices are less accurate than fixed standards. A typical value is 1%.

An inductometer consists of two coils, which are rotatable with respect to each other, so that the mutual inductance between them can be varied continuously, say from $-M_0$ to $+M_0$. If the two coils are connected in series, then the self-inductance of the combination can be varied from $L_1 + L_2 - 2M_0$ to $L_1 + L_2 + 2M_0$ where L_1 and L_2 are the individual self-inductances. Thus the inductometer is used as a variable self-inductor or as a variable mutual inductor.

Figure 6 illustrates two forms of the inductometer. The Ayrton–Perry arrangement, in which the fixed and rotatable coils are wound on sections of spherical surfaces, has the disadvantages of lack of astaticism and a nonlinear variation of M

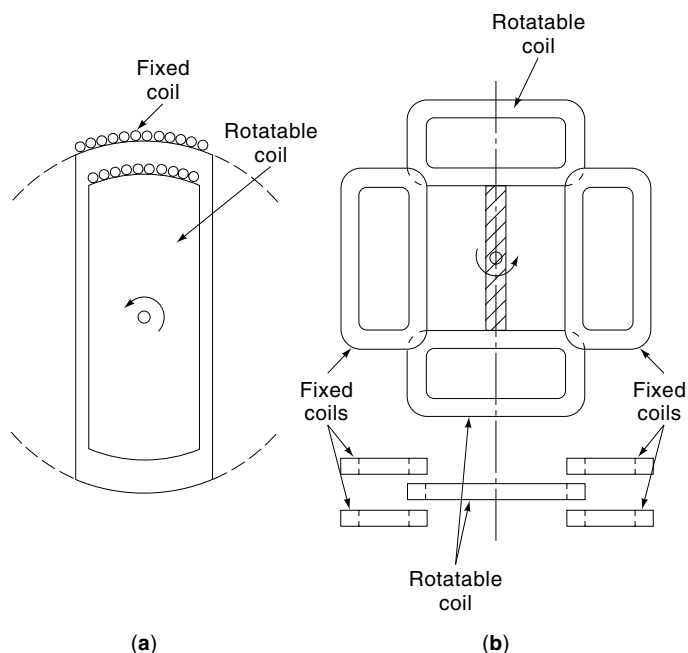


Figure 6. Two forms of inductometer. (a) The Ayrton–Perry inductometer has a nonlinear scale and is not astatic. (b) The Brooks–Weaver inductometer provides a linear scale and is astatic.

with angular rotation. The Brooks–Weaver inductometer has a disk-shaped structure with two pairs of fixed coils and a rotatable third pair sandwiched between them. The left and right groups of fixed coils are connected to form an astatic arrangement. When the flux direction is upward in one group, it is downward in the other group, and one side serves as the return path for the flux of the other group. The coils are shaped to provide a fairly linear variation of M with the angle of rotation. An advantage of the arrangement is that small axial displacements of the rotor with respect to the stator caused by mechanical wear of the bearings do not significantly alter the value of M at any angular position.

METHODS OF MEASUREMENT OF SELF-INDUCTANCE

The salient schemes of measurement, grouped under four heads, are detailed in the following. All measurements are made under sinusoidal steady-state conditions and the values of L_e and R_e (or equivalently Q_e) of the inductor at the test frequency are determined. The inductor current (voltage) in the measuring circuit should be kept at the same level as in the actual application of the device to obtain reliable results for nonlinear inductors.

Methods Employing Indicating Instruments

These methods are easy and convenient in implementation and yield results of moderate accuracy. Figure 7 illustrates two commonly adopted schemes.

In the three-voltmeter method of Fig. 7(a), a known standard resistance R_s is connected in series with the test inductor. The values of R_e and L_e are computed from the three voltmeter readings V_s , V_1 , and V_2 , as indicated in the figure. The large number of algebraic operations involved render the final results prone to gross accumulation of errors.

The wattmeter method of Fig. 7(b) yields the values of L_e and R_e from the readings V , I , and P of the voltmeter, ammeter, and wattmeter. This method is particularly suitable at power frequencies. Its frequency range is limited by that of the dynamometer wattmeter.

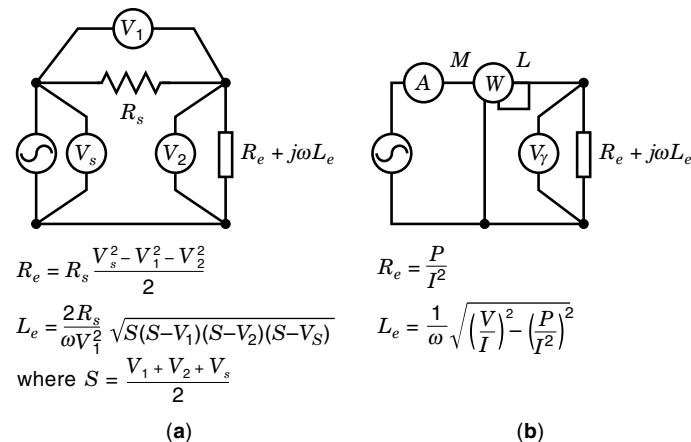


Figure 7. Inductance measurement using indicating instruments. (a) Three voltmeter method. (b) Wattmeter method.

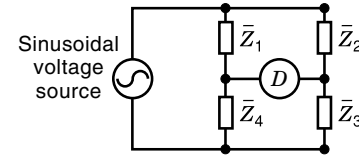


Figure 8. Standard form of a four-arm bridge network.

ac Bridge Measurements

The ac bridge method is one of the most accurate and sensitive means of determining the parameters of a circuit component such as an inductor. The accuracy stems from the fact that the unknown impedance is measured purely in terms of standard circuit components, which can be built to a much higher degree of accuracy than the best indicating instruments and are also less susceptible to deterioration in accuracy from ageing and environmental influences.

The Four-Arm ac Bridge. The four-arm ac bridge network has the general form shown in Fig. 8. The voltage source is typically a sinusoidal oscillator whose frequency is adjusted to the desired value. $\bar{CIZ}_4$ is the impedance to be measured, and the other three arms contain fixed and variable standard elements. The variable elements are adjusted until the detector D senses zero voltage across it. The bridge is then balanced, and $\bar{CIZ}_4$ is given by $\bar{CIZ}_4 = \bar{CIZ}_1 \bar{CIZ}_3 / \bar{CIZ}_2$.

The previous equation in complex numbers yields two real equations in terms of the bridge elements, which are called the balance conditions of the bridge. At least two variable elements should be provided to force bridge balance. Variable resistors are usually preferred for this purpose. If each variable element enters only one of the balance equations exclusively, then the balance adjustments are independent and the balancing process is faster. Furthermore, the balance conditions should preferably be independent of frequency to achieve a sharper balance detection when the source voltage contains harmonics. Detectors include vibration galvanometers, telephone receivers, tuned-amplifier detectors, oscilloscopes and radio receivers. References 1, 2, 6 give a comprehensive account of their relative characteristics and useful frequency ranges.

With the background of the foregoing general remarks, let us now look at a few popular ac bridge networks used for measuring inductors. Table 1 gives the configurations and balance equations of these networks. For measuring high- Q inductors, impractically large values of R_2 are needed in the Maxwell–Wien configuration. This difficulty is overcome by reconnecting R_2 and C_2 in series in arm 2. The resulting configuration, known as the Hay bridge, is quite suitable for use with high- Q inductors. The Owen bridge yields good accuracy but needs a large-valued adjustable capacitor C_1 for high- Q inductors. The Anderson bridge, a modification of the four-arm configuration, is quite suitable for a wide range of inductances and Q -values. A merit of this network is that preliminary balance is obtained with dc excitation and a suitable dc detector by adjusting r and the final balance with ac excitation effected essentially with R_5 accompanied by a slight readjustment of r .

Several important issues common to all bridge measurements are the rapidity with which balance adjustments converge, the sensitivity of detection, the effects of residuals (impurities) in the bridge elements, and the adoption of screening

Table 1. Bridge Networks for Measurements on Self-Inductors

Type	Configuration	Balance Equations	Remarks
Maxwell		$L_e = L_3 \frac{R_1}{R_2}$ $R_e = R_3 \frac{R_1}{R_2}$	Requires an inductometer.
Maxwell±Wien		$L_e = C_2 R_1 R_3$ $R_e = R_3 \frac{R_1}{R_2}$ $Q_e = \omega C_2 R_2$	For measurement of L_e and Q_e , the adjustable elements are R_1 (or R_3) and R_2 . Suitable for low Q ($Q < 10$) inductors.
Owen		$L_e = C_2 R_1 R_3$ $R_e = \frac{C_2}{C_1} R_3$	Suitable for a wide range of inductance values. Determination of L_e in terms of a fixed capacitor is an advantage.
Anderson		$L_e = C_5 R_1 \left[R_3 + R_5 + \frac{R_3 R_5}{R_2} \right]$ $R_e = R_1 \frac{R_3}{R_2} - r$	Suitable for a wide range of inductance. Only resistors are used as balancing elements.

and other strategies to combat the effects of stray admittances and parasitic coupling between elements. The reader may consult Refs. 1, 2, and 5–11 for good coverage of these topics.

Typical commercial bridges for inductance measurements have accuracies ranging from 0.05% to 1%. Besides measurements on conventional coils used as circuit components, inductance bridges are also employed for such purposes as determining inductive output impedances of amplifiers, inductance and Q -values of loudspeaker voice coils and television deflection yokes.

T-networks for RF Measurements. To combat the effects of stray capacitances, which are more pronounced at higher frequencies, it is desirable to ground one end of both the voltage

source and the detector. The four-arm bridge does not permit this without using isolating transformers. Two networks that permit this condition are the bridged-T and the twin-T measuring networks, which are useful for measuring inductances, especially at radio frequencies. These have three-terminal, two-port configurations, and the balance equations necessarily involve the frequency. References 2, 5, 6, and 8 may be consulted for further details.

Transformer Ratio Bridges. A very precise and stable voltage or current ratio standard is established by a suitably designed transformer. The voltage or current ratio is determined by the ratio of the number of turns in two windings of a transformer and is unaffected by temperature, ageing and other

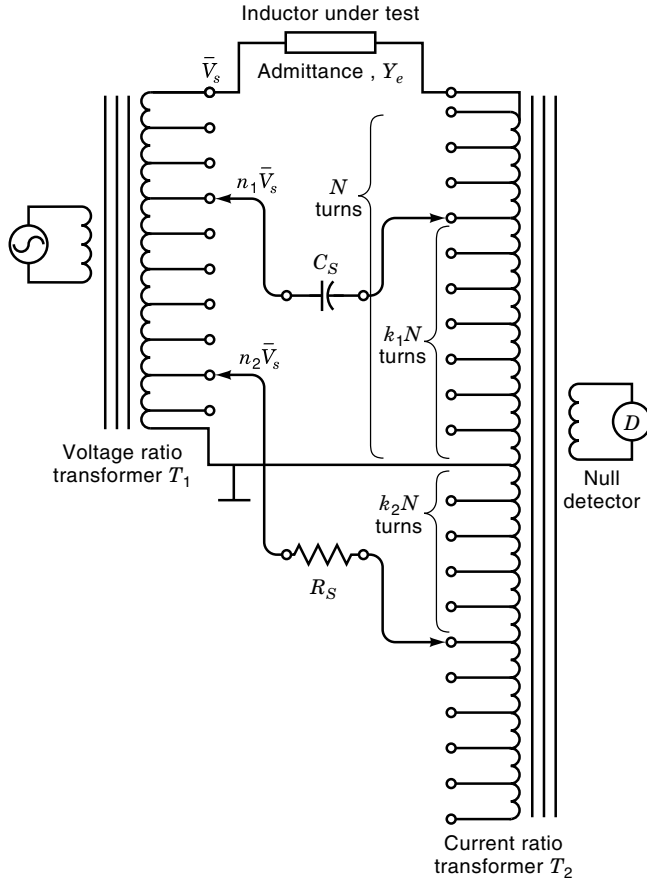


Figure 9. The double-ratio, transformer-bridge circuit measures the admittance $\overline{CY}_e$ of the inductor in terms of fixed standards C_s and R_s and transformer ratios n_1 , n_2 , k_1 , and k_2 .

influences. The imperfections of the materials used for the core and conductor and in the magnetic coupling between the various turns cause only second-order effects, which can be minimized by careful design and choice of materials. Transformer ratio bridges that take advantage of this property have been developed over the last four decades, which incorporate variable-ratio standards of extraordinarily fine resolution (typically 1 in 10^7) and compatible accuracy. The transformer ratio bridges effectively solve the problem of stray admittances in measuring impedances and have the further advantage of needing only fixed-value standard R and C elements. The principle of a transformer ratio bridge scheme of inductor measurement is shown in Fig. 9. The transformer ratios n_1 , n_2 , k_1 , and k_2 are adjusted to secure a null response in the detector D . The flux in the core of T_2 is zero under these conditions. Hence, the primary winding of T_2 presents only a very small impedance equal to its resistance and the currents in C_s and R_s are controlled by $n_1 \overline{CIV}_s$ and $n_2 \overline{CIV}_2$, respectively. The ampere-turn balance at this condition is given by

$$N(\overline{Y}_e \overline{V}_s) + k_1 N(j\omega C_s n_1 \overline{V}_s) - k_2 N(n_2 \overline{V}_s / R_s) = 0$$

or

$$\overline{Y}_e = \frac{k_2 n_2}{R_s} - j\omega k_1 n_1 C_s \quad (11)$$

In practice, T_1 consists of six to eight decade dividers connected in cascade, and hence the ratios n_1 and n_2 can be fixed in steps of 10^{-6} to 10^{-8} . The details are omitted in Fig. 9 for simplicity. Transformer ratio bridges have been made for use at frequencies up to 200 MHz. References 2, 8, 10, and 11 may be consulted for additional details on transformer ratio bridges.

Resonance Methods

If a variable capacitor is employed to resonate with the inductor under test, then the inductance value is deduced from the values of the capacitance at resonance and source frequency. A bridge method and a meter method using this principle are illustrated in Fig. 10. Both methods are particularly suited to measurements in the RF range, which permits the use of small capacitors.

In the resonance bridge of Fig. 10(a), C_4 and a resistor in one of the other three arms is varied to secure the balance. At balance, arm 4 of the bridge is equivalent to a pure resistance, and an accurate value of R_e can be deduced. The frequency f must be known accurately, because it enters the balance equation for L_e . Furthermore, the waveform of the source must be pure for sharp detection of balance.

The Q -meter is a versatile and useful instrument for measuring inductors and capacitors at radio frequencies. Referring to the circuit of Fig. 10(b), a small voltage V_s of fixed value (of the order of 20 mV) is introduced into the series-resonant circuit formed by the test inductor and a variable standard capacitor C . R_s is a small, fixed, noninductive resistance (of the order of 40 mΩ) and the current I through it is sensibly equal to I , which is read by an RF ammeter and adjusted to a predetermined value to yield the required value of V_s . Now the capacitance of C is adjusted to result in the maximum value $(V_0)_m$ of the voltage across it read by the high impedance voltmeter V . For all practical purposes, this is also the resonant condition of the circuit and the voltage magnification $(V_0)_m / V_s$ is equal to the Q -factor of the coil. Accordingly, we deduce L_e and Q_e at frequency f from

$$L_e = 1 / (4\pi^2 f^2 C_0)$$

and

$$Q_e = (V_0)_m / V_s \quad (12)$$

where C_0 is the value of the capacitance for maximum V_0 .

Because V_s has a predetermined value, the output voltmeter is calibrated directly in terms of Q_e . By repeating the above measurement at a revised frequency of $2f$, the self-capacitance C_d of the inductor is deduced as $(C_0 - 4C'_0)/3$, where C'_0 is the value of the variable capacitor needed for resonance at the frequency $2f$.

Digital Methods

Modern electronics has brought a variety of instruments to the art of impedance measurement which provide the convenience of a direct digital readout and accuracies comparable to commercial measuring bridges. The underlying principle of measurement in most of these instruments is based on one or the other of the two schemes shown in Fig. 11.

The inductor under test and a standard resistor R_s of known value are connected in series in the scheme of Fig. 11(a). The

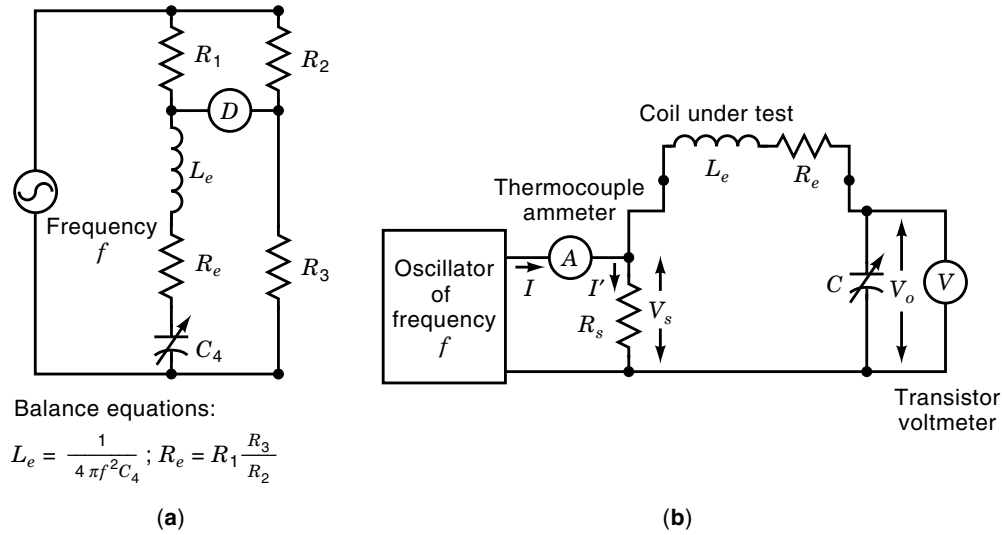


Figure 10. Resonance methods employ a variable capacitor to resonate with the inductor at the test frequency. (a) Resonance bridge. (b) Q-meter.

components of the phasor of the inductor voltage in phase and in quadrature with the phasor of the resistor voltage are measured by suitable analog and digital electronics. The second scheme shown in Fig. 11(b) employs an automatic balancing arrangement to null the detector voltage $\overline{CIV}_d$. This is effected by varying the in-phase and quadrature components of the controlled voltage $\overline{CIV}_z$. In both schemes, the parameters α and β are given by

$$\alpha = R_e/R_s \quad (13a)$$

$$\beta = \omega L_e/R_s \quad (13b)$$

and form the basis of the digital display of the values of R_e and L_e . The details of the circuit implementation in a few representative designs are given in Refs. 11–13.

Modern digital *RLC* meters measure a wide range of impedances at frequencies from a few hertz to megahertz with

a basic accuracy of 0.1% to 1.0%. They are being increasingly used for routine measurements in the laboratory and for production and inspection tests.

MEASUREMENT OF MUTUAL INDUCTANCE

The measurement of mutual inductance is called for in testing mutual inductors and when dealing with mutual inductance transducers for measuring nonelectrical quantities. We shall look at a few representative methods.

Voltmeter-Ammeter Method

In the voltmeter-ammeter method, Fig. 12(a), the current in the primary winding and the induced emf in the secondary are measured by an ammeter and a high-resistance voltmeter, such as an electrostatic voltmeter or an electronic voltmeter. M is calculated as $V_2/\omega I_1$. The value of the supply frequency is needed for this measurement.

Measurement in Terms of Self-Inductances

This is a popular method of measuring mutual inductance. If two coils whose mutual inductance is to be measured are connected in series, the self-inductance of the combination is given by $L_1 + L_2 \pm 2M$, depending on whether the magnetic fluxes set up by a common current in the two coils are additive or not [Fig. 12(b)]. Thus, M is deduced from the measurement of the two resulting values of self-inductance. Any one of the methods of measurement of self-inductance can be used to determine L_a and L_b . If $M < L_1 + L_2$, the method suffers from inaccuracy because then the value of M is deduced from the difference between two nearly equal measured values.

Mutual Inductance Bridges

There are several bridge methods for measuring M (1,2,6,7,9). The Carey-Foster bridge shown in Fig. 12(c) is representative. It measures M in terms of a fixed capacitor. At balance,

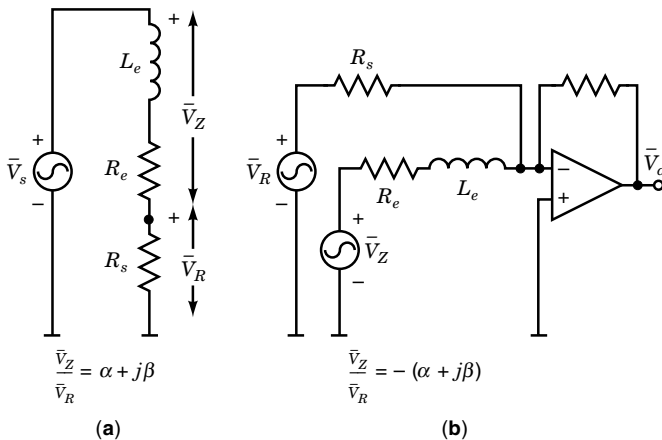


Figure 11. Schemes adopted for digital measurement of impedances. (a) The measurement of α and β enables the determination of R_e and L_e . (b) Feedback forces the controlled voltage $\overline{CIV}_z$ to a value that makes $\overline{CIV}_d = 0$.

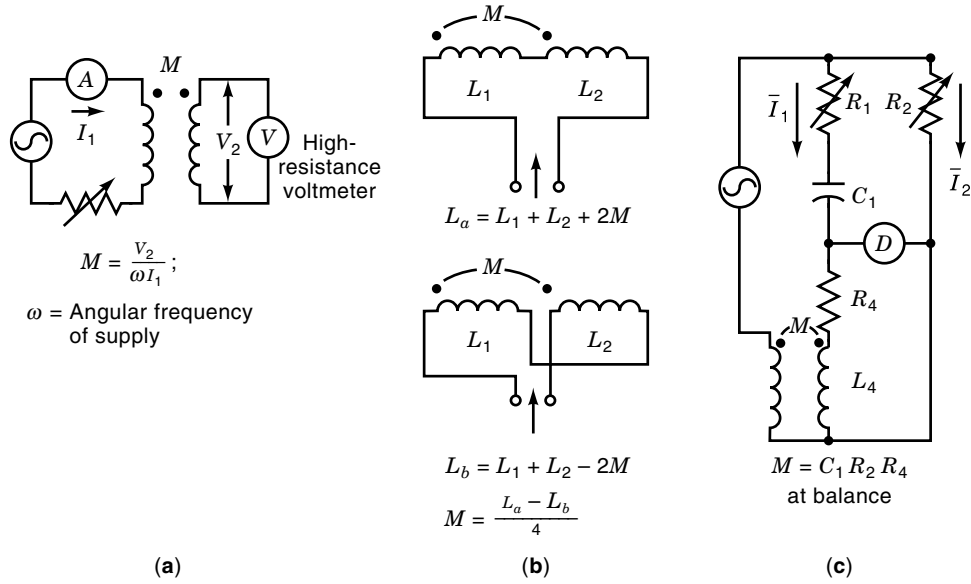


Figure 12. Methods of measuring mutual inductance. (a) The voltmeter-ammeter method is direct and simple but requires the frequency value. (b) M is deduced from the measured self-inductance values L_a and L_b . (c) The Carey-Foster bridge measures M in terms of a capacitance and two resistances.

$(R_1 - j/\omega C_1)\overline{CII}_1 = R_2\overline{CII}_2$, and $(R_4 + j\omega L_4)\overline{CII}_1 - j\omega M(\overline{CII}_1 + \overline{CII}_2) = 0$. Solution of these equations yields

$$M = C_1 R_2 R_4$$

and

$$L_4 = C_1 R_4 (R_1 + R_2) \quad (14)$$

L_4 must be larger than M . Further, the resistance R_4 associated with L_4 must be known accurately.

If m and n are the required values of the potentiometer setting x in the two adjustments, then R_e and L_e are computed from the equations

$$R_e = \frac{(1-m)}{m} R_s \quad (15a)$$

and

$$L_e = \frac{\sqrt{m-n}}{\omega m \sqrt{n}} R_s \quad (15b)$$

SPECIAL TECHNIQUES AND MEASUREMENTS

Quasi-Balanced Bridges

Unlike in a conventional bridge measurement where the repetitive adjustment of two elements in succession is needed, the quasi-balanced bridge requires adjusting only one element, leading to a rapid convergence to what is termed the *quasi-balance* condition. This condition is the detection of the minimum or maximum value of a voltage or even null detection in some forms. The quasi-balance, in principle, yields only one component of an unknown complex impedance. The second component, if needed, is obtained by adjusting another variable element for a different quasi-balance condition. The principles and some examples of quasi-balanced bridges are given in Ref. 9.

The microcontroller-based bridge described in Ref. 14 illustrates the application of the quasi-balance technique in conjunction with modern electronic hardware. Two independent quasi-balance conditions are obtained in succession in the bridge circuit of Fig. 13. The phasor $\overline{CIV}_d$ is brought into quadrature with the phasor $\overline{CIV}_R$ in the first adjustment and with the phasor $\overline{CIV}_s$ in the second. Both adjustments are effected by the potentiometer setting x . The quadrature condition is sensed by the zero output of a phase-sensitive detector (PSD).

A multiplying digital-to-analog converter (MDAC) is used to implement the potential divider action. A microcontroller senses the PSD output, adjusts the inputs to MDAC to secure quasi-balance conditions, measures the source frequency, and

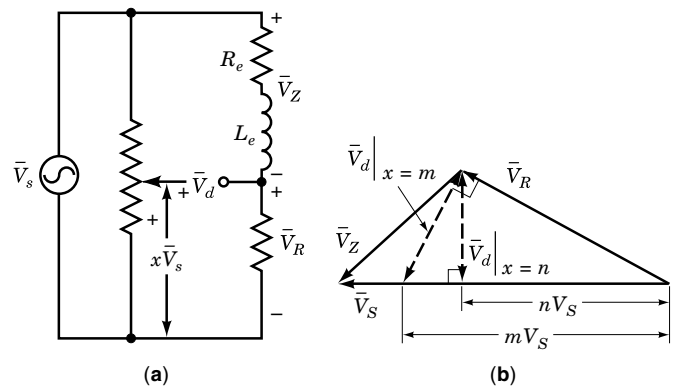


Figure 13. A quasi-balanced bridge measurement of the unknown impedance $R_e + j\omega L_e$. (a) Circuit configuration. The potentiometer setting x is used to obtain quasi-balance. (b) Phasor diagram. $\overline{CIV}_d$ is brought into quadrature with $\overline{CIV}_R$ and $\overline{CIV}_s$ for $x = m$ and $x = n$, respectively.

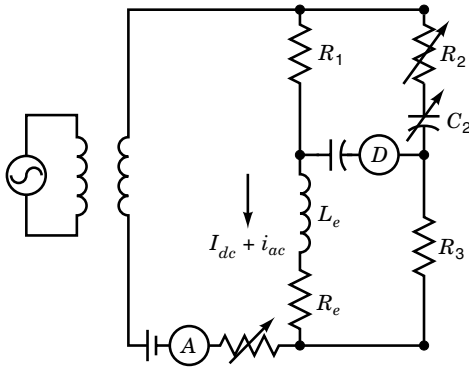


Figure 14. Hay's bridge circuit adapted for the measurement of incremental inductance.

finally displays digital values of R_e and L_e computed from the measurement.

Measurement of Incremental Inductance

A measurement of practical importance is the effective resistance and incremental inductance of iron-cored inductors when they carry a comparatively large direct current with a small superimposed alternating current, such as occurs with filter chokes and transformer windings in electronic circuits. The variation of incremental inductance as a function of the dc bias current needs to be determined. The Owen and Hay bridge circuits are well suited to introduce the dc current into the inductor branch, because the capacitors block the dc in two of the branches. Figure 14 shows the Hay bridge adapted to measure incremental inductance. The resistor in series with the dc supply is first varied to set up the required dc current in the inductor, as read by the dc ammeter A. This bias current also flows through R_1 , whose power rating should be appropriately chosen. The detector responds only to the ac signal because of the blocking capacitor in series with it. Now the bridge is balanced for ac signals by adjusting R_2 and C_2 . The following balance equations, which yield the values of incremental parameters, are the same as for the conventional Hay bridge:

$$L_i = \frac{C_2}{1 + \omega^2 C_2^2 R_2^2} R_1 R_3 \quad (16a)$$

and

$$R_i = \frac{\omega^2 C_2^2 R_2}{1 + \omega^2 C_2^2 R_2^2} R_1 R_3 \quad (16b)$$

where L_i is the incremental inductance and R_i the effective ac resistance at the chosen dc bias current. See Ref. 2 for further details.

Residual Inductance and Its Measurement

The balance equations of the bridge circuits in Table 1 are derived on the assumption that the standard elements used therein are pure. The impurities in these elements, called residuals, are no doubt small in practice, but must nevertheless be considered in high-precision measurements. Although the residuals of standard capacitors normally used can be ignored as negligibly small, the residual inductances associated with

resistors have to be considered, particularly when low-valued inductances are to be measured.

As an illustration of the methods used to tackle the problem of residuals, we consider the Maxwell–Wien bridge of Table I, modified by including a variable standard resistance R_4 in series with the test inductor. Let l_1 , l_2 , l_3 and l_4 be the residual inductance values of the resistors R_1 , R_2 , R_3 and R_4 and let C_2 be considered pure. Reworking the equations of the bridge and making appropriate approximations valid at medium frequencies, we obtain the following balance equation for L_e :

$$L_e = C_2 R_1 R_3 + \frac{1}{R_2} \left[l_1 R_3 + l_3 R_1 - l_2 \frac{R_1 R_3}{R_2} \right] - l_4 \quad (17)$$

Now let the test inductor be removed from the bridge and replaced by a resistor with a resistance value approximately equal to R_e and with a calculable inductance L_s . Such resistors are readily constructed as a loop of two, parallel, high-resistance wires (see pp. 147–153 of Ref. 2). Now the bridge is rebalanced by adjusting C_2 to C'_2 and R_4 to R'_4 . Because all other elements and their residuals remain unchanged,

$$L_s = C'_2 R_1 R_3 + \frac{1}{R_2} \left[l_1 R_3 + l_3 R_1 - l_2 \frac{R_1 R_3}{R_2} \right] - l'_4 \quad (18)$$

where l'_4 is the residual inductance of R'_4 . From Eqs. (17) and (18),

$$L_e = L_s + (C_2 - C'_2) R_1 R_3 + (l'_4 - l_4) \quad (19)$$

Because the substituted resistor has approximately the same resistance as the test inductor, the resistance values R_4 and R'_4 should be approximately equal, and the quantity $(l'_4 - l_4)$ can be neglected. Thus the value L_e is given, independently of the residuals of the bridge elements, by

$$L_e = L_s + (C_2 - C'_2) R_1 R_3 \quad (20)$$

For a detailed treatment of the solution of problem of residuals, consult Refs. 1, 2, and 6.

The substitution principle, which forms the basis of the foregoing procedure, is one of the most effective and accurate tools available, when the original measurement is beset by influencing factors of unknown magnitude. The method outlined in this section is also used to evaluate the residual inductance of a resistor, by treating the given resistor itself as the test inductor.

In-Circuit Measurement

It is useful to measure impedance values of elements already connected in a network without dismantling them. Such measurements, termed *in-circuit* or *in-situ* measurements, are called for in identifying, testing, and trouble-shooting electronic circuits. The shunting effect of other elements in the network has to be avoided in this evaluation.

In one approach (15), a current is injected into the element from an external sinusoidal voltage source. One of the leads to the element carrying this current and another wire carrying an adjustable opposing current derived from the same source through an auxiliary circuit are passed through a clamp-on type

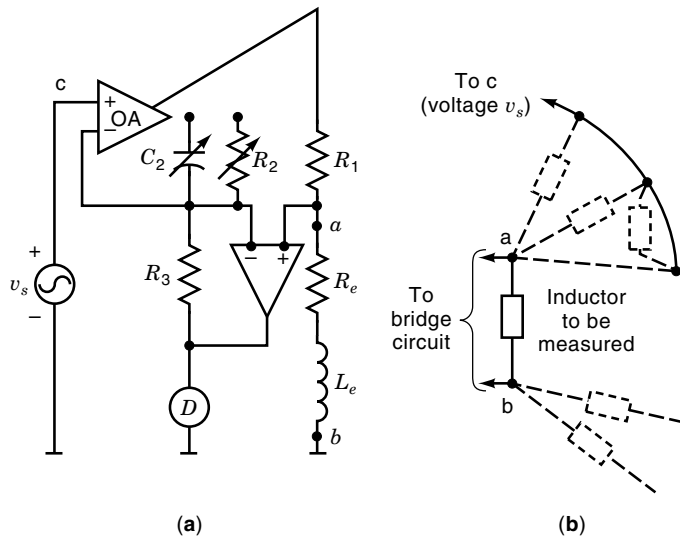


Figure 15. In-circuit measurement of an inductor. (a) The active bridge circuit employed. $R_e = R_1 R_3 / R_2$ and $L_e = R_1 R_3 C_2$ at balance. (b) Network modifications for the in-circuit measurement of an inductor connected between nodes a and b in a network.

current probe. The circuit constants of the auxiliary circuit, adjusted to obtain a null output from the probe, are used to evaluate the parameters of the element under test.

A second approach is to block the currents in the disturbing network elements by forcing zero voltage on them or to make these currents inconsequential and then measure the desired impedance by a suitable method (16,17). The active bridge circuit of Fig. 15, discussed in Ref. 17, apart from other advantages, enables the in-circuit measurement of an inductor at any specified voltage. To this end, the source v_s is adjusted to the desired voltage and the test inductor is connected into the bridge with all nodes in the network adjacent to one terminal of the test inductor lumped together and connected to the source as shown in Fig. 15(b). This method is applicable only when no other element exists in the network, which is connected directly in parallel with the test inductor.

BIBLIOGRAPHY

1. F. K. Harris, *Electrical Measurements*, New York: Wiley, 1966.
2. B. Hague and T. R. Foord, *Alternating Current Bridge Methods*, 6th ed. London: Pitman, 1971.
3. C. H. Dix and A. E. Bailey, Electrical standards of measurement—Part 1: D. C. and low frequency standards, *Proc. IEE*, **122**: 1018–1036, 1975.
4. C. H. Page, A new type of computable inductor, *J. Res. Nat. Bur. Stand.*, **67B**: 31–39, 1963.
5. F. E. Terman and J. M. Petit, *Electronic Measurements*, 2nd ed. New York: McGraw-Hill, 1952.
6. M. B. Stout, *Basic Electrical Measurements*, 2nd ed. New Delhi: Prentice-Hall of India, 1973.
7. E. W. Golding and F. C. Widdis, *Electrical Measurements and Measuring Instruments*, 5th ed. London: Pitman, 1963.
8. H. P. Hall, Impedance measurement, in B. M. Oliver and J. M. Cag, eds., *Electronic Measurements and Instrumentation*. New York: McGraw-Hill, 1971, Chap. 9.

9. K. B. Karandeev, *Bridge and Potentiometer Methods of Electrical Measurements* (English translation by B. Kuznetsov), Moscow: Peace Publishers, 1966.
10. B. A. Gregory, *An Introduction to Electrical Instrumentation*. London: Macmillan, 1973.
11. M. T. Sanderson, Electrical measurements, B. E. Noltingk, ed., *Instrumentation Reference Book*. Oxford: Butterworth-Heinemann, 1995, Part 3, Chap. 1.
12. K. M. Ibrahim and M. A. H. Abdul Karim, Digital impedance measurement by generating two waves, *IEEE Trans. Instrum. Meas.*, **34**: 2–5, 1985.
13. M. A. Atmanand, V. Jagadeesh Kumar, and V. G. K. Murti, A novel method of measurement of L and C , *IEEE Trans. Instrum. Meas.*, **44**: 898–903, 1995.
14. M. A. Atmanand, V. Jagadeesh Kumar, and V. G. K. Murti, A microcontroller-based quasi-balanced bridge for the measurement of L , C , and R , *IEEE Trans. Instrum. Meas.*, **45**: 757–761, 1996.
15. R. L. Forgacs, In-circuit impedance measurement using current sensing, *IEEE Trans. Instrum. Meas.*, **34**: 6–14, 1985.
16. M. Rahman, M. T. Ahmed, and V. G. K. Murti, A TRA bridge technique for in-circuit impedance measurement, *IEEE Trans. Instrum. Meas.*, **33**: 252–256, 1984.
17. V. G. K. Murti, Application of the generalised impedance converter in electrical instrumentation. *IEEE TENCON Conf. Proc.*, Singapore, 1984.

V. JAGADEESH KUMAR

V. G. K. MURTI

P. SANKARAN

Indian Institute of Technology

CAPACITANCE MEASUREMENT

Capacitance is the property that exists between two charged bodies separated by a dielectric material. The capacitance C is equal to the ratio of the absolute value of the charge Q to the absolute value of the voltage between charged bodies as

$$C = Q/V \quad (1)$$

where C is the capacitance in farads (F), Q is charge in Coulombs (C) and V is voltage (V).

The unit of capacitance is the Farad, which is a very large unit; therefore, practical capacitors have capacitances of microfarads (μF or 10^{-6} F), nanofarads (nF or 10^{-9} F), and picofarads (pF or 10^{-12} F).

In general, the capacitance can be determined by using Laplace's equations $\nabla^2 V(x, y, z) = 0$ with appropriate boundary conditions. The boundary conditions specify the electrode voltages V_1 and V_2 of the plates. For two electrodes, Laplace's equation yields the voltage $V(x, y, z)$ and the electric field $E(x, y, z) = -\nabla V(x, y, z)$. The charge of each electrode can also be obtained by integration of the flux density over each electrode surface as

$$Q = \int \epsilon(x, y, z) \mathbf{E}(x, y, z) \cdot d\mathbf{A} \quad (2)$$

If the capacitor is made from two parallel plates, as shown in Fig. 1, the capacitance value in terms of dimensions may be

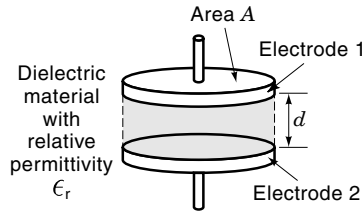


Figure 1. A typical capacitor made from two parallel plates. The capacitance between two charged bodies depends on the permittivity of the medium, distance between the bodies, and the effective area. It can also be expressed in terms of the absolute values of the charge and the absolute values of the voltages between bodies.

expressed by

$$C = \epsilon A/d = \epsilon_r \epsilon_0 A/d \quad (3)$$

where ϵ is the dielectric constant or permittivity, ϵ_r is the relative dielectric constant (in air $\epsilon_r = 1$), ϵ_0 is the dielectric constant of vacuum (8.854188×10^{-12} F/m or $10^{-9}/36\pi$ F/m), d is the distance between the plates in m, and A is the effective area of the plates in m^2 .

As can be seen in Eq. (3), the value of the capacitance is largely dependent on the permittivity of the dielectric material used. The permittivity of commonly used materials in the construction of capacitors, is listed in Table 1.

The capacitance C depends on the size and shape of charged bodies and their positioning relative to each other as shown in Table 2. In many electric and electronic systems, it is necessary to deal with the useful capacitances as well unwanted components. In many circuits, stray capacitances are introduced externally or internally at various stages. Cables and other external components introduce additional capacitances that need to be dealt with for desirable performance of the system. In these cases, Table 1 is useful to identify and analyze possible sources of additional capacitances where charged bodies are involved.

Characteristics of Capacitors

Capacitors are characterized by dielectric properties, breakdown voltages, temperature coefficients, insulation resistances, frequency and impedances, power dissipation, and quality factors: reliability, aging, etc.

The dielectrics of capacitors can be made from polar or non-polar materials. Polar materials have dipolar characteristics; that is they consist of molecules whose ends are oppositely charged. This polarization causes oscillations at certain frequencies resulting in high losses at those frequencies.

Table 1. Permittivity (Dielectric Constants) of Materials Used in Capacitors

Material	Permittivity
Vacuum	1.0
Air	1.0006
Polythene etc.	2.0 to 3.0
Impregnated paper	4.0 to 6.0
Glass and Mica	4.0 to 7.0
Ceramic (low K)	to 20.0
Ceramic (medium K)	80.0 to 100.0
Ceramic (high K)	1000.0 up

Dielectric absorption and leakage currents occur due to properties of dielectric materials. Absorption introduces a time lag during the charging and discharging of capacitors, thus reducing the capacitance values at high frequencies and causing unwanted time delays, particularly in pulse circuits. The leakage current, on the other hand, prevents indefinite storage of energy in the capacitor. An associated parameter of leakage currents is leakage resistance, which is measured in megohms but is usually expressed in megohm-microfarads or ohms-farads. The leakage resistance and capacitance introduce time constants which can vary from a few days for polystyrene to several seconds in some electrolytic capacitors. The leakage current is not only dependent on the properties of the dielectric materials but also on the construction and structure of the capacitors, and this is particularly true for capacitors having values less than $0.1 \mu\text{F}$ with very thin dielectric material between the electrodes.

If the capacitor is subjected to high operating voltages, the electric field in the dielectric exceeds the breakdown value and damages the dielectric permanently. The dielectric strength depends on the temperature, the frequency, and the applied voltage. An example of this dependence on the applied voltage is given in Fig. 2. It is commonly known that the use of capacitors below their rated values increases the reliability and their expected life time. The standard voltage ratings of most capacitors are quoted by the manufacturers as 50, 100, 200, 400, and 600 V. Tantalum and electrolytic capacitors have ratings of 6, 10, 12, 15, 20, 25, 35, 50, 75, 100 V, and higher.

Usually, values of surge voltages are given to indicate the ability of capacitors to withstand high transients. Typically, the surge voltages for electrolytic capacitors are 10% above the rated voltage, 50% for aluminum capacitors, and about 250% for ceramic and mica capacitors. The rated reverse voltages of electrolytic capacitors are limited to 1.5 V and, in some cases, to 15% of the rated forward voltages.

The temperature characteristics of capacitors are largely dependent on the temperature properties of the dielectric materials used, as given in Fig. 3. The temperature coefficients of

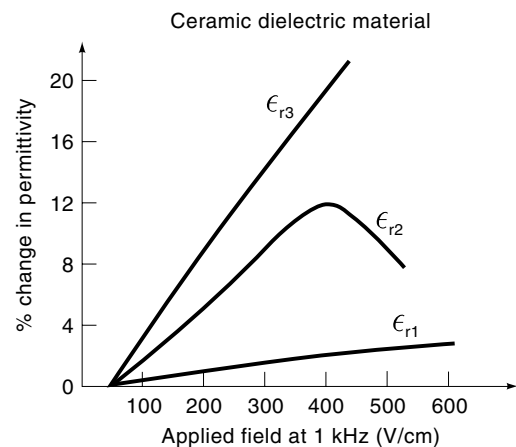
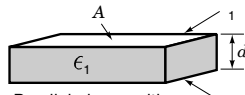
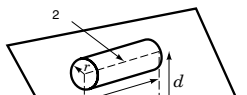
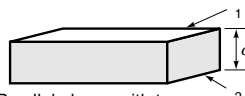
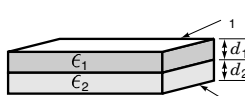
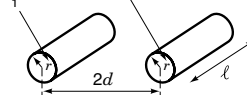
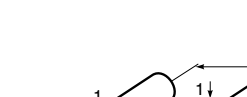
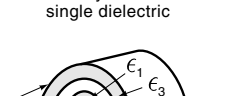
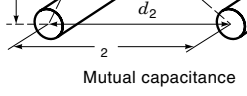


Figure 2. Changes in relative permittivity against field strength. The dielectric strength depends on the temperature, frequency, and applied voltage. Increases in the applied voltage cause higher changes in the dielectric strength. If the capacitor is subjected to higher operating voltages, the electric field in the dielectric exceeds the breakdown value which can damage the dielectric permanently.

Table 2. Capacitances of Various Electrode Systems

 Parallel plates with homogeneous dielectric $C = \frac{\epsilon A}{d}$	 $C = \frac{\pi \epsilon \ell}{\ln \left[\frac{d}{r} + \sqrt{\left(\frac{d}{r} \right)^2 - 1} \right]} = \frac{2\pi \epsilon \ell}{\cosh^{-1} \left(\frac{d}{r} \right)}$
 Parallel plates with two constituent dielectrics $C = \frac{(\epsilon_1 r + \epsilon_2) A}{(1 + r) d}$ where r value is the volumetric ratio	
 Parallel plates with two different dielectrics $C = \frac{\epsilon A}{\frac{d_1}{\epsilon_1} + \frac{d_2}{\epsilon_2}}$	 Parallel cylinders with equal radius $C = \frac{\pi \epsilon \ell}{\ln \left[\frac{d}{r} + \sqrt{\left(\frac{d}{r} \right)^2 - 1} \right]} = \frac{2\pi \epsilon \ell}{\cosh^{-1} \left(\frac{d}{r} \right)}$
 Coaxial cylinders with single dielectric $C = \frac{4\pi \epsilon \ell}{\ln(r_2/r_1)}$	 Mutual capacitance $C_{11} = \frac{\pi \epsilon \ell}{\ln(d_1/r)} + \frac{\pi \epsilon}{2} \frac{\ln(b/a)}{[\ln(a/r)] \ln(b/r)}$ $C_{21} = C_{12} = \frac{\pi \epsilon \ell}{2} \frac{\ln(b/a)}{[\ln(a/r)] \ln(b/r)}$
 Coaxial cylinders with three dielectrics $C = \frac{2\pi \epsilon_0 \ell}{\ln \left[\left(\frac{r_2}{r_1} \right)^{\frac{\epsilon_3}{\epsilon_1}} \times \left(\frac{r_3}{r_2} \right)^{\frac{\epsilon_2}{\epsilon_2}} \times \left(\frac{r_4}{r_3} \right)^{\frac{\epsilon_3}{\epsilon_3}} \right]}$ $= \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \right)^{-1}$	 Concentric spheres $C = \frac{4\pi \epsilon r_1 r_2}{r_2 - r_1}$

glass, teflon, mica, polycarbonate, etc. are very small, whereas in ceramic capacitors, they can be very high.

The insulation resistance of capacitors is important in many circuits. The insulation resistance is susceptible to temperature and humidity. Particularly, unsealed capacitors show

large and rapid changes against temperature and humidity. For most capacitors, at high temperatures, the change in insulation resistance is an exponential function of temperature ($R_{T1} = R_{T2} e^{K(T1-T2)}$). The temperature dependence of insulation resistance of common capacitors is shown in Fig. 4.

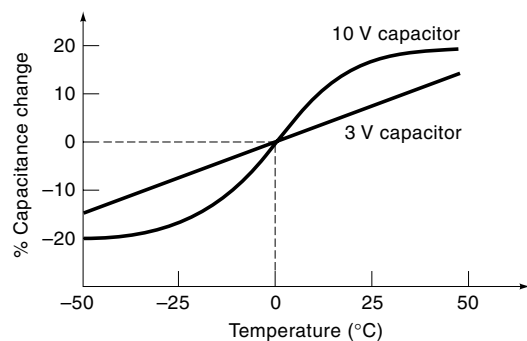


Figure 3. Temperature dependence of capacitors. The temperature characteristics of capacitors are largely dependent on the temperature properties of the dielectric materials used. The variations in capacitance due to temperature also depend on the type of capacitor and the operational voltage. The temperature coefficients of glass, teflon, mica and polycarbonate are very small, whereas in ceramic capacitors, they are relatively high.

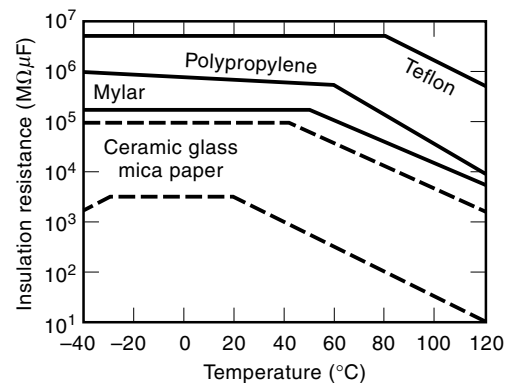


Figure 4. Temperature dependence of insulation resistance. The insulation resistance of many capacitors is not affected at low temperatures. However, under high temperature conditions, the change in insulation resistance may be approximated by an exponential relation. The insulation resistance is also susceptible to variations in humidity.

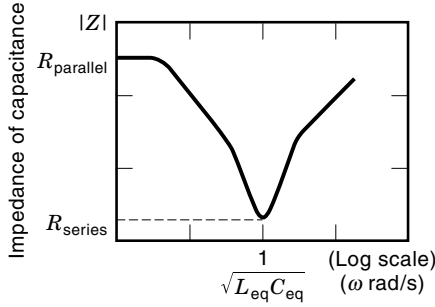


Figure 5. Frequency and impedance relation of capacitors. The losses and inherent inductance affect the ideal operation of capacitors, and the capacitance impedance becomes a function of frequency. Depending on their construction, all capacitors will resonate at a certain frequency.

Power Dissipation and Quality Factors. An ideal capacitor should have entirely negative reactance, but losses and inherent inductances prevent ideal operation. Depending on the construction, every capacitor resonates at certain frequencies due to equivalent resistance and inductances. A typical impedance characteristic of a capacitor is depicted in Fig. 5.

The electric equivalent circuit of a capacitor consists of a pure capacitance (C_p), plate inductances (L_1, L_2), plate resistances (R_1, R_2), and a parallel resistance R_p which represents the resistance of the dielectric called leakage resistance as shown in Fig. 6. The capacitors that have high leakage currents flowing through the dielectric have relatively low values for R_p . Very low leakage currents are represented by extremely large values of R_p . Examples of these two extremes are electrolytic capacitors that have high leakage current (low R_p), and plastic film capacitors which have very low leakage current (high R_p). Typically, an electrolytic capacitor might easily have $R_p < 1 \text{ M}\Omega$, giving several microamperes of leakage current, while a plastic film capacitor could have a resistance greater than $100,000 \text{ M}\Omega$.

In some cases, it is usual to represent a low leakage capacitor (high R_p) by a series RC circuit, while those with high leakage (low R_p) are represented by a parallel RC circuit. However, when the capacitor is measured in terms of the series C and R quantities, it is usually desirable to convert them into the parallel equivalent circuit quantities. This is because the (parallel) leakage resistance best represents the quality of the capacitor dielectric.

An ideal capacitor stores energy without dissipating any power. However, due to equivalent resistances, R_{eq} , some power will be dissipated. The power factor of a capacitor may be

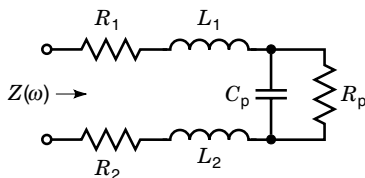


Figure 6. Capacitor equivalent circuit. A practical capacitor has resistances and inductances. Often, the electrical equivalent circuit of a capacitor can be simplified by a pure capacitance C_p and a parallel resistance R_p by neglecting resistances R_1, R_2 and inductances L_1, L_2 . In low leakage capacitors where R_p is high, the equivalent circuit may be represented by a series RC circuit.

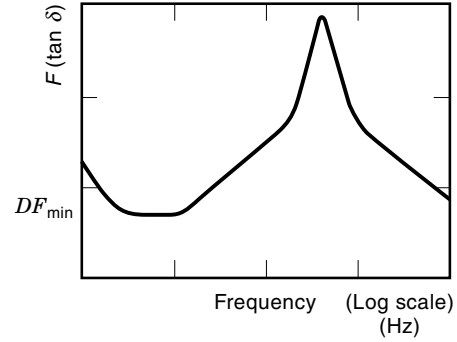


Figure 7. Power dissipation factors. In ideal operation, capacitors should store energy without dissipating power. Nevertheless, due to resistances, some power will be dissipated. The dissipation depends on frequency. The standard measurement of the dissipation factor δ is determined by applying 1.0 Vrms at 1 kHz .

expressed as

$$\text{PF} = \cos \theta = R_{eq} / |Z_{eq}| \quad (4)$$

where θ is the phase angle, and Z_{eq} is the equivalent total impedance.

One important characteristic, the dissipation factor of capacitors, is expressed as

$$\text{DF} = \tan \delta = R_{eq} / X_{eq} \quad (5)$$

where δ is the angle of loss, and X_{eq} is the equivalent reactance.

The dissipation factor is dependent on the frequency. Capacitors are designed such that this dependence is minimal. The measurement of the dissipation factor δ is made at 1 kHz and 1.0 Vrms applied to the capacitor. A typical dissipation factor curve is depicted in Fig. 7.

CAPACITIVE REACTANCE

Capacitors are used as two terminal elements, shown in Fig. 8(a), with the current voltage relationship given by

$$i(t) = C dv(t)/dt \quad (6)$$

From the $v(t), i(t)$ relationship, the instantaneous power of this element is then

$$p(t) = v(t)i(t) \quad (7)$$

The stored energy in the capacitor at time t seconds may be calculated as

$$\begin{aligned} w(t) &= \int C v(t) \{dv(t)/dt\} dt \\ &= [C v^2(t)]/2 \end{aligned} \quad (8)$$

Hence, the energy stored in the capacitor in the time interval t_1, t_2 will be equal to

$$W = (1/2)C[v^2(t_2) - v^2(t_1)] \quad (9)$$

If $v(t_2) = v(t_1)$, the energy stored in the capacitor in this time interval is equal to zero. This implies that the energy stored

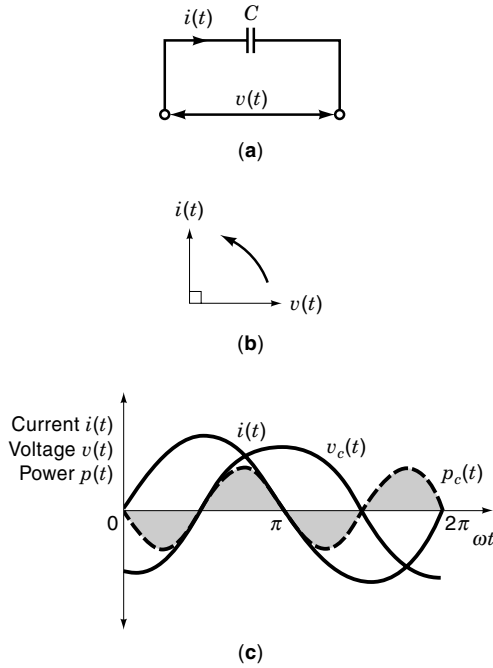


Figure 8. A capacitor as a two-terminal circuit element: (a) connection of a capacitor in electrical circuits, (b) current voltage relationship under sinusoidal operations, and (c) the power, voltage, and current relationships. The power has positive and negative values with twice the frequency of applied voltage.

in the capacitor is returned to the circuit during the time interval; that is, the capacitor transforms the energy without dissipation.

In circuits, capacitors form passive components. That is, if the voltage is known, the currents can immediately be determined by differentiation of the voltages. The Laplace transform gives the relationship

$$I(s) = sC V(s) \quad (10)$$

From this, the input-output relationship yields the impedance of the capacitor as

$$Z(s) = 1/sC$$

or

$$Y(s) = sC \quad (11)$$

In the stationary condition, $s \rightarrow 0$, $Z \rightarrow \infty$, and $Y \rightarrow 0$.

In the sinusoidal condition, $s = j\omega = 2\pi f$, where f = frequency, and hence,

$$\begin{aligned} Z(j\omega) &= 1/(j\omega C) \\ &= -j/(\omega C) \end{aligned} \quad (12)$$

and

$$Y(j\omega) = j\omega C \quad (13)$$

The capacitor can then be characterized, under the sinusoidal condition, by a reactance of $X_c = 1/\omega C$, measured in ohms. The current leads the voltage by 90° as shown in the phasor diagram in Fig. 8(b).

In sinusoidal operations, the instantaneous power $p(t) = v(t)i(t)$ can be calculated as

$$v(t) = V_{\max} \cos \omega t = \sqrt{2}V \cos \omega t \quad (14)$$

By using the relationship given by Eq. (6), the current can be written as

$$i(t) = C dv/dt = -\omega C \sqrt{2}V \sin \omega t \quad (15)$$

giving

$$p(t) = v(t)i(t) = -2\omega C V^2 \sin \omega t \cos \omega t = V^2 \sin 2\omega t / X_c \quad (16)$$

This indicates that the average power is zero because of the $\sin 2\omega t$ term, but there is a periodic storage and return of energy, and the amplitude of that power is V^2/X_c . The power, voltage, and current relationship in a capacitor is shown in Fig. 8(c).

Series and Parallel Connection of Capacitors

The formulas for series and parallel connections of capacitors, as shown in Fig. 9(a) and 9(b), respectively, can be obtained from a general consideration of series and parallel connections of impedances. For the series connection, the impedances are added such that

$$1/sC = 1/sC_1 + 1/sC_2 + \dots + 1/sC_n \quad (17)$$

where $C_1, C_2, \dots, C_n$ are the capacitances of the capacitors connected in series as in Figure 9(a). The equivalent capacitance is then given by

$$C = \{1/C_1 + 1/C_2 + \dots + 1/C_n\}^{-1} \quad (18)$$

That is, in series connection, the final capacitance value will always be smaller than the smallest value of the capacitance in the circuit.

In a similar way, the equivalent capacitance of parallel connected capacitors is

$$C = C_1 + C_2 + \dots + C_n \quad (19)$$

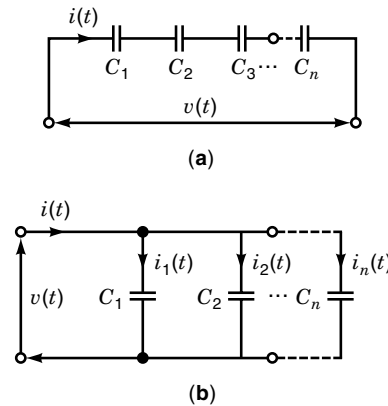


Figure 9. Series and parallel connection of capacitors: (a) series connection, and (b) parallel connection. In series connection, the final capacitance value will always be smaller than the smallest value of the capacitor as the circuit element, whereas in parallel connection, the final value is greater than the largest capacitance.

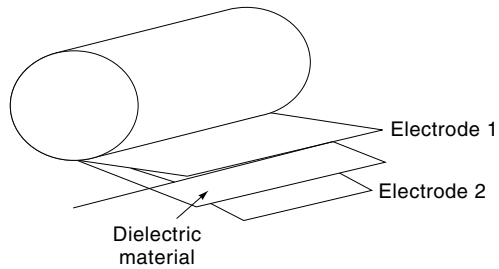


Figure 10. Construction of a typical capacitor. Dielectric sheets are placed between electrode foils and convolutely wound. The moisture of the dielectric material is removed at high temperatures by vacuum drying before the capacitor is impregnated with oil, paraffin, or wax.

and the final value of C is always larger than the largest capacitance in the circuit.

STANDARD CAPACITORS

Commonly used fixed capacitors are constructed with paper, mica, polymeric, and ceramic dielectric materials. Variable capacitors are generally made of air or ceramic dielectric materials. Capacitors can broadly be classified as electrolytic, ceramic, polymer, mica, variable, and integrated circuit capacitors.

Paper Capacitors

Usually, paper capacitors are made with thin ($5\ \mu\text{m}$ to $50\ \mu\text{m}$ thickness) wood pulp. A number of sheets are used together to eliminate possible chemical and fibrous defects that may exist in each sheet. The paper sheets are placed between thin aluminum foils and convolutely wound, as shown in Fig. 10. The moisture of the paper is removed by high temperature vacuum drying before the capacitor is vacuum impregnated with oil, paraffin, or wax. When impregnated with silicon oil, they can withstand voltages up to 300 kV.

Electrolytic Capacitors

This implies any capacitors in which the dielectric layer is formed by an electrolytic method. Electrolytic capacitors in dry foil form may be similar in construction to paper film capacitors; that is, two foil layers separated by an impregnated electrolyte paper spacer are rolled together. In this case, one of the plates is formed by metallization of one side of a flexible dielectric film. A foil, for example aluminum, is used as the other plate. The

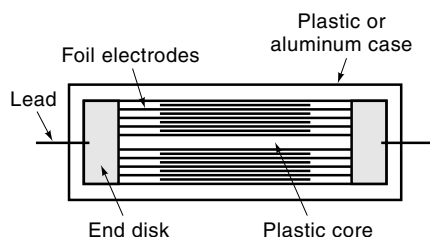


Figure 11. Construction of an electrolytic capacitor. The two foil layer electrodes are separated by an impregnated electrolyte paper spacer and rolled together on a plastic core. Usually, a flexible metallized dielectric film is used as one of the plates and an ordinary foil is used for the other. The capacitor is then hermetically sealed in an aluminum or plastic can.

capacitor is then hermetically sealed in an aluminum or plastic can, as shown in Fig. 11. These capacitors can be divided into two main subgroups:

Tantalum Electrolytic. The anode consists of sintered tantalum-powder, and the dielectric is Ta_2O_5 , which has a high value of ϵ_r . A semiconductor layer MnO_2 surrounds the dielectric. The cathode made from graphite is deposited around the MnO_2 before the capacitor is sealed. These capacitors are highly stable and reliable with good temperature ranges, and they are suitable in high frequency applications.

Aluminum Electrolytic Capacitors. Aluminum-foil is oxidized on one side as Al_2O_3 . The oxide layer is the dielectric, having a thickness of about $0.1\ \mu\text{m}$ and a high electric field strength ($7 \times 10^5\ \text{Vmm}^{-1}$). A second layer acting as a cathode made from etched Al-foil is inserted. The two layers are separated by a spacer; then, the layers are rolled and mounted.

Electrolytic capacitors must be handled with caution, since in these capacitors, the electrolyte is polarized. That is, the anode should always be positive with respect to the cathode. If not connected correctly, hydrogen gas will form, which damages the dielectric layer causing a high leakage current or blow-up. These capacitors may be manufactured to values up to $10,000\ \mu\text{F}$.

Ceramic and Glass Capacitors

The dielectric is a ceramic material with deposited metals. They are usually rod or disk shaped. They have good temperature characteristics and are suitable in many high frequency applications. There are many different types, such as

1. *Low K ceramic:* These capacitors are made with materials that contain a large fraction of titanium dioxide (TiO_2). The relative permittivity of these materials varies from 10 to 500, with negative temperature coefficients. The dielectric is $\text{TiO}_2 + \text{MgO} + \text{SiO}_2$, suitable in high frequency applications in filters, tuned circuits, coupling and bypass circuits, etc.
2. *High K ceramic:* The dielectric contains a large fraction of barium titanate, BaTiO_3 , mixed with PbTiO_3 or PbZrO_3 giving relative permeabilities of 250 to 10,000. They have high losses and also have high voltage time dependence with poor stability.
3. *Miniature ceramic capacitors:* These are used in critical high-frequency applications. They are made in the ranges of 0.25 pF to 1 nF.
4. *Dielectric ceramic capacitors:* The material is a semiconducting ceramic with deposited metals on both sides. This arrangement results in two depletion layers which make up the very thin dielectric. In this way, high capacitances can be obtained (22 nF to 100 nF). Due to thin depletion layers, only small dc voltages are allowed. They are used in small and lightweight equipment such as hearing aids.

Glass capacitors are made with dielectric glass materials. The properties of dielectric glass are similar to ceramic materials.

Polymer Capacitors

Various polymers such as polycarbonate, polystyrol, polystyrene, polyethylene, polypropylene, etc. are used as the dielectric. The construction is similar to that of paper

capacitors. Polystyrene capacitors, in particular, are very stable and are virtually frequency independent. They have low voltage rating and are used in transistorized applications as tuning capacitors.

Mica Capacitors

A thin layer of mica, usually muscovite mica (≥ 0.003 mm), is stapled with Cu-foil or coated with a layer of deposited silver. It is then vacuum impregnated and coated with epoxy. The field strength of these capacitors is very high (10^5 Vmm $^{-1}$), and their resistivity is $\rho = 10^6$ to 10^{15} Ω m. These capacitors are available in values from 1.0 pF to several microfarad for high voltage (from 100 V to 2000 V) and high frequency applications. They have tolerances between ± 20 to $\pm 0.5\%$.

Variable Capacitors

These capacitors usually have air as the dielectric and consist of two assemblies of space plates positioned together by insulation members, such that one set of plates can be rotated, as shown in Fig. 12. Their main use is in adjustment of the resonance frequency of tuned circuits in receivers and transmitters, filters, etc. By shaping the plates, various types of capacitances can be obtained, such as linear capacitance, in which capacitance changes as a linear function of rotation, and logarithmic capacitance.

Variable capacitors may be grouped as precision types, general-purpose types, transmitter types, trimmer types, and special types such as phase shifters.

Precision type variable capacitors are used in bridges, resonant circuits, and many other instrumentation systems. The capacitance swing can be from 100 pF to 5000 pF. They have good long term stability with very tight tolerances.

General-purpose type variable capacitors are used as tuning capacitors in radio and other broadcasting devices. They are available in many laws such as straight line frequency, straight line wavelength, etc. The normal capacitance swing is from 400 pF to 500 pF. In some cases, a swing of 10 to 600 pF is available.

Transmitter type variable capacitors are similar to general-purpose variable capacitors, but they are specially designed for high voltage operations. The swing of these capacitors can go from as few picofarads up to 1000 pF. The most common laws are liner frequency and straight line capacitances. In some

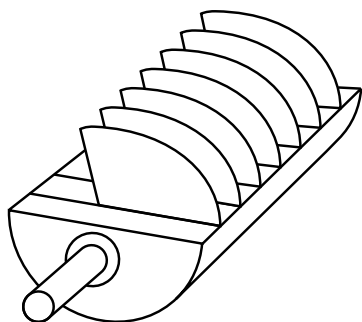


Figure 12. A variable capacitor. It consists of two assemblies of spaced plates positioned together by insulation members such that one set of plates can be rotated. The majority of variable capacitors have air as the dielectric. They are used mainly in adjustment of resonance frequency of tuned circuits in receivers and transmitters. By shaping the plates suitably, they can be made to be linear or logarithmic.

cases, oil filling or compressed gases are used to increase operating voltages.

Trimmer capacitors are used for coil trimming at intermediate radio frequencies. They can be air spaced rotary types (2 pF to 100 pF), compression types (1.5 pF to 2000 pF), ceramic-dielectric rotary types (5 pF to 100 pF), and tubular types (up to 3 pF).

INTEGRATED CIRCUIT CAPACITORS

There are many types of capacitors in IC technology. The most commonly used are depletion capacitors and thin-film capacitors. Junction capacitors are the easiest to form in integrated circuits and are primarily used as decoupling and bypass capacitors. The typical value of a junction capacitor is about 300 pF/mm ^{2} . Capacitors of the order of picofarads occupy large areas.

The thin-film capacitors are made mostly within MOS integrated circuits, as monolayer capacitors containing tantalum or other suitable deposits. The plates of integrated circuit capacitors are generally formed by two heavily doped polysilicon layers formed on a thick layer of oxide. The dielectric is usually made from a thin layer of silicon oxide. Important parameters for IC capacitors are the tolerances, voltage coefficients, temperature coefficients, and capacitance values. These capacitors are largely temperature stable with a temperature coefficient of about 20 ppm/ $^{\circ}$ C. The voltage coefficients are usually less than 50 ppm/V. Integrated circuit capacitive sensors are achieved by incorporating a dielectric sensitive to physical variables. Typical values of these capacitors are 1 pF to 27 nF for temperature compensating ceramic, 100 pF to 3000 pF for tantalum oxide, 390 pF to 0.47 μ F for general purpose ceramic, and 0.1 μ F to 10 μ F for tantalum electrolyte. Operating voltages range from 25 V to 200 V for ceramic, 12 V to 35 V for tantalum electrolyte, and 12 V to 25 V for tantalum oxide.

Integrated circuit capacitors are adapted for use in micro-electronic circuits. They include some miniature ceramic capacitors, tantalum oxide solid capacitors, and tantalum electrolyte solid capacitors. The ceramic and tantalum oxide chips are not encapsulated and are fitted with end caps for direct surface mounting onto the circuit board. Ceramic chip capacitors are most suitable for RF applications. Many different types of high- Q , multi-layer ceramic chip capacitors are offered by different manufacturers. They are available from 0.5 pF to 1000 pF, with voltage ratings from 100 V to 500V dc, depending on capacitance value.

The thin film integrated circuit computable capacitor may contain many capacitors with typical values of up to 1, 2, 4, 8, and 16 pF capacitance. Binary ratios starting at 0.25 pF are also available. In many designs, a common back metallization layer forms one plate of the capacitor, individual contacts form the other plate, and silicon dioxide dielectric features high Q over a wide temperature range. These capacitors are commonly used in circuits involving active filters, such as the Chebyshev or Butterworth filter (multiple feedback design) for high pass, low pass, stopband, or band pass applications.

CAPACITIVE MEASUREMENTS

Capacitive measurement techniques are extensively used in industrial and scientific applications. They are based on changes

in capacitances in response to physical property variations. These sensors find many diverse applications from humidity and moisture to displacement measurements.

Capacitive Displacement Measurements

The measurement of distances or displacements is an important aspect of many industrial, scientific, and engineering systems. Capacitive displacement sensors have high linearity and have wide ranges from a few centimeters to a few nanometers.

The basic sensing element of a typical displacement sensor consists of two simple electrodes with capacitance C . The capacitance is a function of the distance d (cm) between the electrodes of a structure, the effective surface area A (cm²) of the electrodes, and the permittivity ϵ ($=8.85$ pF/m for air) of the dielectric between the electrodes; therefore

$$C = f(d, A, \epsilon) \quad (20)$$

There are three basic methods for realizing a capacitive displacement sensor, that is, by varying d , A , or ϵ as discussed below.

Variable Area Displacement Measurements

The displacements may be sensed by varying the surface area of the electrodes of a flat plate capacitor, as illustrated in Fig. 13. In this case, the capacitance is

$$C = \epsilon_r \epsilon_0 (A - wx)/d \quad (21)$$

and the transducer output is linear with displacement x . This type of sensor is normally implemented as a rotating capacitor for measuring angular displacement. The rotating capacitor structures are also used as output transducers for measuring electric voltages as capacitive voltmeters.

Variable Distance Displacement Measurements

If a capacitor is made from two flat coplanar plates separated by a variable distance x , ignoring fringe effects, the capacitance can be found by

$$C(x) = \epsilon A/x = \epsilon_r \epsilon_0 A/x \quad (22)$$

where ϵ is the dielectric constant or permittivity, ϵ_r is the relative dielectric constant (in air and vacuum, $\epsilon_r \approx 1$), $\epsilon_0 =$

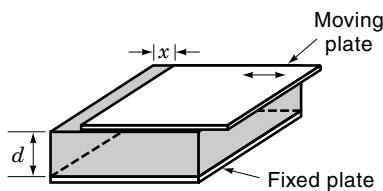


Figure 13. A variable area capacitive displacement sensor. The sensor operates on the variation in the effective area between the plates of a flat plate capacitor. The transducer output is linear with displacement x . This type of sensor is normally implemented as a rotating capacitor for measuring angular displacement.

8.854188×10^{-12} F/m ($10^{-9}/36 \pi$ F/m) is the dielectric constant of vacuum, x is the distance of the plates in m, and A is the effective area of the plates in m².

The capacitance is a nonlinear function of the distance x , and has the characteristics of a hyperbolic transfer function. The sensitivity of capacitance to changes in plate separation may be found as

$$dC/dx = -\epsilon_r \epsilon_0 A/x^2 \quad (23)$$

Equation (23) indicates that the sensitivity increases as x decreases. Nevertheless, using Eqs. (22) and (23), it can be proved that the percent changes in C are proportional to the percent changes in x which can be expressed as

$$dC/C = -dx/x \quad (24)$$

This type of sensor is often used for measuring small incremental displacements without making contact with the displaced object.

Differential Capacitive Measurements

These sensors are basically three terminal capacitors, as shown in Fig. 14. Slight variations of these sensors find many applications, including differential pressure measurements. In some versions, the central plate moves in response to physical variables with respect to the fixed plates. In others the central plate is fixed, and the outer plates are allowed to move. The output from the center plate is zero at the central position and increases as it moves left or right. The range is equal to twice the separation d . For a displacement d , we have

$$\begin{aligned} 2\delta C &= C_1 - C_2 = \epsilon_r \epsilon_0 lw/(d - \delta d) - \epsilon_r \epsilon_0 lw/(d + \delta d) \\ &= 2\epsilon_r \epsilon_0 lw \delta d / (d^2 + \delta d^2) \end{aligned} \quad (25)$$

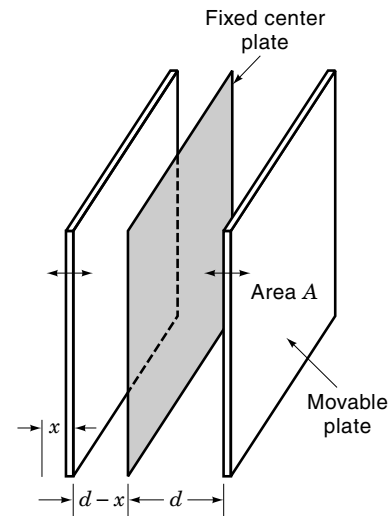


Figure 14. A differential capacitive sensor. This is a three-terminal capacitor with one fixed center plate and two outer plates. Its response to physical variables is linear. In some versions, the central plate moves in response to physical variables with respect to two outer plates; in the others, the central plate is fixed, and the outer plates are allowed to move.

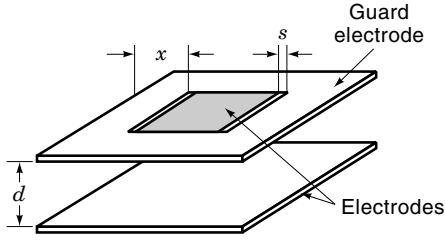


Figure 15. A typical smart capacitive position sensor. These types of microstructure position sensors contain three electrodes, two of which are fixed, and the third electrode moves infinitesimally relative to the others. Although the response is highly nonlinear, the integrated chip contains linearization circuits. They have a 0 to 1 mm measuring range with 1 μm accuracy.

and

$$\begin{aligned} C_1 + C_2 = 2C &= \epsilon_r \epsilon_0 l w / (d - \delta d) + \epsilon_r \epsilon_0 l w / (d + \delta d) \\ &= 2\epsilon_r \epsilon_0 l w d / (d^2 + \delta d^2) \end{aligned} \quad (26)$$

giving

$$\delta C / C = \delta d / d \quad (27)$$

This indicates that the device is inherently linear in contrast to the nonlinear response of the two-plate types. However, some nonlinearity can still be observed in practice due to defects in the structure.

Integrated Circuit Smart Capacitive Position Sensor

Figure 15 shows a typical microstructure capacitive displacement sensor. It has a 1 mm measuring range, 1 micron accuracy, and 0.1 s total measuring time. The specification corresponds to a possible range of 1 pF to 50 pF to displacement.

The relative deviation in the capacitance C_x between the two electrodes caused by the finite guard electrode can be expressed as

$$d < \epsilon^{-P} \frac{x}{d} \quad (28)$$

where x is the width of the guard, while d is the distance between the electrodes. Nonlinearity characteristics are exhibited in this deviation. Therefore, d is required to be less than 100 ppm.

In some special cases, the sensor capacitance C_x can be expressed as

$$C_x = \frac{\epsilon A_x}{d_0 + \Delta d} \quad (29)$$

where A_x is the area of the electrode, d_0 is the initial distance between them, ϵ is the dielectric constant, and Δd is the displacement to be measured.

Capacitive Pressure Sensors

A commonly used two-plate capacitive pressure sensor is made from one fixed metal plate and one flexible circular diaphragm as shown in Fig. 16. The flat circular diaphragm is clamped around its circumference, and is bent into a curve by an applied

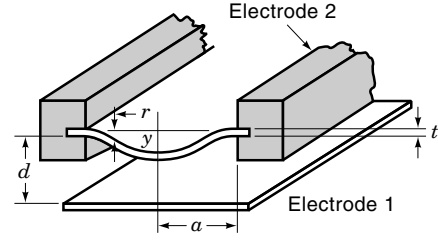


Figure 16. A capacitive pressure sensor. The pressure sensors are made from a fixed metal plate and a flexible diaphragm. The flat flexible diaphragm is clamped around its circumference. The bending of the flexible plate is proportional to applied pressure P . The deformation of the diaphragm results in changes in capacitance.

pressure P . The deflection y of this system at any radius r is given by

$$y = 3(1 - \nu^2)(a^2 - r^2)P / 16Et^3 \quad (30)$$

where

a is the radius of diaphragm
 t is the thickness of diaphragm
 E is the Young's modulus and
 ν is the Poisson's ratio

Deformation of the diaphragm means that the average separation of the plates is reduced. Hence, the resulting increase in the capacitance ΔC can be calculated by

$$\Delta C / C = (1 - \nu^2)a^4 P / 16Et^3 \quad (31)$$

where d is the initial separation of the plates, and C is the capacitance at zero pressure.

Capacitive Accelerometers and Force Transducers

Accelerometers based on capacitive sensing elements typically use the proof mass as one plate of the capacitor and the other plate as the base, as illustrated in Fig. 17. When the sensor

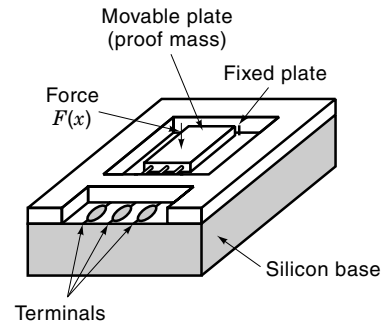


Figure 17. A capacitive force transducer. A typical capacitive micro-machined accelerometer has one of the plates as the proof mass. The other plate is fixed, thus forming the base. When the sensor is accelerated, the proof mass tends to move, thus varying the distance between the plates and altering the voltage across the capacitor. This change in the voltage is made to be directly proportional to the applied acceleration.

is accelerated, the proof mass tends to move, and the voltage across the capacitor changes. This change in the voltage corresponds to the applied acceleration.

In Fig. 17, let $F(x)$ be the positive force in the x direction. Neglecting all losses (due to friction, resistance, etc.), the energy balance of the system can be written for an infinitesimally small displacement dx , electrical energy dE_e , and field energy dE_f of the electrical field between the electrodes as:

$$dE_m + dE_e = dE_f \quad (32)$$

in which

$$dE_m = F(x)dx \quad (33)$$

also

$$dE_m = d(QV) = QdV + VdQ \quad (34)$$

If the supply voltage V across the capacitor is kept constant, it follows that $dV = 0$. Since $Q = VC(x)$, the Coulomb force is given by

$$F(x) = -V^2 \frac{dC(x)}{dx} \quad (35)$$

Thus, if the movable electrode had complete freedom of motion, it would have assumed a position in which the capacitance is maximal, and also if C were a linear function of x , the force $F(x)$ would be independent of x .

Capacitive silicon accelerometers are available in a wide range of specifications and sizes. A typical light weight sensor will have a frequency range of 0 to 1000 Hz with a dynamic range of ± 2 to $\pm 500 g$.

Capacitive Level Measurement

The level of a nonconducting liquid can be determined by capacitive techniques. This method is generally based upon the difference between the dielectric constant of the liquid and that of the gas or air above it. Two concentric metal cylinders are used for capacitance as shown in Fig. 18. The height of the liquid, l , is to be measured relative to the total height, h . Appropriate provision is made to ensure that the space between the cylindrical electrodes is filled by the liquid to the same height as the rest of the container. The usual operational conditions dictate that the spacing between the electrodes, s , should be much less than the radius of the inner electrode, r . Furthermore, the tank height should be much greater than r . When these conditions apply, the capacitance is approximately

$$C = \frac{\epsilon_1(l) + \epsilon_g(h-l)}{4.6 \log[1 - (s/r)]} \quad (36)$$

where ϵ_1 and ϵ_g are the dielectric constant of the liquid and gas (or air), respectively. The denominator of the above equation contains only terms that relate to the fixed system. A typical application is the measurement of the amount of gasoline in a tank in airplanes. The dielectric constant for most compounds commonly found in gasoline is approximately equal to 2, while that of air is approximately unity. These sensors are often incorporated with an ac deflection bridge.

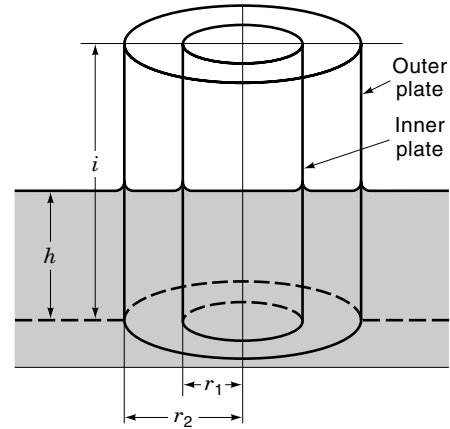


Figure 18. A capacitive liquid level sensor. Two concentric metal cylinders are used as electrodes of a capacitor. The value of the capacitance depends on the permittivity of the liquid and that of the gas or air above it. The total permittivity changes depending on the liquid level. These devices are usually applied in nonconducting liquid applications.

Capacitive Humidity and Moisture Sensors

The permittivity of atmospheric air as well as of many solid materials is a function of moisture content and temperature. The capacitive humidity device is based on the changes in the permittivity of the dielectric material between the plates of capacitors. There are many types of capacitive humidity sensors. Aluminum types and tantalum types will be introduced here as examples.

Aluminum Type Capacitive Humidity Sensors. The majority of capacitive humidity sensors are aluminum oxide type sensors. In these sensors, high purity aluminum is chemically oxidized to produce a prefilled insulating layer of partially hydrated aluminum oxide which acts as the dielectric. A water permeable but conductive gold film is deposited onto the oxide layer, usually by vacuum-deposition, and forms the second electrode of the capacitor.

Another type of aluminum-aluminum oxide sensor has a pore structure as illustrated in Fig. 19. The oxide with its pore

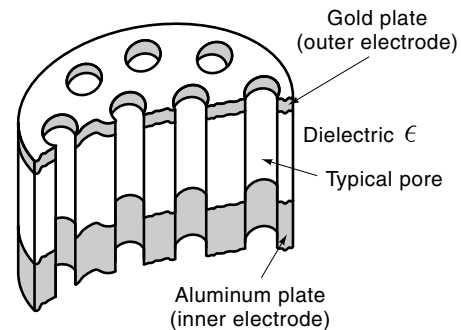


Figure 19. A typical capacitive humidity sensor. The sensors have pore or cracked mosaic structures for the moisture in air or gas to reach the dielectric material. The characteristics of the dielectric material change with the amount of water absorbed, thus reducing the resistance and increasing the capacitance. The quantity measured can be either resistance, capacitance, or impedance.

structure forms the active sensing material. Moisture in the air reaching the pores reduces the resistance and increases the capacitance. The decreased resistance can be thought of as being due to an increase in conduction through the oxide. An increase in capacitance can be considered to be due to an increase in the dielectric constant. The quantity measured can be either resistance, capacitance, or impedance. The high humidity end is best measured by capacitance, since resistance changes are very small in this region.

Tantalum Capacitive Humidity Sensors. In some versions of capacitive humidity sensors, one of the capacitor plates consists of a layer of tantalum deposited on a glass substrate. A layer of polymer dielectric is then added, followed by a second plate which is made from a thin layer of chromium. The chromium layer is under high tensile stress, so that it cracks into a fine mosaic structure which allows water molecules to pass into the dielectric. A sensor of this type has an input range of 0 to 100% relative humidity, RH. The capacitance is 375 pF at 0% RH and a linear sensitivity of 1.7 pF/%RH. The error usually is less than 2% due to nonlinearity and 1% due to hysteresis.

Capacitive humidity sensors enjoy wide dynamic ranges, from 0.1 ppm to saturation points. They can function in saturated environments for longer lengths of time, which would adversely affect many other humidity sensors. Their ability to function accurately and reliably extends over a wide range of temperatures and pressures. Capacitive humidity sensors also exhibit low hysteresis and high stability with little maintenance requirements. These features make capacitive humidity sensors viable for many specific operating conditions and ideally suitable for a system where uncertainty of unaccounted conditions exists during the operations. However, in some applications, contaminants can block the flow of water vapor into the sensor material, thus affecting the accuracy of the instrument. Many sensors come with some form of casing to provide protection.

Capacitive Moisture Sensors

Measurements in capacitive moisture sensors are based on the changes in the permittivity of granular or powder type dielectric materials such as wheat and other grains containing water. Usually, the sensor consists of a large cylindrical chamber, for example 150 mm deep and 100 mm in diameter. The chamber is filled with samples under test. The variation in capacitance values with respect to water content is processed. The capacitor is incorporated as a part of an oscillatory circuit operating at a suitable frequency.

Capacitive moisture sensors need to be calibrated for samples made from different materials, as the materials, themselves, demonstrate different permittivity. Accurate temperature is necessary as the dielectric constant may be highly dependent on temperature. Most of these devices are built to operate at temperature ranges of 0°C to 50°C, supported by tight temperature compensation circuits.

CAVITY RESONATORS

Cavity resonators are enclosures that confine electromagnetic energy in space as illustrated in Fig. 2(a). Cavity resonators find extensive use in microwave and laser optic applications. The energy is stored in the form of electric and magnetic energy

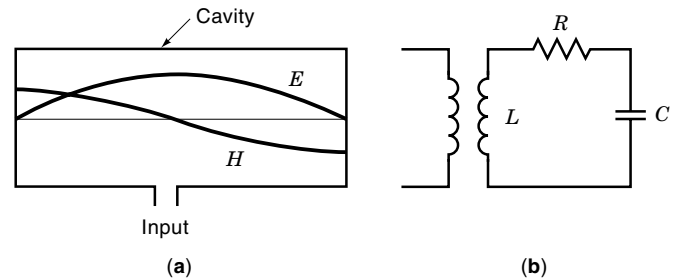


Figure 20. Cavity resonator and equivalent circuit. (a) A cavity resonator is a container confining electromagnetic energy in space. The resonant frequency is set by the particular field configuration and by the cavity dimensions. A resonance is obtained when the cavity length l is equal to the multiple of the half of the wavelength. (b) The energy is stored in electric and magnetic energy forms inside the cavity, largely determined by equivalent capacitance, inductance, and resistance. In its simplest form, the equivalent circuit can be viewed as an inductive coupling supplying the capacitor C .

inside the cavity, largely determined by equivalent capacitance, inductance, and resistance, very similarly to circuits made from coil and capacitance combinations. The coupling of the cavity resonator is equivalent to the inductive coupling as shown in Fig. 20(b).

Cavity resonators come in many shapes, from rectangular to circular or semicircular. The most commonly used cavity resonators are rectangular types, Fig. 20(a). The cavity shapes can be related to associated waveguide modes. With each mode, a resonant frequency may be determined by the particular field configuration involved and by the cavity dimensions. A given resonator has an infinite number of resonant modes which correspond to definite resonant frequencies. At resonant frequencies, the maximum amplitude of the standing wave occurs. The mode of the minimum resonant frequency is called the dominant mode and the higher resonant frequencies are called higher order modes. The modes in cavities are classified as transverse electric and transverse magnetic. In the case of rectangular prisms, if the wave is travelling in the l direction, resonance will be obtained whenever the multiple of half a wavelength is equal to the cavity length l .

The resonant frequency of a cavity resonator can be derived from Maxwell's equations for the electrical and magnetic fields within the resonator. In practical applications, the resonator is obtained first by constructing a resonator of convenient size and measuring the resulting resonant frequency for the construction. The ratio of the desired wavelength to the practical wavelength gives a scale factor that can be applied to every dimension of the test model. The resonant frequencies of a cavity resonator can be changed in a number of ways, such as (1) altering the mechanical dimensions, (2) changing the coupling reactance into the resonator, and (3) using copper paddles.

The Q factor of a cavity resonator is significant, as in the case of any other resonant circuits. The Q factor can be defined by

$$Q = 2\pi \frac{\text{Energy stored}}{\text{Energy loss per cycle}} \quad (37)$$

The energy stored is proportional to the square of the magnetic flux density integrated throughout the volume of the resonator.

The energy loss is proportional to the skin depth and to the square of the magnetic flux density integrated over the surface of the cavity. As a result, the Q factor depends on the ratio of volume to surface area of the resonator.

CAPACITANCE MEASUREMENTS

Null Bridge Instruments and Resonance Methods

Bridges are used to make precise measurements of unknown capacitances and associated losses in terms of known external capacitances and resistances. Most commonly used bridges are series-resistance-capacitance bridges, parallel-resistance capacitance bridges, Wien bridges, and Schering bridges.

The Series RC Bridge. Figure 21 is a Series-Resistance Capacitance Bridge which is used for the comparison of a known capacitance with an unknown capacitance. The unknown capacitance is represented by its series equivalent circuit C_x and R_x . A standard adjustable resistance R_1 is connected in series with a standard capacitor C_1 . The voltage drop across R_1 balances the resistive voltage drop when the bridge is balanced. The additional resistor in series with C_x increases the total resistive component, so that small values of R_1 will not be required to achieve balance. Generally, the bridge balance is most easily achieved when capacitive branches have substantial resistive components. To obtain balance, R_1 and either R_3 or R_4 are alternately adjusted. This type of bridge is found to be most suitable for capacitors with a high-resistance dielectric, hence very low leakage currents. At balance,

$$Z_1/Z_3 = Z_2/Z_4 \quad (38)$$

Substituting impedance values gives

$$\frac{(R_1 - j/\omega C_1)}{R_3} = \frac{R_x - j/\omega C_x}{R_4} \quad (39)$$

Solving for the real terms gives

$$R_x = R_1 R_4 / R_3 \quad (40)$$

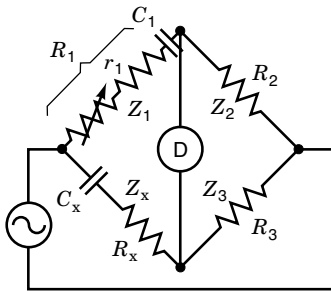


Figure 21. A series RC bridge. In these bridges, the unknown capacitance is compared with a known capacitance. The bridge balance is most easily achieved when capacitive branches have substantial resistive components. The resistors R_1 and either R_3 or R_4 are adjusted alternately to obtain the balance. This type of bridge is found to be most suitable for capacitors with a high-resistance dielectric, hence very low leakage currents.

and solving for the imaginary terms gives

$$C_x = C_1 R_3 / R_4 \quad (41)$$

An improved version of the series RC bridge is the substitution-bridge, which is particularly useful to determine the values of capacitances at radio frequencies. In this case, a series connected RC bridge is balanced by disconnecting the unknown capacitance and resistance C_x and R_x and replacing it by an adjustable standard capacitor C_s and adjustable resistor R_s . After having obtained the balance position, the unknown capacitance and resistance C_x and R_x are connected in parallel to the capacitor C_s . The capacitor C_s and resistor R_s are adjusted again for the rebalance of the bridge. The changes in the ΔC_s and ΔR_s lead to unknown values as

$$C_x = \Delta C_s \quad \text{and} \quad R_x = \Delta R_s (C_{s1} / C_x)^2 \quad (42)$$

where C_{s1} is the value of C_s in the initial balance.

Parallel-Resistance-Capacitance Bridge. Figure 22 illustrates a parallel-resistance capacitance bridge. In this case, the unknown capacitance is represented by its parallel equivalent circuit, C_x in parallel with R_x . The Z_3 and Z_4 impedances are pure resistors where either or both may be adjustable. The Z_2 impedance is balanced by a standard capacitor C_1 in parallel with an adjustable resistor R_1 . The bridge balance is achieved by adjustment of R_1 and either R_3 or R_4 . The parallel-resistance capacitance bridge is found to be most suitable for capacitors with a low resistance dielectric, hence relatively high leakage currents. At balance,

$$\frac{1/[1/R_1 + j\omega C_1]}{R_3} = \frac{1/[1/R_x + j\omega C_x]}{R_4} \quad (43)$$

Solving for real terms gives

$$R_x = R_1 R_4 / R_3 \quad (44)$$

and solving for the imaginary terms gives

$$C_x = C_1 R_3 / R_4 \quad (45)$$

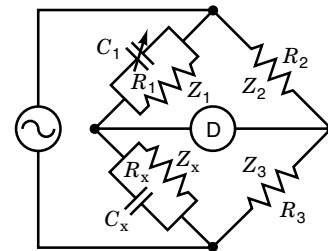


Figure 22. A parallel-resistance-capacitance bridge. The unknown capacitance is represented by its parallel equivalent circuit: C_x in parallel with R_x . The bridge balance is achieved by adjustment of R_1 and either R_3 or R_4 . The parallel-resistance capacitance bridge is found to be most suitable for capacitors with a low resistance dielectric, hence relatively high leakage currents.

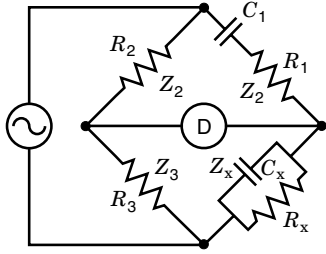


Figure 23. The Wien bridge. This bridge is used to compare two capacitors directly. It finds applications, particularly in determining the frequency in RC oscillators. In some cases, capacitors C_1 and C_x are made equal and ganged together, so that the frequency at which the null occurs varies linearly with capacitances.

The Wien Bridge. Figure 23 shows a Wien bridge. This is a special resistance-ratio bridge which permits two capacitances to be compared once all the resistances of the bridge are known. At balance, it can be proven that the unknown resistance and the capacitance are

$$R_x = R_3(1 + \omega^2 R_1^2 C_1^2) / (\omega^2 R_1 R_2 C_1^2) \quad (46)$$

and

$$C_x = C_1 R_2 / [R_3(1 + \omega^2 R_1^2 C_1^2)] \quad (47)$$

It can also be shown that

$$\omega^2 = 1 / R_1 C_1 R_x C_x \quad (48)$$

As indicated in Eq. (48), the Wien bridge has an important application in determining the frequency of RC oscillators. In frequency meters, C_1 and C_x are made equal, and the two capacitors are joined together so that the frequency at which the null occurs varies linearly with capacitances.

The Schering Bridge. Figure 24 illustrates the configuration of a Schering bridge. This bridge is used for measuring the capacitance, the dissipation factors, and the loss angles. The unknown capacitance is directly proportional to the known capacitance C_1 . That is, the bridge equations are

$$C_x = C_1 R_2 / R_3 \quad \text{and} \quad R_x = C_2 R_3 / C_1 \quad (49)$$

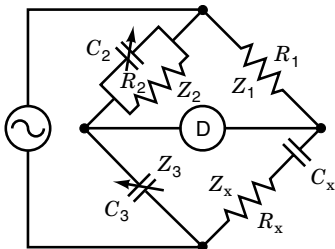


Figure 24. The Schering bridge. This bridge is particularly useful for measuring the capacitance, associated dissipation factors, and the loss angles. The unknown capacitance is directly proportional to the known capacitance C_1 . The Schering bridge is frequently used as a high-voltage bridge with a high-voltage capacitor as C_1 .

A Schering bridge is frequently used as a high-voltage bridge with a high-voltage capacitor as C_1 . It is also used as a high frequency bridge, since variable capacitors may be used for both adjustments.

Resonance Methods

Capacitors can be connected in parallel or series with inductors to form LC networks. In a series LC network, as the frequency, ω , increases the impedance of the circuit decreases. The decrease in the impedance reaches a point where the reactance becomes zero. That is

$$\omega_0 L - 1/\omega_0 C = 0 \quad \text{or} \quad \omega_0 = 1/\sqrt{LC} \quad (50)$$

The frequency ω_0 is called the resonant frequency of the circuit. The value of ω_0 is specified entirely in terms of the parameters of the energy storing elements of the circuit. In many instances, the unknown value of a capacitance is determined once the resonance frequency and the inductance are known.

ACCURACY AND SIGNAL PROCESSING

Generally, capacitive type measurements require relatively more complex circuitry in comparison to many other type of sensors, but they have the advantage of mechanical simplicity. They are also sensitive with minimum mechanical loading effects. For signal processing, these sensors are usually incorporated either in ac deflection bridge circuits or oscillator circuits. In practice, capacitive sensors are not pure capacitances but have associated resistances representing losses in the dielectric. This has an important influence in the design of circuits, particularly in oscillator circuits. Here, the most commonly used circuit based on charge amplifier is discussed.

Operational Amplifiers and Charge Amplifiers

One method of eliminating the nonlinearity in the relationship between the physical variable, for example two plate displacement sensors, and capacitance C is by the use of operational amplifiers, as illustrated in Fig. 25. In this circuit, if the input impedance of the operational amplifier is high, the output is not saturated, and the input voltage is small, it is

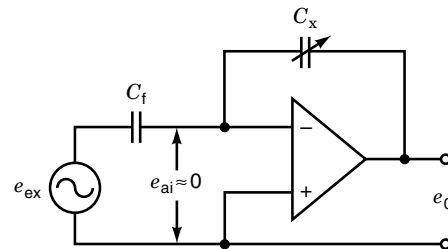


Figure 25. An operational amplifier signal processor. This method is useful to eliminate the nonlinearity in the signals generated by capacitive sensors. By this type of arrangement, the output voltage can be made to be directly proportional to variations in the signal representing the nonlinear operation of the device.

possible to write

$$1/C_f = \int i_f dt = e_{ex} - e_{ai} = e_{ex} \quad (51)$$

$$1/C_x = \int i_x dt = e_0 - e_{ai} = e_0 \quad (52)$$

$$i_f + i_x - i_{ai} = 0 = i_f + i_x \quad (53)$$

Manipulation of these equations gives

$$e_0 = -C_f e_{ex} / C_x \quad (54)$$

Substituting the value of C_x

$$e_0 = -C_f x e_{ex} / \epsilon A \quad (55)$$

Equation (55) shows that the output voltage is directly proportional to the plate separation x , thus giving linearity for all variations in motion.

However, a practical circuit requires a resistance across C_f to limit output drift. The value of this resistance must be greater than the impedance of C_f at the lowest frequency of interest.

A practical charge amplifier circuit is presented in Fig. 26. In this case, the effective feedback resistance R_{ref} is given by

$$R_{\text{ref}} = R_3(R_1 + R_2)/R_2 \quad (56)$$

It is possible to reduce the output drift substantially by suitably selecting the resistors. The accuracy of this circuit can be improved further by cascading two or more amplifiers. In this way, a substantial improvement in the signal-to-noise ratio can also be achieved.

In many applications, the resultant change in capacitance of capacitive transducers is measured with a suitable ac bridge, such as a Wien bridge or a Schering bridge. However, in a majority of cases, improvised versions of bridges are used as oscillator circuits for capacitive signal processing. The transducer is configured as part of the oscillatory circuit that causes changes in the frequency of the oscillations. This change in fre-

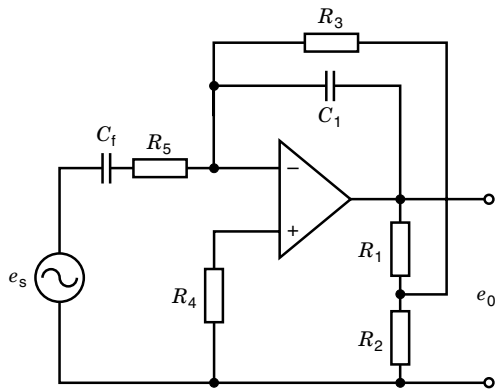


Figure 26. A practical charge amplifier. The effective feedback resistance is a function of other resistances. It is possible to reduce the output drift substantially by selecting the resistors suitably. The accuracy of this circuit can be improved further by cascading two or more amplifiers substantially, improving the signal-to-noise ratio.

quency is scaled to be a measure of the magnitude of the physical variable.

SELECTION, RELIABILITY, ACCURACY, AND STANDARDS OF CAPACITORS

Selection of Capacitors

The capacitors used in electronic circuits can be classified as low-loss, medium-loss, and high-tolerance capacitors. The low-loss capacitors have good stability, such as mica, glass, low-loss ceramic, and low-loss plastic film capacitors. These capacitors are expensive and used for precision applications, for example, telecommunication filters. Medium-loss capacitors are paper, plastic film, and medium-loss ceramic capacitors. Their applications include coupling, decoupling, bypass, energy storage, and some power electronic applications, for example, motor starting, lighting, power line applications, and interference suppressions. High tolerance capacitors, such as aluminum and tantalum electrolytic capacitors, deliver high capacitances. Although their sizes are relatively greater, they are reliable and have longer service life. They are used in polarized voltage applications, radio, television, other consumer goods as well as military equipment and harsh industrial environmental usage.

The values of general-purpose capacitors tend to be grouped closely together in a bimodal distribution manner within their tolerances. Usually, the tolerances of standard capacitors are 5, 10 and 20% of their rated values. Nevertheless, tolerances of precision capacitors are much tighter in the range of 0.25, 0.5, 0.625, 1, 2, and 3% of their rated values. These capacitors are much more expensive than those in the standard range.

For capacitors in the small pF range, the tolerances can be given as ± 1.5 , ± 1 , ± 0.5 , ± 0.25 , and ± 0.1 pF. Usually, low tolerance ranges do not include additional values; the capacitors' values simply have tighter distribution.

Capacitors and capacitive sensors are offered by many manufacturers, and some examples are listed in Table 3.

Reliability and Accuracy of Capacitors

Experience shows that a substantial part of component failures in electronic equipment is due to capacitors. The major cause of this is the improper selection and inappropriate application of capacitors. The following factors are, therefore, important criteria in the selection of capacitors in circuit applications: (1) The capacitance values and tolerances are determined by operating frequencies or by the value required for timing, energy storage, phase shifting, coupling, or decoupling. (2) The type and nature of the sources, such as ac, dc, transient, surges, and ripples determine the applied voltages. (3) The stability is determined by operating conditions like temperature, humidity, shock, vibration, and life expectancy. (4) Life expectancy, leakage current, dissipation factor, impedance, and self-resonant frequency determine electrical properties. (5) Types and construction, for example, size, configuration, and packaging determine mechanical properties.

Some of the common causes of capacitor failure are due to voltage and current overloads. The voltage overload produces an excessive electric field in the dielectric that results in its breakdown and destruction. The current overload caused by rapid voltage variations results in current transients. If these currents are of sufficient amplitude and duration, the dielectric

Table 3. List of Manufacturers

ANALITE Inc. 24-T Newtown Plaza Plainview, NY 11803 Tel: 1-800-229 3357	Magnetek 902 Crescent Avenue Bridgeport, CT 06607 Tel: 1-800-541 9997 Fax: 203-335 2820
Chenelex Barr Road, P.O. Box 82 Norwich, NY 13815 Tel: 607-344 3777 Fax: 607-334 9076	Maxwell Laboratories Inc. 8888 Balboa Avenue San Diego, CA, 92123 Tel: 619-576 7545 Fax: 619-576 7545
Comet North America Inc. 11 Belden Avenue Norwalk, CT 06850 Tel: 203-852 1231 Fax: 203-838 3827	Metuchen Capacitors Inc. 139 White Oak Lane Old Bridge, NJ 08857 Tel: 908-697 3366 or 1-800-679 0514 Fax: 1-800-679 9959
CSI Capacitors 810 Rancheros Drive San Marcos, CA 92069-3009 Tel: 619-747 4000 Fax: 619-743 5094	Murata Electronics Marketing Communications 2200 Lake Park Drive Smyrna, Georgia 30080 Tel: 1-800-394 5592 Fax: 1-800-4FAXCAT
FSI/FORK Standards Inc. 668 Western Avenue Lombard, IL 60148-2097 Tel: 708-932 9380	NWL Capacitors 204 Caroline Drive, PO Box 97 Snow Hill, NC 28580 Tel: 919-747 5943 Fax: 919-747 8979
Gordon Engineering Corp. 67 Del Mar Drive, Brookfield, CT 06804 Tel: 203-775 4501	Okaya Electric America Inc. 503 Wall Street, Valparaiso, Indiana 46383 Tel: 219-477 4488 Fax: 219-477 4856
Hecon Corp. 15-T Meridian Rd, Eatontown, NJ 07724 Tel: 1-800-524 1669	Rechner Electronic Industries Inc. 8651 Buffalo Avenue, Box 7 Niagara Falls, NY 14304 Tel: 1-800-544 4106
Kistler Instrumentation Corp. Amherst, NY 14228-2171 Tel: 716-691 5100 Fax: 716-691 5226	RDP Electrosense, Inc. 2216-Dept. B Pottstown, PA Tel: 1-800-334 5838
Locon Sensor Systems, Inc. 1750 S. Eber Road, PO Box 789 Holland, Ohio 43526 Tel: 419-865 7651 Fax: 419-865 7756	

can be deformed or damaged, resulting in drastic changes in capacitance values, thus leading to equipment malfunction. Overheating and high temperatures accelerate dielectric aging. This causes the plastic film to be brittle and also introduces cracks in the hermetic seals. Moisture and humidity due to a severe operating environment causes corrosion, reduces the dielectric strength, and lowers insulation resistances.

Standards and Computable Capacitors

Standards of Capacitors. Standard capacitors are constructed from interleaved metal plates using air as the dielectric material. The area of the plates and distance between them are determined and constructed with precision. The National Bureau of Standards maintains a bank of primary standard air capacitors that can be used to calibrate secondary and working standards for laboratories and industry. Generally, smaller capacitance working standards are obtained from air capacitors, whereas larger working standards are made from solid dielectric materials. Usually, silver-mica capacitors are selected as working standards. These capacitors are very stable, have very low dissipation factors, small temperature coefficients,

and very little aging effect. They are available in decade mounting forms.

Computable Capacitors. Another form of standard capacitors is the Thompson-Lampard calculable cross-capacitor. Thompson has shown that in cylindrical capacitors, the capacitance per unit length may be computed with good accuracy. The cylindrical cross-capacitor of Lampard and Thompson and the modified versions find wide use in many standards laboratories. There are many other types of standard calculable capacitors. For example, recently, new types of capacitors have been added to the classical-Kelvin parallel plates and coaxial cylindrical-calculable standards. Also, cubical capacitors, for which the direct capacitance between opposite faces can be estimated with high accuracy, are used as an independent check of the farad as established by means of a cross-capacitor.

The calculable capacitor is capable of determining the value of a 0.5 pF capacitor to within 2.2 parts in 10^8 . There are several computer-based models available that can be used to investigate the effect of changing the lengths and diameters of capacitors and their frequency dependence.

BIBLIOGRAPHY

1. J. P. Bentley, *Principles of Measurement Systems*, 2nd ed., Harlow, UK: Longman Scientific and Technical, 1988.
2. E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
3. D. Christiansen, (ed.), *Electronics Engineers' Handbook*, 4th ed., New York: McGraw-Hill, 1997.
4. L. J. Giacoletto (ed.), *Electronics Designers' Handbook*, 2nd ed., New York: McGraw-Hill, 1977.
5. J. P. Holman, *Experimental Methods for Engineers*, 5th ed., New York: McGraw-Hill, 1989.
6. F. F. Mazda (ed.), *Electronics Engineers' Reference Book*, 5th ed., London: Butterworths, 1983.
7. J. C. Whitaker, (ed.), *The Electronics Handbook*, Boca Raton, FL: CRC Press, 1996.
8. J. G. Webster, (ed.), *Measurement, Instrumentation and Sensors Handbook*, Boca Raton, FL: CRC Press, 1999.

HALIT EREN
Curtin University of Technology

Q-FACTOR MEASUREMENT

This article reports on traditional and modern methods of measurement of the Q -factors of resonators. Many of the techniques which will be described were first developed to determine the Q -factors of microwave resonators. Bearing this in mind, the language of the microwave engineer will be used throughout, although the reader should note that the basic principles are just as applicable to measuring the Q -factors of radio-frequency, millimeter, sub-millimeter, and optical resonators. The chapter will commence with a definition of the Q -factor, and a clear description of the resonator parameters which need to be measured to define the Q -factor of a resonance. The methods of measuring these parameters will then be outlined, including the particular advantages and disadvantages of a technique, the equipment required for the measurement, and the potential

for accuracy. Finally, some of the latest developments in the field of Q measurement will be presented.

Measurement of the Q -factor is the measurement of one of the defining characteristics of a resonance. Resonances are seen in devices which are able to store energy for a length of time which is long when compared to the period of oscillation of the electromagnetic energy being stored. The Q -factor is a direct measure of the quality of the resonator, or in other words, the length of time for which the energy can be stored.

To store energy, a resonator must have a method to trap energy. Microwave resonators make use of two main methods to trap energy: storage of the fields in a volume which is surrounded by conductive material, and storage inside a high dielectric material by total internal reflection at the dielectric surface. The first class of resonators are termed microwave cavities, while the second class of resonators are termed dielectric resonators. Of course, it is also possible to build hybrid resonators which use both mechanisms to trap energy. Any electromagnetic field pattern which satisfies Maxwell's equations in the resonator volume and satisfies the boundary conditions imposed by the resonator surfaces is called a normal mode or resonance of the resonator (1).

For energy to enter and leave a resonator, there must be some electrical or magnetic pathway between the resonator and an external transmission line. This pathway is most commonly provided in three ways:

1. Magnetic loop or electric probe on the end of a coaxial transmission line which protrudes into a cavity resonator,
2. A direct opening in the metallic wall between a cavity and a waveguide transmission line (iris coupling),
3. Placing a dielectric resonator in close proximity to a microstrip conductor (evanescent field or reactive coupling).

The number of coupling ports and the type of coupling used on those ports are used to further characterize microwave resonators. Resonators with a single port using probe or iris coupling are termed reflection resonators. Resonators with two or more ports using probe or iris coupling are termed transmission resonators. Two port resonators making use of reactive coupling are termed reactive or absorption resonators.

Microwave resonators, and accurate Q measurements to characterize those resonators, are vitally important in microwave engineering and physics:

1. Filters can be constructed from resonators because of their selective behavior in the frequency domain. A filter is a device which can transmit or reflect a desired signal while rejecting other nearby unwanted signals.
2. Measurements of electrical conductivity can be performed by an accurate measurement of the Q -factor of a microwave resonator which has been constructed of the material of interest (2,3). This measurement technique has undergone a renaissance recently with the great interest in high critical temperature superconductors (4).
3. Microwave properties of materials can be measured by the influences they have on the Q -factor and resonant frequency of a microwave resonance when they are loaded into a microwave resonator (1,5,6). Highly sensitive measurements of the Q and frequency are required to gain accurate material characterization. The dielectric loss and

dielectric constants of fluids and gases have been reported by a number of workers using this method (7,8). Magnetic and electrical properties of solids have also been reported (9,10,11,12).

4. Oscillators which have a high frequency stability in the short term require high- Q resonators; the best stability which has so far been achieved was with a superconducting cavity microwave oscillator operating near 2 K with a Q -factor around 10^{11} (13), and more recently, by a sapphire microwave oscillator at 4 K operating with a Q above 4×10^9 (14).

Microwave Q -factors of interest range from 2 (15) up to 10^{11} for a high perfection superconducting cavity at cryogenic temperatures (13). Coupling factors to these resonances can vary from 10^{-3} for extremely lightly coupled wavemeter resonators up to 10^3 for superconducting cavities at cryogenic temperatures. No single technique can be expected to operate over this full range and always give high resolution and highly accurate measurements. The desire to generate accurate measurements over this full range of interest has led to the large number of Q measurement methods reported here.

Many traditional methods for obtaining the Q factor have fallen from favor either because of improvements in technology, or because the original cumbersome technique can now be circumvented. For example, one of the major reasons for devising time domain Q measuring techniques was to avoid difficulties arising from the typically poor frequency stability of microwave signal generators of the time. However, modern quartz-based synthesizer technology provides a frequency stability which is adequate to measure the Q -factors of all but the very highest Q resonators. A second example is found with the rare use of voltage standing wave ratio (VSWR) measurement techniques in modern laboratories. These techniques are slow (16) and cumbersome as they require a device to be moved mechanically to make measurements. Modern network analyzers can make equivalent measurements essentially instantaneously without the need for mechanical changes to the circuit. Nevertheless, traditional Q measurement techniques still have importance because of their simplicity, low cost instrumentation, and instructional value. For these reasons, they will still be discussed below. For highly detailed instructions on performing these methods, the reader should refer to the classical texts (17,18).

DEFINITIONS

The letter Q was first used to describe the ratio of reactance to resistance of low frequency inductors and capacitors (18). Because of the usual technique of describing microwave resonators in terms of low frequency circuit elements, the letter Q became associated with describing the ratio of reactance in a resonant circuit to the resistance in that circuit. A simple calculation (see Eq. 3) shows that there is also a close relationship between Q and the time for energy to dissipate in such an equivalent circuit. This led to the incorrect but widespread belief that Q is the initial letter of the expression *Quality factor* (abbreviated as Q -factor), a term related to the length of time that a resonant circuit stores energy. The agreed modern definition of the Q -factor of any general resonant circuit (including electronic, optical, and mechanical resonators) is (18) 2π times

the ratio of energy stored in the system to the energy dissipated per cycle or in symbols,

$$Q = \frac{2\pi U}{\Delta U} \quad (1)$$

where U is the maximum energy of the system, and ΔU is the energy dissipated in a cycle. This equation can be solved to find how the energy decays as a function of time. The energy dissipated in one cycle can be found from

$$\Delta U = -\frac{\partial U}{\partial t} \frac{1}{f_0} \quad (2)$$

where f_0 is the resonant frequency. Substituting Eq. (2) into Eq. (1), the time dependence of U is found as

$$U(t) = U_0 e^{-(\omega_0/Q)t} \quad (3)$$

where U_0 is the initial energy of the resonant system, and $\omega_0 \equiv 2\pi f_0$ is the angular resonant frequency. From inspection of Eq. (3), another more practical definition of Q can be stated: 2π times the number of cycles required for the energy in the system to decay to $1/e$ of some initial amount or

$$Q = \frac{2\pi \tau}{T_0} \quad (4)$$

where τ is the time taken for the energy to decrease to $1/e$ of its value (called the ring-down time), and $T_0 \equiv 1/f_0$ is the resonant period. As will be seen below, direct use is made of Eq. 4 in determining Q factors by a time domain technique. All other measurement techniques make use of the response of the resonator to forced oscillation. To understand these techniques, we will need to derive expressions which will describe the behavior of the resonator under this condition.

Microwave Resonators

A diagram of a transmission microwave resonator is given in Fig. 1(a). The figure shows an empty metallic cavity where coupling between the resonator and external transmission lines is provided by small openings in the walls of the cavity. A general low frequency analogue to this microwave resonator, its coupling arrangement, and the external transmission lines is given by Fig. 1(b). Each resonance of the cavity is represented by a parallel LCR circuit (19). The k th resonant frequency is given by the expression:

$$(f_0)_k = \frac{1}{2\pi \sqrt{L_k C_k}} \quad (5)$$

and the unloaded Q of the k th resonance can be determined by substituting the energy dissipation provided by R_k and the energy storage provided by C_k and L_k into Eq. 1:

$$(Q_u)_k = R_k \sqrt{\frac{L_k}{C_k}} \quad (6)$$

The iris coupling has been modeled as transformer of unknown turns ratio (17,20). X_{e1} and R_{e1} are the equivalent reactance

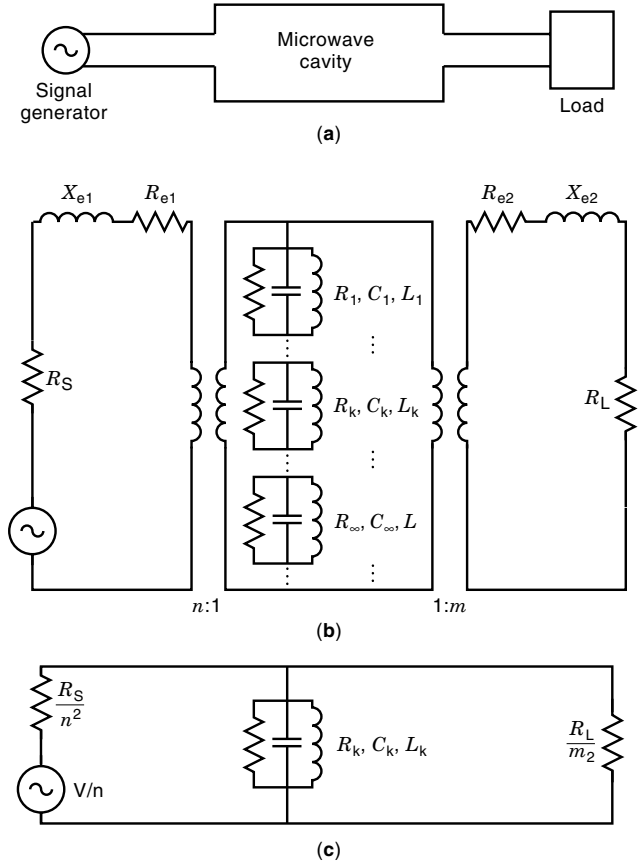


Figure 1. (a) A diagram of a general microwave cavity resonator with iris coupling on the input and output ports. (b) An equivalent LCR circuit for part (a). The cavity resonator has an infinite number of resonances which have been represented by the series of LCR circuits: $(R_1, C_1, L_1), \dots, (R_k, C_k, L_k), \dots, (R_\infty, C_\infty, L_\infty)$. The input and output coupling networks have been modeled with a series loss (R_{e1}, R_{e2}) and series inductance (X_{e1}, X_{e2}) to take into account any losses or reactance of the coupling network. The source and load series impedances are denoted by R_S and R_L . The iris coupling has been modeled as a transformer of unknown turns ratio. (c) A highly simplified version of part (b), where the traditional assumptions of Q -factor measurement have been made. The coupling reactance and loss are assumed to be zero. The source and load are assumed to have a resistance which is perfectly matched to the line impedance. All other resonances of the cavity are assumed to be sufficiently far from the resonance of interest that they can be safely ignored. All external components have been transformed across the transformers and into the resonant circuit.

and shunt resistance of the input coupling network and transmission line, while X_{e2} and R_{e2} are the equivalent parameters for the resonator output network. R_S is the series impedance of the signal generator, and R_L is the equivalent load resistance of the equipment that is monitoring the resonator. Any measurement of the resonant system represented by Fig. 1(b) will yield a result which is a combination of the properties of the transmission line, the coupling mechanism, and the resonator. However, under certain conditions, it is possible to measure the Q of a particular mode while ignoring the influences of the load, the coupling mechanism, the transmission line, and other modes. In this case, the external impedances can be transformed into the resonant circuit giving the circuit shown in Fig. 1(c). This circuit is far more amenable to analysis and is

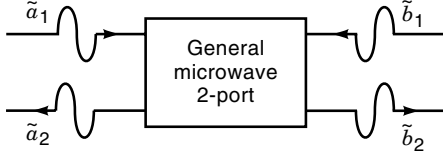


Figure 2. Schematic diagram representing the scattering of microwave voltage waves by a microwave two-port.

the model usually assumed by traditional measurement techniques as it greatly simplifies the post-measurement analysis [attempts were sometimes made to allow for coupling losses; i.e., $R_{e1} \neq 0$ (18)]. These simplifications will lead to important errors if the resonator is strongly overcoupled, the Q of the resonance is relatively low, the generator or source are poorly matched to the transmission line, or there is an excessive amount of reactance in the coupling mechanism or transmission line. For the sake of a simple introduction, we will use the circuit shown in Fig. 1(c) and reserve discussion of the more complex situation until the “Modern Measurement Techniques” section later.

To describe how microwave radiation interacts with a resonator, we will use a scattering matrix formalism. Figure 2 shows a voltage wave, $\tilde{a}_1$, incident on an arbitrary microwave two port device. The wave is scattered by the two port, some of its energy going into a reflected wave, $\tilde{a}_2$, and part into a transmitted wave, $\tilde{b}_2$. We express all three waves as

$$A \exp[j(\omega t - \beta z)] \quad (7)$$

where ω is the applied angular microwave frequency, A is the normalized voltage amplitude of the wave, β is the propagation constant, and z is the measurement position. We define the resonator scattering terms s_{11} and s_{21} as

$$s_{11} \equiv \frac{\tilde{a}_2}{\tilde{a}_1} \quad \text{and} \quad s_{21} \equiv \frac{\tilde{b}_2}{\tilde{a}_1} \quad (8)$$

We can, of course, define equivalent parameters for port 2. By substituting Eq. 7 into Eq. 8, we note that s_{11} and s_{21} will depend on the measurement position. This leads to the difficulty that we need to specify the measurement plane each time we quote the resonator scattering parameters. Theoretically, we could avoid this difficulty by always referring the measurement to some special pre-agreed measurement plane, for example, the plane which is coincident with the coupling device. However, practically, it is difficult to make a measurement at that physical position. Instead, we choose to define some special planes which have a well defined measurable characteristic. If a measurement is made of s_{11} at a frequency which is far from the resonant frequency, we can always find a position on the line which responds as if there were a short circuit or an open circuit on the transmission line. These planes are termed the *detuned short* or *detuned open* positions (17,18,20). From the detuned short position, the equivalent low frequency circuit shown in Fig. 1(c) is an accurate representation of the real microwave circuit (given that all the above assumptions still hold). Modern measurements can avoid this difficulty of determining the reference plane (see the Modern measurement techniques section). The scattering parameters at the detuned short position

for the circuit shown in Fig. 1(c) have been calculated as (21,22)

$$S_{11}(\epsilon) = \frac{\beta_1 - \beta_2 - 1 - jQ_u\epsilon}{1 + \beta_1 + \beta_2 + jQ_u\epsilon} \quad (9)$$

$$S_{21}(\epsilon) \equiv S_{12}(\epsilon) = \frac{2\sqrt{\beta_1 + \beta_2}}{1 + \beta_1 + \beta_2 + jQ_u\epsilon} \quad (10)$$

where $\epsilon \equiv f/f_0 - f_0/f$, f is the incident frequency, f_0 is the resonant frequency as defined above in Eq. 5, Q_u is the unloaded Q of the resonator as defined in Eq. 6, and the coupling factors are defined as the ratio of the internal resonant resistance to the external resistances when transformed into the resonant circuit; that is,

$$\beta_1 = \frac{n^2 R_k}{R_s} \quad \text{and} \quad \beta_2 = \frac{m^2 R_k}{R_L} \quad (11)$$

where n is the effective turns ratio of the input coupling mechanism, and m is the effective turns ratio of the output coupling mechanism. If $\beta < 1$, then a resonator is termed undercoupled; if $\beta > 1$, then a resonator is termed overcoupled; and if $\beta = 1$, then the resonator is critically coupled. Eq. 9 also applies to single port reflection resonators if β_2 is set to zero.

It is useful to split Eqs. 9 and 10 into their real and imaginary parts as this will give us the in-phase response and quadrature phase response of the resonator, respectively. In other words, the real part of Eq. 9 gives us the portion of the reflected wave which is in phase with the incident signal and the imaginary part will give the fraction of the reflected signal, which is $\pi/2$ out of phase with the incident signal.

$$\text{Real}[s_{11}] = \frac{(1 + \beta_1 + \beta_2)(\beta_1 - \beta_2 - 1) - \epsilon^2 Q_u^2}{(1 + \beta_1 + \beta_2)^2 + \epsilon^2 Q_u^2} \quad (12)$$

$$\text{Imag}[s_{11}] = \frac{-2\beta_1 \epsilon Q_u}{(1 + \beta_1 + \beta_2)^2 + \epsilon^2 Q_u^2} \quad (13)$$

$$\text{Real}[s_{21}] = \frac{2\sqrt{\beta_1 \beta_2}(1 + \beta_1 + \beta_2)}{(1 + \beta_1 + \beta_2)^2 + \epsilon^2 Q_u^2} \quad (14)$$

$$\text{Imag}[s_{21}] = \frac{-2\sqrt{\beta_1 \beta_2} \epsilon Q_u}{(1 + \beta_1 + \beta_2)^2 + \epsilon^2 Q_u^2} \quad (15)$$

An interesting and novel way of summarizing this information is by using a three dimensional parametric plot of the resonator response as a function of frequency (see Fig. 3 for a plot of s_{11}). If we make two dimensional projections of this plot along the principal axes, we will see response curves that are more familiar to the microwave engineer. To display this, we have projected the three dimensional plot onto the faces of the box in Fig. 3 and then rotated these faces into view. Viewing the curve which has been projected onto the plane perpendicular to the real axis we see the imaginary response of the resonator [Fig. 3(b)], looking along the imaginary axis, we observe the real response of the resonator [Fig. 3(c)], and finally, looking parallel to the frequency axis, we observe a circle—which is termed the Q circle of the resonator [Fig. 3(d)]. This horizontal projection plane is more readily recognized as the Smith Chart (23). The Smith Chart gives the magnitude of s_{11} by the distance that a plotted point is from the origin, while the phase of s_{11} (ϕ) is given by the angle that the line joining the point to the origin makes with the real axis. The most important advantage of

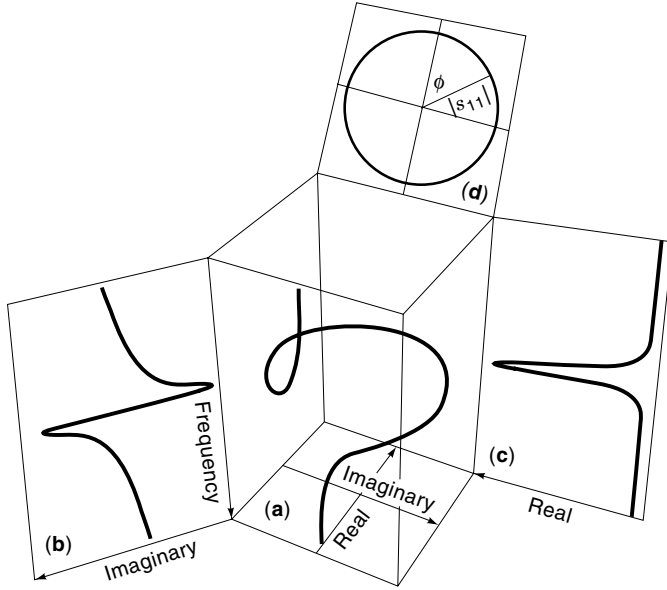


Figure 3. (a) A three-dimensional display of the scattering parameter, s_{11} , from a reflection resonator. The curve has been projected onto the faces of the box, and then these faces have been rotated into view. (b) Shows the imaginary (or quadrature phase) reflection response of the resonator. (c) Shows the real (or in-phase) reflection response of the resonator. (d) Shows the Q -circle of the resonator on a Smith Chart. The phase and magnitude of the reflection coefficient at a particular frequency are given by the distance that a plotted point is from the origin and the angle it makes with the real axis.

this three-dimensional description of the resonator response is the clear demonstration of the intimate relationship which exists between the many descriptions of microwave resonators given in the literature: the real and imaginary components of the scattering matrix, the Smith Chart, and the magnitude and phase of the response.

Q -factor measurements of two port microwave resonators frequently employ measurements of the transmitted and reflected power as a function of frequency. In cases of moderate to high Q , the derived frequency variable ϵ can be simplified with negligible error to

$$\epsilon = \frac{2\Delta f}{f_0} \quad (16)$$

where $\Delta f \equiv f - f_0$. We will term Δf the frequency detuning. The squared magnitude of s_{11} and s_{21} (which gives the reflected and transmitted power coefficients respectively) can be derived from Eqs. 9 and 10 as

$$|s_{11}|^2 = \frac{(1 - \beta_1 + \beta_2)^2 + 4 \left(\frac{Q_u \Delta f}{f_0} \right)^2}{(1 + \beta_1 + \beta_2)^2 + 4 \left(\frac{Q_u \Delta f}{f_0} \right)^2} \quad (17)$$

$$|s_{21}|^2 = \frac{4\beta_1\beta_2}{(1 + \beta_1 + \beta_2)^2 + 4 \left(\frac{Q_u \Delta f}{f_0} \right)^2} \quad (18)$$

Plots of these functions have been given on Fig. 4.

Loaded Q (Q_L), Unloaded Q (Q_U) and External Q (Q_e)

There are three distinct Q -factors of which we commonly talk when we discuss the losses in a microwave circuit. The unloaded Q (Q_U) is the Q -factor that would be measured as the electromagnetic coupling to the resonator tended to zero. That is, it is associated with dissipation which occurs inside the resonator (R_k in Fig. 1(c) and see Eq. 6). The external Q (Q_e) is the dissipation which occurs external to the resonator but which is coupled into the resonance by the coupling to the external circuit [R_s and R_L in Fig. 1(c)]. We can derive an expression for the external Q -factors in the same way as we found the unloaded Q factor (Eq. 6):

$$Q_{e1} = \frac{R_s}{n^2} \sqrt{\frac{C}{L}} \quad \text{and} \quad Q_{e2} = \frac{R_L}{m^2} \sqrt{\frac{C}{L}} \quad (19)$$

The loaded Q (Q_L) is the Q -factor measured by any scalar Q measurement method. It includes the contributions of all losses in the circuit which provide a dissipation mechanism for the resonance (both internal and external to the resonant circuit). We can write the relationship between the loaded Q , the unloaded Q , and the external Q factors as

$$\frac{1}{Q_L} = \frac{1}{Q_U} + \frac{1}{Q_{e1}} + \frac{1}{Q_{e2}} \quad (20)$$

Using the definitions of the coupling factors (Eq. 11) and the definitions of the unloaded and external Q factors (Eqs. 6 and 19), we can simplify Eq. 20 to

$$Q_u = (1 + \beta_1 + \beta_2) Q_L \quad (21)$$

General Q -Factor Measurement Techniques

Let us simplify some of our original expressions to generate some practical ways to measure the Q factor. From an examination of Eq. 17, we see that the maximum power transmitted through the resonator occurs when the incident frequency is exactly equal to the resonant frequency. As we move away from the resonant frequency, the transmitted power falls as the square of the frequency difference. An important measure

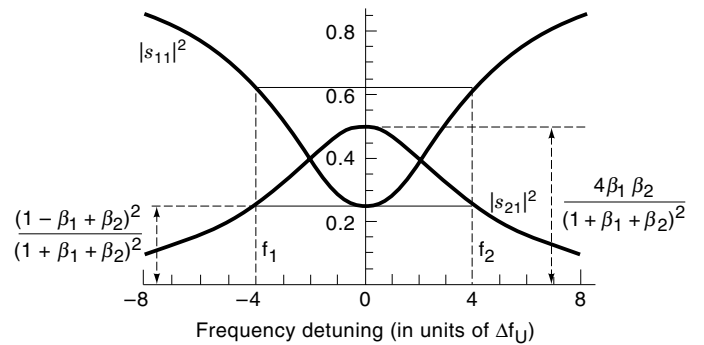


Figure 4. The power reflection and transmission coefficients of a two-port resonator with $\beta_1 = 1$ and $\beta_2 = 2$. The loaded Q of the resonator is given by $Q_L = f_0/(f_2 - f_1)$, where f_1 and f_2 are the points at which the power transmission coefficient is half the maximum power transmission, or the points where the power reflection coefficient is the average of the on and off-resonance power reflection coefficients.

of such a resonant curve would be its sharpness. We choose to measure the sharpness of the curve by the frequency difference between the two frequency values which transmit half the maximum power; that is, we need to find the frequency detuning Δf which satisfies the equation:

$$|s_{21}(\Delta f)|^2 = \frac{1}{2} |s_{21}(0)|^2 \quad (22)$$

The solution to Eq. 22 is

$$\Delta f = \pm \frac{1}{2} \frac{(1 + \beta_1 + \beta_2) f_0}{Q_U} = \pm \frac{1}{2} \frac{f_0}{Q_L} \quad (23)$$

Let us define a new parameter, Δf_L : the loaded half bandwidth which is equal to the positive solution of Eq. 23; that is,

$$\Delta f_L = \frac{1}{2} \frac{f_0}{Q_L} \quad (24)$$

Equation 23 gives us a method of measuring Q_L in transmission resonators. By measuring the resonant frequency, the maximum transmission power and the two frequency points at which the transmitted power is down by half, we can calculate the loaded Q .

To derive Q_u from this measurement, we will need to determine the coupling factors and use Eq. 21. By measuring the reflected power coefficient on resonance and the transmitted power coefficient on resonance, we have two equations with two unknowns (Eqs. 17 and 18). These can be solved to find the coupling coefficients:

$$\beta_1 = \frac{(\Gamma_0 \pm 1)^2}{1 - \Gamma_0^2 - T_0^2} \quad \text{and} \quad \beta_2 = \frac{T_0^2}{1 - \Gamma_0^2 - T_0^2} \quad (25)$$

where $\Gamma_0 = |s_{11}(0)|$ and $T_0 = |s_{21}(0)|$. The first of these two equations can also be applied to finding the coupling coefficient for a one port resonator by letting β_2 be equal to zero. The equation can then be rewritten as

$$\beta_1 = \frac{1 \pm \Gamma_0}{1 \mp \Gamma_0} \quad (26)$$

It is not possible to determine which of the two solutions in Eq. (25) or (26) is, in fact, the correct one if we are limited to a scalar power transmission method of determining the Q -factor. We need to measure some phase information (16), or vary the physical coupling in a known way (18,24), to remove this ambiguity. If one is not interested in knowing the coupling factors individually but just for the purpose of determining the unloaded Q factor from the loaded Q factor, then one can use (22)

$$1 + \beta_1 + \beta_2 = \frac{2}{s_{11}(0) + s_{22}(0)} \quad (27)$$

If it is possible to measure the complex components of s_{11} either directly, with a network analyzer or using a phase bridge, or indirectly, through measurements of the line VSWR, then there are some alternative methods to estimating the Q -factor of a resonator. For simplicity, we will perform the following calculations for a one port reflection resonator ($\beta_2 = 0$). However, it should be noted that the equations which govern reflection,

transmission, and reaction resonators are all essentially equivalent (25), and so, most Q measurement techniques can be applied to all three types of resonator (26,25).

We need to determine the relationship which exists between the real and imaginary components of s_{11} when at a particular frequency detuning. This relationship may then be used to measure the Q . Let us define an unloaded half bandwidth in the same way as we did for a loaded half bandwidth (Eq. 24):

$$\Delta f_U = \frac{1}{2} \frac{f_0}{Q_U} \quad (28)$$

Substituting Δf_L and Δf_U into the real and imaginary parts of s_{11} (Eqs. 12 and 13), we find the following relationships:

$$\text{Im}[s_{11}(\pm \Delta f_L)] = \mp \{1 + \text{Re}[s_{11}(\pm \Delta f_L)]\} \quad (29)$$

$$\text{Im}[s_{11}(\pm \Delta f_U)] = \mp \left\{ 1 - \sqrt{2 - (\text{Re}[s_{11}(\pm \Delta f_U)])^2} \right\} \quad (30)$$

If the frequency detuning is varied until one of the above equations are satisfied, the value of the detuning will be equal to the loaded or unloaded half bandwidth. By back substitution into Eqs. 28 and 24, the unloaded and loaded Q may be found.

GENERAL SOURCES OF ERRORS IN Q MEASUREMENT

We can divide measurement error into two categories: systematic and random. Systematic error yields a consistent result on repeated measurements of a constant resonance but with a value which differs from the true value. Such a measurement would be termed one with high resolution but relatively poor accuracy. Random errors (noise) will result in a scattering of results on repeated measurements of a constant resonance. If the measurement system is dominated by random noise, it is relatively easy to estimate its accuracy by just making multiple measurements of a constant resonance and quoting some statistical measure of the deviation. An improvement in the quality of the measurement can also usually be made by averaging multiple measurements. An estimation of the accuracy of a system which is dominated by systematic errors is far more difficult. Typically, one would try to use two or more completely independent measurement techniques to measure the same resonance and quote the differences between the results.

Systematic Q Measurement Errors

Any Q estimation technique which samples the scattering parameters at a number of different frequencies requires that we place conditions on the amplitude and frequency stability of the source to achieve a given accuracy. Any constant drift in the source frequency during the measurement will result in distortions to the resonance curve shape with a consequent distortion of the loaded Q value. Drifts in the amplitude of the source will give rise to distortions in the measured coupling factors which will give errors in the estimation of the unloaded Q . The relative error in the Q -factor under these conditions depends on the order in which the measurements are taken. However, for a worst case estimate, if we assume the measurements are taken in ascending or descending frequency order

and are equally spaced, then we can estimate the fractional error in estimation of Q_L as

$$\frac{\Delta Q_L}{Q_L} = \pm \frac{tkQ_L}{f_0} \quad (31)$$

where t is the time required to sample the response between $+\Delta f_L$ and $-\Delta f_L$, and k is the frequency drift rate of the source in Hz/s.

Drifts in the resonant frequency of the resonance during the measurement will also cause distortions to the measured Q values in much the same way as drifts in the source frequency. This is especially important when the temperature coefficient of the resonator is large, and the resonator Q -factor is high. For example, in the case of a sapphire dielectric resonator at 77 K, the fractional frequency temperature coefficient is around $10^{-5}/K$, while Q_L is around 5×10^7 (27). Temperature drifts of the order of 200 μK during the period of measurement will limit the accuracy of the Q -factor measurement to 10%.

Almost all frequency domain Q -measurements have traditionally relied on being able to treat the resonance as a single resonance which is well described by an equivalent LCR network. In the situation where there are losses or reactance in the coupling network, or there are interfering resonances in the resonator or in the external transmission lines, there will be distortions to the classic resonance shape with consequent errors in the measured Q factors. It is difficult to give a summary of the errors induced by ignoring these effects as there are so many possibilities. However, for example, the Q circles method (see below) will give a Q_U which is 15% in error for each 1 dB of return loss which is ignored in the lines leading to a critically coupled resonator.

In resonators which are remote from the experimental apparatus, or in systems with poorly matched loads and sources, resonances can appear in the connecting transmission lines which cause distortions to the resonance shape. The best solution to this problem is to suppress these spurious resonances by using microwave isolators to better match the load and source. To overcome large amounts of coupling reactance or interfering resonances in the resonator, itself, one must refer to some of the methods covered in the "Modern measurement techniques" section below.

If the resonator contains materials which have properties (either reactive or resistive) which are altered by the amplitude of the incident microwave fields, there is a strong possibility the resonator will exhibit a level of nonlinear behavior (28,29). For example, resonators which have been loaded with maser materials such as ruby can exhibit Q -factors which are negative under certain conditions (28). If high power microwave sources are being used, it is also possible for temperature changes to be induced in the resonator which cause resonant frequency changes, or for electrical breakdown to occur. Most superconducting cavities demonstrate a very strong Q dependence on power (3). Under all these circumstances, great care must be taken to ensure that correct results are obtained. For example, the Q -factor can be measured at a number of different incident powers to determine if a consistent result is being obtained.

Finally, any errors in the calibrations of the measuring instruments will lead to inaccuracies in the measurement. For example, Sucher (18) has calculated the error in measurement of the Q of a transmission resonator from an error in the

attenuation measurement (ΔA) of the half power transmission points to be

$$\frac{\Delta Q_L}{Q_L} = -0.23\Delta A \quad (32)$$

Thus, a 0.1 dB error in measurement leads to an inaccuracy of 2.3%.

Random Q Measurement Errors

Random noise in parts of the measurement apparatus will lead to random variations in the measured Q value of an unchanging resonance. Many of the mechanisms mentioned in the previous section are also applicable to this section if, rather than smooth drift of the variable, it exhibits noisy behavior. For instance, rapidly varying frequency noise (phase noise) of the microwave source will create an amplitude noise in the reflected or transmitted power. This will limit the experimenter's ability to identify a particular point on a resonance curve. For example, to measure a typical X-band dielectric resonator Q of 10^5 to an accuracy of 1% requires that the position of the measured frequency points is determined to better than 1 kHz. Such frequency stability is only achieved by modern quartz synthesized microwave signal sources.

All Q measurements will be ultimately limited by the intrinsic noise of the amplitude or phase sensor in the analyzing instruments. The fractional noise in the Q measurement is given approximately by the fractional noise in the power measurement. It is very rare that this ultimate limit is achieved in practice as the other noise sources (particularly the frequency noise in the source) conspire to increase the measurement noise level.

TRADITIONAL METHODS OF Q MEASUREMENT

The methods of Q measurement have been classified into the categories shown in Fig. 5. A major distinction has been made between spatial and temporal techniques because of the very

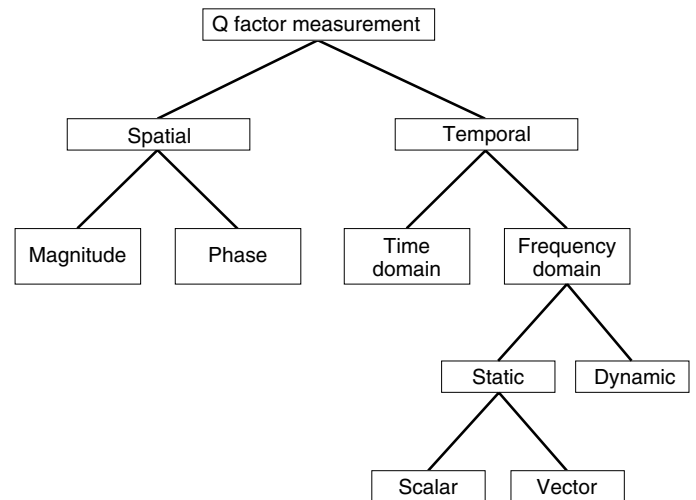


Figure 5. The Q factor measurement techniques described in this chapter have been divided into the classification scheme which is shown in this figure.

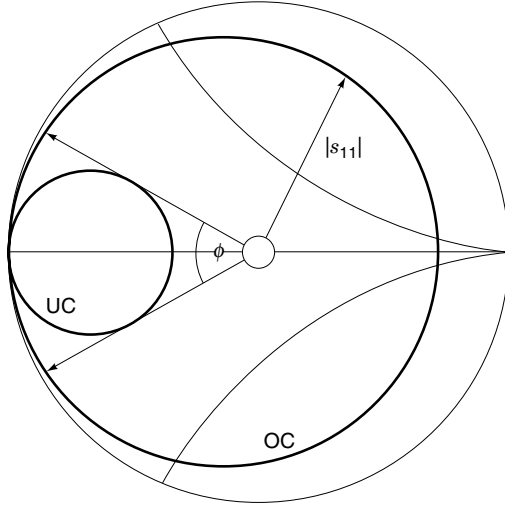


Figure 6. Smith Chart plots of the s_{11} responses of over (OC) and undercoupled (UC) resonances. The reflection phase of the OC resonance changes by 2π across the resonance, while the reflection phase of the UC resonance reaches a maximum value of $\phi/2$ before returning to zero and then reaching a minimum value of $-\phi/2$ before once again returning to zero. The relative changes in magnitude of the reflection coefficient are much larger in the case of the undercoupled resonator.

different technology required for the two types of measurement. Spatial techniques rely on using a slotted line device to measure the voltage standing wave ratio (VSWR) in the incident microwave line. The VSWR gives a measure of the magnitude of the reflection coefficient of the resonator, and the positions of the voltage minima in the VSWR pattern give a measure of the reflection phase.

Temporal measures of the Q include measurements made in both the frequency and time domains. Frequency domain measures are the most commonly used techniques today. These techniques rely on measuring either or both of the components of the scattering parameters as a function of frequency. This can be done in a step-wise measurement of the scattering parameter at a number of frequency points, or it can be done with a continuously swept source as long as the time of the sweep is much longer than the ring-down time of the resonance. The term *static* in Fig. 5 implies that measurements are made slowly compared to the ring-down time, and that the measured output signal is synchronous with the source frequency sweep. Dynamic measurements are also made with a sweep which is slow compared to the ring-down time of the resonance, but in these measurements, the output signal is at a harmonic of sweep rate, and so they are intrinsically AC measurements.

Spatial Methods: Phase and Amplitude

If a measurement were made of the voltage along the incident line of the microwave resonator in Fig. 2, we would find a standing wave formed by the interference of the incoming and outgoing waves. The ratio of the minimum to maximum voltage as measured along this transmission line is termed the VSWR and can be directly measured with a slotted line device. The locations of the voltage minima (or maxima) in this pattern are determined by the phase of the reflected signal. The amplitude

of the VSWR is given by

$$\text{VSWR} = \frac{1 + |s_{11}|}{1 - |s_{11}|} \quad (33)$$

To compare s_{11} for under and overcoupled resonators, a plot of each is given on a Smith Chart in Fig. 6 (as viewed from the detuned short position). One notes that $|s_{11}|$ (and hence the VSWR also) of an overcoupled resonator is large and relatively constant although the reflection phase changes by 2π across the resonance. The undercoupled resonance demonstrates a relatively small phase variation although the relative VSWR change is large. This knowledge arms the engineer for an accurate measurement of the Q using a slotted line device. If the resonator is overcoupled, then a measurement of the position of the voltage minima should be made, while if the resonator is undercoupled, then a measurement of the magnitude of the VSWR will yield the more accurate result.

Substitution of Eq. 9 into Eq. 33 yields the result that the VSWR on resonance is equal to $\beta^{\pm 1}$ for a one port resonator. A measurement of the VSWR alone is unable to resolve the sign of the exponent although an additional quick measurement can determine this (see below).

The VSWR at $\pm \Delta f_L$ and $\pm \Delta f_U$ for the undercoupled case can be calculated as a function of the VSWR on resonance (V_0) to be

$$\text{VSWR}(\pm \Delta f_L) = \frac{\left| \frac{j + V_0}{1 + V_0} \right| + 1}{\left| \frac{j + V_0}{1 + V_0} \right| - 1} \quad \text{and} \quad (34)$$

$$\text{VSWR}(\pm \Delta f_U) = \frac{\left| \frac{1 - (1 + j)V_0}{1 + (1 + j)V_0} \right| + 1}{\left| \frac{1 - (1 + j)V_0}{1 + (1 + j)V_0} \right| - 1}$$

Thus, if we measure V_0 , substitute it into the equation above, and then measure the frequency detuning at which the VSWR is equal to the calculated value, we can immediately determine Q_U and Q_L .

For overcoupled resonators, the Q is accurately determined by measuring the position of the voltage minima (or maxima) as the resonant frequency is changed (for a tunable resonator) or as the incident signal frequency is changed. A plot of the position of the voltage minima for two one-port resonators ($\beta_1 = 0.6$ and 1.6) is given in Fig. 7 as a function of resonant frequency. The relatively large change in the position of the minima for overcoupled resonators is clearly visible. The points on the overcoupled resonator curve which are most easily identified experimentally occur where the minima shift is $\pm \pi/2$ (this is half the maximum shift). Given that the phase shift on reflection from the resonator is given by

$$\phi = \text{ArcTan} \left[\frac{\text{Im}[s_{11}]}{\text{Re}[s_{11}]} \right] \quad (35)$$

we see that a phase shift of $\pm \pi/2$ occurs when the real part of s_{11} (Eq. 12) is equal to zero. Solve this to find the $\pi/2$ frequency

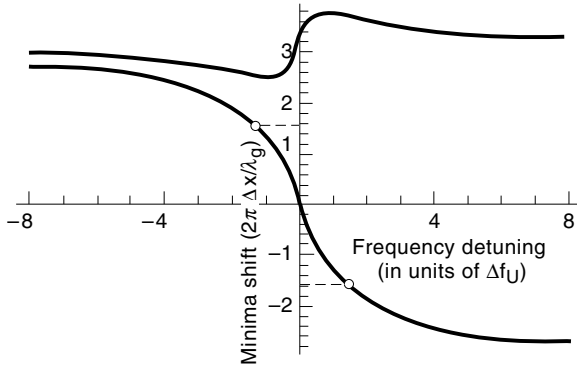


Figure 7. Calculated position shift (in radians) of the minima in the VSWR pattern as a function of resonant frequency for two resonances with $\beta_1 = 0.6$ (upper curve) and 1.6 (lower curve). The position shift is seen to be much larger for the overcoupled resonator which is in accordance with our prediction from the Smith Chart plot in Fig. 6.

detuning points:

$$\Delta f_{\pi/2} = \pm \frac{1}{2} \frac{f_0}{Q_U} \sqrt{\beta_1^2 - 1} \quad (36)$$

So a measurement of the frequency detuning which gives a $\pi/2$ shift combined with a measurement of the VSWR when the resonator is tuned to resonance will yield the unloaded Q of the resonator.

Figure 7 also suggests a simple method to determine whether a resonator is under or overcoupled. For undercoupled resonators, we note that the position of the minima is the same when the resonator is tuned exactly to resonance and when it is tuned far from resonance. This is not the case for overcoupled resonators. Thus if the slotted line probe is adjusted onto a voltage minimum at resonance, and then the resonant frequency is shifted so that it is far from resonance, only in the case of an undercoupled resonator will the probe still be at a voltage minimum.

Time Domain Techniques

Decrement Method. The decrement method is one of the older methods of measuring the Q factors of high Q resonators (17,18). Before the advent of quartz based microwave synthesizers, the frequency stability of the best Gunn oscillators and klystrons meant that measurements of Q factors higher than 10^{4-5} would be inaccurate. To overcome this difficulty, a method to excite the resonance and measure the time for the energy to dissipate was developed.

A microwave oscillator which can have its output power switched on or off by an external signal is driven by a pulse generator with a repetition rate much slower than the resonator ring-down time (Eq. 4). This pulsed generator also triggers an oscilloscope which is monitoring the energy leaking from the resonator from either the input or output port using a crystal detector. The power leaking from the resonator is proportional to the stored energy in the resonator. If the crystal detector has an accurate square law response, that is, its output voltage is proportional to the incident power, then the oscilloscope screen display is an accurate representation of Eq. 3, and the Q is directly determined by measuring the time taken for the signal to

drop by a factor of e . This time interval can then be substituted into Eq. 4 and the loaded Q determined.

Care must be taken to ensure that the microwave source frequency remains resonant and that the source output power is constant; otherwise, the oscilloscope trace moves vertically causing difficulty with an accurate measurement of the time interval. To avoid this difficulty, a method making use of a continuously swept source has been suggested (30). The source is swept across the resonance, and when it becomes resonant, an electronic circuit monitoring the transmitted power automatically switches the source off and triggers the oscilloscope. Using modern equipment, decrement techniques can easily measure loaded Q -factors near 10 GHz in the range upwards from 10^{4-5} (ring down times longer than a few microseconds).

Inaccuracies enter the decrement techniques if the bandwidth of the power detector and/or the oscilloscope is inadequate to correctly capture the transient response of the resonator. The measured Q factor will also be in error if the crystal detector does not give an output which is strictly proportional to input microwave power.

Unfortunately, all time domain methods give no information on the coupling and, hence, can only measure the loaded Q of the resonators. This difficulty may be overcome if it is possible to vary the coupling to the resonator. In this case, the coupling can be made progressively smaller until the measured Q does not change with any further reduction in coupling. At this point, the measured loaded Q will be equal to the unloaded Q to within the experimental resolution. The accuracy of the decrement techniques is directly proportional to the fractional error in the estimation of the ring-down time.

Fast Sweep Decrement Method

The fast sweep decrement method (31) is a modification of the decrement method with three advantages: the researcher can make use of more widely available CW sweep microwave generators rather than the pulsed generator required above, the required frequency stability of the microwave source is much less stringent, and no sophisticated electronics are required.

This measurement is suitable for a reflection resonator and requires a microwave detector to monitor the signal reflected from the resonator. The source is swept over a frequency range which is much wider than the bandwidth of the resonance with a sweep repetition rate (seconds per sweep cycle) which is much longer than the ring-down time of the resonator. The frequency sweep rate (Hz/s), however, is set such that the source sweeps across the resonance in a time which is much shorter than the ring-down time. An instant after the sweep has passed through the resonance, the detector mounted on the reflected port receives two signals: the reflected off-resonance signal which has a frequency corresponding to the frequency of the source at this instant, as well as the weak signal leaking from the resonator at the resonant frequency. The detector mixes these two signals giving rise to a characteristic frequency chirp (caused by the frequency difference having a time dependence) which has an exponential envelope following the decreasing strength of the leaking signal. From (31), we see that the envelope of the chirp has an exponential time constant which is twice as long as that seen in the normal decrement method; that is,

$$Q_L = \pi \tau_e f_0 \quad (37)$$

where τ_e is the time required for the signal envelope to decrease to $1/e$ of its amplitude. A measurement of τ_e yields Q_L .

A colleague (32) has developed an automatic Q measuring system which can provide relatively high resolution results based on this technique. By monitoring the dc and ac voltage output from the reflected power detector, one can calculate the incident power on the resonator as well as a value which is proportional to the amplitude of the chirp. The amplitude of the chirp is proportional to both the incident power and the Q of the resonance. By forming the ratio of the ac and dc voltage, one can calculate a number which is proportional to the loaded Q factor of the resonance. The constant of proportionality can be determined by an initial calibration using one of the other techniques for measuring Q . From then on, a new value for the loaded Q is generated after every sweep cycle. The technique gives measurements with a resolution of around 5% under typical conditions. The technique has proved very useful when the Q is changing quickly as a function of time.

Scalar Static Frequency Domain Techniques

Half-Power Points Technique. We have already outlined a method to determine the loaded Q from a transmission measurement where we determined the frequency point at which half the on-resonance power is transmitted and then directly substituted into Eq. 23 to calculate Q_L . However, an implicit assumption is made in Eq. 23 that one can identify the resonant frequency accurately to calculate the value of Δf . However, because the resonance shape is flat at the center of resonance, it is frequently difficult to identify this point accurately. Furthermore, the fractional error induced in the Q is extremely sensitive to an error in the calculation of Δf arising from a misidentification of the resonant frequency (Δf_0):

$$\frac{\Delta Q_L}{Q_L} = \frac{\Delta f_0}{f_0} Q_L \quad (38)$$

Since we can measure source frequency differences very accurately using an electronic frequency counter, a better method of measuring the Q relies on measuring the two frequencies f_1 and f_2 at which the transmission power is down by half from the on-resonance value (17,18):

$$f_1 - f_0 = \frac{1}{2} \frac{f_0}{Q_L} \quad \text{and} \quad f_2 - f_0 = \frac{1}{2} \frac{f_0}{Q_L} \quad (39)$$

If we form the difference of these two equations, we get

$$f_1 - f_2 = \frac{f_0}{Q_L} \quad (40)$$

where we can determine Q_L without the need for an accurate determination of f_0 (see Fig. 4). An additional measurement of the on resonance reflection and transmission coefficients (Eq. 25) gives us the coupling information needed to calculate Q_U .

Nonetheless, it is important to note that for an accurate measurement of the Q using this method, the requirements on the attenuation measurement are extremely stringent (Eq. 32). An alternative microwave bridge method has been proposed which can find the half transmission points very accurately (33).

Calibrated Attenuator Method. One of the difficulties of the Half-Power points technique is the necessity for an accurately

calibrated power detector. To avoid this difficulty, we can make use of a calibrated attenuator. This method relies on finding the maximum transmission point, while there is a 50% attenuator placed in the signal path between the microwave source and the detector. This attenuation is then removed, and the source detuned to the two frequency points at which the original output voltage is regained on the detector. Once again, the loaded Q is calculated using Eq. 40.

Reflected Power Technique. To determine the power reflected from a one port resonator at $\pm \Delta f_L$, we substitute into Eq. 17 to get

$$|s_{11}(\pm \Delta f_L)|^2 = \frac{1 + \beta_1^2}{(1 + \beta_1)^2} \quad (41)$$

Thus, to measure the loaded Q , we need to measure the reflection coefficient on resonance and calculate the coupling using Eq. 26. We then substitute this into Eq. 41 and calculate the power reflection coefficient at the loaded half-bandwidth frequency points. Finally, we measure the frequency detuning corresponding to this power reflection in the resonator, and from this frequency detuning, we can calculate Q_L and Q_U .

Straight-Line Methods. Another method for measuring the Q factor from the scalar reflection or transmission coefficients has been suggested where transformed frequency coordinates are used so that the resonance shape appears to be linear (34,35,18). This method lends itself to either a manual graphical solution or a simple least squares fit on a computer. We can rearrange Eq. 17 for a one port resonator [we can also do the same thing for the transmission coefficients (34)] to define a new variable:

$$\delta_U(\Delta f) = \left(\frac{Q_U \Delta f}{f_0} \right) = \pm \sqrt{\frac{(1 + \beta_1)^2 |s_{11}(\Delta f)|^2 - (1 - \beta_1)^2}{1 - |s_{11}(\Delta f)|^2}} \quad (42)$$

If we have calculated β_1 from a measurement of the on-resonance reflection coefficient (see Eq. 25), then for each measurement of s_{11} we can calculate a value for the variable δ_U (all values are known on the right hand side of Eq. 42). If we make a plot of these values of δ_U as a function of frequency detuning (Δf), we will find a straight line which has a slope equal to Q_U/f_0 . From the slope, we can clearly generate the unloaded Q (35).

When implemented in a computer based system, this technique has the potential for excellent high accuracy measurements of the order of 0.01% for Q -factors of the order of 10^4 (34). This accuracy arises because the measurement makes use of reflection data which is in close proximity to the center of resonance and because of the enormous amount of data which can be taken. By limiting the data to that near the center of resonance, there is a reduction in the effect of interference from other resonator and transmission lines modes. To achieve the accuracy reported, the researchers used nearly 1000 data points to characterize the resonance. This technique also has another advantage in that the degree of linearity of the plot is a measure of the level of interference from other modes. This immediately gives the experimenter an idea of the veracity of the measurement.

A similar method to the above mentioned one has been suggested in connection with transmission measurements at

93 GHz (36,37). One of the potential problems with the above method is finding the exact center of resonance. Any errors in this identification introduce a large error because Δf in Eq. 42 will be incorrect (see Eq. 38). If we define a new variable, m ,

$$m^2 \equiv \frac{|s_{21}(0)|^2}{|s_{21}(\Delta f)|^2} \quad (43)$$

then we can rewrite Eq. 18 as

$$m^2 = 1 + 4 \left(\frac{Q_L \Delta f}{f_0} \right)^2 \quad (44)$$

If we measure the left hand side of Eq. 43 at a known frequency detuning, we can immediately calculate Q_L from Eq. 44. However, any error in the measurement of the resonant frequency will introduce large errors. We would like to perform the calculation in a way which is insensitive to the exact value of the resonant frequency. If we make a measurement of m^2 at two frequency points on either side of the resonance, we can rewrite Eq. 44 as

$$\begin{aligned} \Delta f_1 = f_1 - f_0 &= \frac{-f_0}{2Q_L} \sqrt{m_1^2 - 1} \quad \text{and} \\ \Delta f_2 = f_2 - f_0 &= \frac{f_0}{2Q_L} \sqrt{m_2^2 - 1} \end{aligned} \quad (45)$$

If we take the difference of these two equations, we can derive a further equation to define Q which is only dependent on a frequency difference which, in principle, we can measure very accurately:

$$Q_L(f_1, f_2) = \frac{f_0}{2(f_1 - f_2)} (\sqrt{m_1^2 - 1} + \sqrt{m_2^2 - 1}) \quad (46)$$

If we make N measurements of m^2 on either side of the resonant frequency, then we have N^2 pairs of m values. From each pair, we can calculate a value for Q_L using Eq. 46. The variance of the ensemble of Q_L values is a measure of the closeness with which the data follows the theoretical resonance curve. If there is any systematic variation of Q_L as a function of frequency, it suggests the effects of interfering resonances. This method, is therefore, a very useful indicator of the isolation of the resonance (38).

Single Frequency Technique. As mentioned above, resonators and the resonator Q factor are commonly used to measure the electrical, optical, or magnetic properties of materials. These measurements are usually undertaken by measuring the properties of a resonator before and after making some change to that resonator such as, for example, inserting a small sample into the resonator, changing the illumination level on a sample already in the resonator, or changing the resonator temperature. In these cases, we are more interested in a measurement of the change in Q rather than an absolute measurement of the Q -factor. The single frequency measurement technique determines changes in the Q -factor essentially instantaneously and doesn't require measurements in the time or frequency domain (39,40).

If the change we have made to the resonator is such that the electromagnetic field distribution near the coupling ports is unchanged, then we can expect that the coupling ratios m

and n in Fig. 1 will remain unchanged. If the change causes a modification to the losses in the resonator (R_k in Eqs. 6 and 11), then both the unloaded Q and coupling will change, but the ratio of these two parameters will remain constant:

$$\frac{Q_u}{\beta_1} = \frac{R_s}{n^2} \sqrt{\frac{L_k}{C_k}} = C \quad (47)$$

where C is some constant. Thus, if we continuously monitor the reflected power at resonance, we can determine any changes in the coupling, and from Eq. 47, the fractional change in Q_u will be exactly equal to the fractional change in coupling.

Dual Resonator Techniques. A novel method for determining all resonator parameters from a simultaneous measurement of two resonators has been described by a number of authors (41,42,43). In these techniques, the power transmission and/or reflection coefficient of a resonator is plotted as a function of the transmission coefficient of a tunable reference cavity. The output plots are a function of the difference in resonant frequency of the two resonators and the ratio of their Q factors. Analytic expressions for the shapes of these curves have been derived in terms of the difference frequency and Q factor ratio, and these can be adjusted to fit experimentally derived measurements (41,43). The accuracy of this method is similar to those of the other techniques reported here ($\sim 2\%$), although it is necessary to know the Q factor of the reference cavity in advance of these measurements using one of the other methods. The big advantage of the technique is that it is unnecessary to use calibrated power detectors, although the detectors must have a matched response. The second major advantage arises because the data is represented on a parametric plot so any source frequency instability is suppressed. The quality of the measurement is determined by the stability of the reference cavity.

Integration Technique. If a frequency stable source and/or calibrated power measuring device is unavailable so that accurate measurements of the half power transmission points are impossible then an integration method can be used (44). If we integrate the transmitted power coefficient (Eq. 18) as a function of frequency, we obtain the following expression:

$$\int_{-\infty}^{\infty} |s_{21}|^2 d\Delta f = \frac{P_0 \pi f_0}{2Q_L} \quad (48)$$

where P_0 is the power transmitted at the resonant frequency. This equation can be implemented experimentally if we use a microwave sweep oscillator to sweep the frequency over many loaded bandwidths and use an analog integrator to integrate the output of a crystal detector on the transmission port of the resonator. An additional measurement of the output voltage of the detector when the generator is on-resonance determines the only other unknown parameter so that Q_L can be determined.

The integration method has also been proposed as an accurate method to determine the small changes in Q associated with determining losses in materials via resonator perturbation methods (45).

Numerical Methods. One of the earliest computer-based methods of measuring the Q factor was reported by Pandrangi et al. (46). A computer directed a voltage controlled oscillator

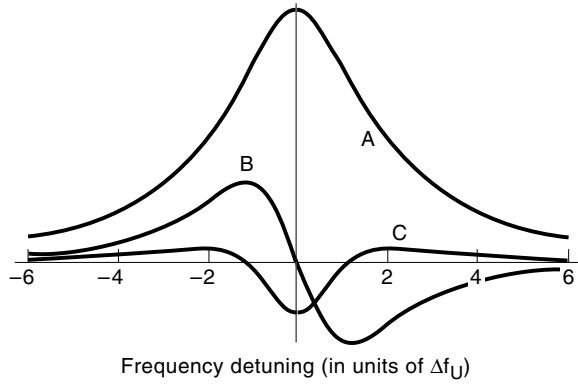


Figure 8. A plot of the frequency dependence of the power transmission coefficient (A), as well as its first (B) and second derivative (C), with respect to frequency (arbitrary vertical scale). Note that the curve possesses two inflection points (points where the second derivative is equal to zero).

(VCO) to step through the resonance curve of a transmission resonator at around 1 GHz. The incident and transmitted power were monitored by detectors which were, in turn, monitored by the computer. The power transmission coefficient was obtained by a division of these signals and a polynomial fitted to a smoothed version of this data. By differentiation of this polynomial, the inflection points of the resonance curve were found (see Fig. 8). To find how the inflection points are related to the Q , we need to find when the second frequency derivative of Eq. 18 is equal to zero; that is,

$$\frac{\partial^2}{\partial \Delta f^2} \{|s_{21}(\Delta f_{inf})|^2\} = 0 \quad (49)$$

which has as its solution

$$\Delta f_{inf} = \pm \frac{f_0}{2\sqrt{3}Q_L} \quad (50)$$

The experimental results were limited by frequency instability in the VCO to around 5%.

Frequency Lock Technique. An interesting method for accurately measuring the phase as a function of frequency is presented in the paper by Linzer et al. (47). In this method, a Klystron is frequency locked to a resonator. Typically, any frequency locking system is set up to detect the imaginary part of the reflection or transmission coefficient and then use this to stabilize the oscillator. If a phase offset is introduced into the feedback mechanism so that the locking system is now partially sensitive to the real part of the reflection coefficient as well, then it will lock at some distance from the center of resonance. The particular dependence of the oscillator frequency on this phase shift is determined by the loaded Q of the resonator, and hence by a measurement of the frequency offset in the system as a function of the feedback phase shift, the loaded Q can be measured with an accuracy of around 1–2%.

Vector Static Frequency Domain Methods

Q Circles Method. If one measures both the phase and magnitude (or real and imaginary components) of s_{11} using either

the VSWR method, a reflection bridge (48), or a network analyzer, then it is possible to plot s_{11} directly on a Smith Chart. If plots of the loci of Eqs. 29 and 30 for all Δf_L and Δf_U are also placed on the chart, then the intersections of the s_{11} curve and these two solution sets give the loaded and unloaded half bandwidths. This graphical solution method is outlined in detail by Ginzton (17).

Straight-Line Methods. If one can measure the real and imaginary components of s_{11} or s_{21} of a resonator separately, then one can plot the following ratios as a function of Δf for each of the different resonator types:

$$\frac{\text{Im}[s_{11}(\Delta f)]}{\text{Re}[s_{11}(\Delta f)]} [\text{reaction Eq. (49)}],$$

$$\frac{\text{Im}[s_{21}(\Delta f)]}{\text{Re}[s_{21}(\Delta f)]} (\text{transmission}),$$

$$\frac{\text{Im}[s_{11}(\Delta f)]}{\text{Re}[s_{11}(\Delta f)] + 1} (\text{reflection})$$

All of these ratios yield a straight line with a slope of $2Q_L/f_0$. Clearly by a measurement of the slope of these straight line plots, we can measure Q_L .

Dynamic Frequency Domain Techniques

There are two main dynamic frequency domain techniques, the harmonic or inflection point technique and the intermodulation method. Both of these techniques rely on the fact that at two frequency points on either side of the resonant frequency, there are points where the second derivative of the resonance curve is equal to zero (see Fig. 8 and Numerical Methods above). It is important to note that the sweep rates (Hz s^{-1}) for both of these techniques need to be much slower than $(f_0/Q_L)^2$ for the measurement to be accurate. In this sense, the measurements are still quasi-static, although we have chosen to term them dynamic because the measurement is made at a frequency which differs from the sweep frequency.

Harmonic (Inflection Points) Technique. If a frequency modulated signal is incident on the resonance curve shown in Fig. 8, and we monitor the power of the transmitted signal, it is clear that part of the frequency modulation (FM) will be converted to amplitude modulation (AM) by the resonator. If the frequency deviation is small, the amplitude of the transmitted AM at the fundamental FM frequency will be proportional to the slope of the curve measured at the carrier frequency. If the resonance has nonzero curvature; that is, if the second derivative is not zero, then the transmitted signal will also contain an AM component at the second harmonic of the FM. Thus, a measurement of the amplitude of the second harmonic AM as a function of frequency will yield a curve which is proportional to that shown as the second derivative curve (C) on Fig. 8. Since extremely sensitive synchronous detection techniques exist (such as lock in amplifiers), this method can very accurately find the points for which there is no second harmonic AM (17,18). Using Eq. 50 this value can be easily converted to the loaded Q of the resonator.

Intermodulation Method. A potential problem arises in the previous method if the FM signal is not completely free from

higher harmonic content. If the incident signal already contains the second harmonic component, then the measurement of the inflection points will be in error. To avoid this, we can make use of an intermodulation method. If we simultaneously frequency modulate an incident signal at two different frequencies, then it is clear that we will generate AM components in the transmitted signal at the two fundamental FM frequencies. However, if there is any non-linearity of the curve, then we will also generate AM components at the difference and sum frequency of the two components. If we set our system to detect one of these intermodulation components, such as the difference frequency, then we can be sure that this could only have been generated via the nonlinearity of the resonance curve (17,18). Once again, when identification of the inflection points is completed, we can calculate Q_L using Eq. 50.

MODERN MEASUREMENT TECHNIQUES

Because of the advent of powerful personal computers for data analysis, combined with the wide availability of microwave equipment with computer interfaces, many new techniques have been devised recently which allow the lifting of the restrictions which were previously placed on Q measurement. For example, accurate measurements can now be made in the presence of a transmission line or coupling reactance. An additional advantage of computer aided measurement is that large numbers of measurements, which would be tedious to take manually, are easily made. The data averaging which occurs from using such large numbers of data samples gives improved immunity to random noise in the Q measurement. This section will describe some of these latest methods and is divided into sections which highlight the major advantage of the technique.

Scalar High Resolution Measurements

Luiten et al. (50) have developed a new data averaging technique which doesn't require the use of expensive network analyzers and relies on measurement of the square of the magnitude of the reflection or transmission coefficient as a function of frequency. The acquired data is then fitted using a nonlinear algorithm to the expected shape of a completely general resonance (Eqs. 17 or 18). The fitting mechanism is adaptive, which gives the program a robustness and an ability to follow resonances if the resonance parameters change smoothly with time. This feature was used to great advantage to measure the temperature dependence of thermal, electrical, and magnetic properties of TiO_2 and Al_2O_3 crystals (10). The extreme frequency stability of the source in this case (fractional frequency stability $\sim 10^{-15}$) meant that the Q measurements had a resolution of 0.03% despite the loaded bandwidth being only 10 Hz.

Kajfez (24) has developed a technique which uses a scalar network analyzer to measure the values of the reflection coefficient as a function of frequency. This technique takes into account moderate coupling losses and, therefore, allows the off-resonance reflection coefficient to differ from 1. When compared to vector curve-fitting methods, this technique, like all scalar techniques, suffers from the disadvantage that it is impossible to tell whether a resonator is over or under-coupled. Nonetheless, its accuracy appears to be comparable to the vector methods ($\sim 0.2\%$).

Vector High Resolution Measurements

A simple modification to the Q -circles method has been suggested by Hearn et al. (51) to take into account losses in the coupling network. It can be shown that Eq. 30 is equivalent to the condition that the magnitude of the real and imaginary parts of the impedance of the resonator are equal. In the case where there is loss in the coupling mechanism or the transmission lines, a correction needs to be made; otherwise, the estimated Q will be less than the true value. Hearn et al. (51) have suggested a simple correction which reduces the errors of the traditional measurement by around 80 times in the case of high coupling loss and small coupling. Unfortunately, this correction cannot be implemented with a simple Smith Chart overlay, although it is easily implemented in a automated measurement system.

Sanchez et al. (49) report an absorption resonator Q measurement system based on a computer driven network analyzer. This system empirically removes the effects of any loss in the coupling network, or transmission lines, by rescaling the data by a factor obtained from the transmission coefficient far from resonance. A fit is performed to the Q -circle from which the coupling and Q is found. The Q_U measurement error is around 1%.

Kajfez (52,53) has developed a method which uses a network analyzer combined with a computer analysis to generate and fit Q -circles [Figs. 3(d) and 6]. Instead of graphically finding the solutions, he uses a least square fitting routine to an expression which includes the effects of frequency independent line and coupling reactance as well as any losses in the coupling network. This technique avoids the ambiguity that exists between over and under-coupled resonators in scalar techniques. The technique's accuracy is estimated as around 0.1% when utilizing approximately 1000 data points.

In the presence of significant line inductance, we find that instead of the ideal circle which should be seen in a plot of the real and imaginary components of the resonator impedance, we find a curve which is nearly circular but with a smaller diameter and which crosses over itself (54). By finding the two frequencies corresponding to the crossing over point as well as those corresponding to the maximum and minimum of the imaginary impedance component (the critical points), we can determine Q_U very accurately. These measurements are relatively easy using a modern network analyzer and will generate Q -factors which have an accuracy of about 0.5%.

Measuring Resonance Parameters in Presence of Other Resonances

Kajfez et al. (55) have developed an overlay for a network analyzer which takes into account the effects of inductance and loss in the coupling and transmission network as well as the effects of any distant resonant modes. These effects are modeled as a frequency dependent reactance in a series with the mode of interest. The effect of this frequency dependent series inductance is to cause the Q circle to develop distortions which make its apparent diameter smaller than it would otherwise be. If an accurate determination of the reference plane of the measurement is made, this technique can also measure the inductance of the coupling network.

Sauer et al. (56) have modeled the effects of reflections which exist in the lines between the measuring apparatus and

the resonator under test. These reflections create subsidiary line resonances which distort the resonance shape and create errors in the measurement of the Q -factor. The model can accurately fit experimental data taken in the presence of interfering line resonances yielding a Q -factor with an accuracy of $\sim 1\%$.

An extensive analysis and experimental procedure has been developed which allows for the accurate measurement of Q factors and coupling coefficients of two or more modes which are in extremely close proximity (19,57). The process relies on developing an expression which fully represents the circuit shown in Fig. 1(b) with enough free parameters to represent all visible resonance curves in the data as well as an additional frequency dependent term to represent any coupling reactance. The free parameters are then adjusted by a least squares fitting procedure to the experimental data. Even in the situation where a double loop is visible on the Smith Chart because two resonances are within a few bandwidths of each other, the accuracy of the Q measurement is around 2%.

Measuring Extremely Low Resonators

Resonators measurements are generally made through transmission lines or coupling networks which have poorly understood characteristics. The interaction of these external networks causes distortions to the resonance and, therefore, to the apparent Q -factor. This is an especially difficult problem when the Q factor of the resonator is low as the frequency rate of change of the resonator impedance becomes comparable to that of the external circuit parameters. Kajfez et al. (58) have developed an expression for determining Q_U when it is measured through an arbitrary two port. Using a similar approach, Drozd et al. (15) have developed a general Q -factor measuring method based on the full s_{11} parameters as determined by a computer driven network analyzer. The loaded Q is shown to be given by

$$Q_L = \frac{\omega_0}{2R_k} \left| \frac{dX_k}{d\omega} \right| \quad (51)$$

where X_k is the total reactance of the k th resonance and R_k is the total resistance of the k th resonance. The frequency derivative of the reactance is estimated from a number of measurements of s_{11} in the immediate neighborhood of center frequency of the resonance. Measuring near to the center of resonance gives immunity from the influences of stray inductance in the transmission line and coupling mechanism. This method proves to be far more accurate than a measurement of the half-power points or of the critical points, both of which are relatively distant from the resonant frequency. In addition, this method does not make the common assumption of Eq. 16 which is really only valid when the Q is greater than a few thousand. As a result, this method yields accurate results even for Q -factors of the order of 2.

Oscillator Phase Noise Determination

Fuchs (59) has made use of Leeson's (60) relation that in a simple microwave oscillator, the loaded Q -factor of the resonator directly determines the phase noise of the oscillator. By artificially introducing high levels of component phase noise, the oscillator noise floor becomes dominant over the phase noise measurement system's internal noise. From a plot of the phase

noise of the oscillator, the loaded Q can be directly determined to an accuracy of around 10%.

Theoretical Q Estimates

Theoretical estimation of Q factors has always been extremely difficult for all but the simplest metallic cavities with known surface resistance. Recently, a number of attempts have been made to model more complex hybrid dielectric resonators and estimate the Q factors on purely theoretical grounds (61,62,38). These estimates have value in that they could allow an independent verification of the veracity of Q measurement techniques. Unfortunately, as the accuracy of these calculations is still only around 10%, their predictive power is not yet useful in determining the accuracy of various measurement techniques.

ACKNOWLEDGEMENT

The author would like to thank Tony Mann for a critical reading of the manuscript. He would also like to thank all the personnel of the Sapphire Clock and Low Phase Noise Oscillator projects at UWA for providing an innovative and challenging environment in which to work.

BIBLIOGRAPHY

1. H. B. G. Casimir, On the theory of electromagnetic waves in resonant cavities, *Philips Res. Rep.*, **6**: 162–182, 1951.
2. D. Kajfez and A. Gundavajhala, Measurement of material properties with a tunable resonant cavity, *Elec. Lett.*, **29** (22): 1936–1937, 1993.
3. C. M. Lyneis, *Experimental Studies of Microwave Properties of 8.6GHz Superconducting Niobium, Niobium-Tantalum, and Tantalum Cavities*, Ph.D. Thesis, Department of Physics, Stanford University, California, 1974, unpublished.
4. J. Ceremuga, J. Krupka, and T. Kosciuk, Resonant measurements of surface resistance of high- T_c superconducting films: How good or bad are they?, *J. Superconductivity*, **8** (6): 681–689, 1995.
5. J. J. Hinds and W. H. Hartwig, Materials-properties analyzers using superconducting resonators, *J. Appl. Phys.*, **42** (1): 170–179, 1971.
6. M. N. Afsar, J. R. Birch, and R. N. Clark, The measurement of the properties of materials, *Proc. IEEE*, **74**: 183–199, 1986.
7. G. Birnbaum and J. Franeau, Measurement of the dielectric constant and loss of solids and liquids by a cavity perturbation technique, *J. Appl. Phys.*, **20**: 817–818, 1949.
8. G. Johri, R. Ngoui, and J. Roberts, A study of experimental and theoretical shifts in a resonant microwave cavity loaded with polar and non-polar molecules in the gas phase, *J. Microw. Power Electromagn. Energy*, **24** (4): 227–235, 1989.
9. D. Grissom and William H. Hartwig, Dielectric dissipation in NaCl and KCl below 4.2K, *J. Appl. Phys.*, **37** (13): 4784–4789, 1966.
10. A. N. Luiten, A. G. Mann, and D. G. Blair, Paramagnetic susceptibility and permittivity measurements at microwave frequencies in cryogenic sapphire resonators, *J. Phys. D: Appl. Phys.*, **29**: 2082–2090, 1996 and A. N. Luiten et al., Microwave properties of a Rutile resonator between 2–10K, *Meas. Sci. Tech.*, 1998, in press.
11. N. Klein et al., Dielectric properties of rutile and its use in high temperature superconducting resonators, *J. Appl. Phys.*, **78** (11): 6683–6686, 1995.
12. R. J. Deri, Dielectric measurements with helical resonators, *Rev. Sci. Instrum.*, **57** (1): 82–86, 1986.

13. S. R. Stein and J. P. Turneaure, Superconducting cavity stabilized oscillators with improved stability, *Proc. IEEE*, **63**: 1249–1250, 1975.
14. A. N. Luiten et al., Latest results of the UWA cryogenic sapphire oscillator, *Proc. 49th Freq. Control Symp. (FCS)*, San Francisco, 433–437, 1995.
15. J. M. Drozd and W. T. Joines, Determining Q Using S Parameter Data, *IEEE Trans. Microw. Theory Tech.*, **MTT-44**: 1996.
16. J. R. Ashley and F. M. Palka, Reflection Coefficient Measurement of Microwave Resonator Q Factors, *Microw. J.*, **14**: 35–39, 1971.
17. E. L. Ginzton, *Microw. Meas.*, New York: McGraw-Hill, 1957.
18. M. Sucher, Measurement of Q , in volume **II** *Handbook of Microwave Measurements*, M. Sucher and J. Fox, (eds.) New York: Wiley, 417–493, 1963.
19. W. P. Wheless and D. Kajfez, Experimental Characterization of Multimoded Microwave Resonators Using Automated Network Analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-35**: 1987.
20. K. Kurokawa, *An Introduction to the Theory of Microwave Circuits*, New York: Academic Press, 1969.
21. Z. Galani et al., Analysis and Design of a single resonator GaAs FET oscillator with Noise Degeneration, *IEEE Microw. Theory Tech.*, **MTT-32**: 1556–1565, 1984.
22. D. H. Han, Y. S. Kim, and M. Kwon, Two Port cavity Q measurement using scattering parameters, *Rev. Sci. Instrum.*, **67** (6): 2179–2181, 1996.
23. P. H. Smith, *Electronic Applications of the Smith Chart in Waveguide, Circuit and Component Analysis*, New York: McGraw-Hill, 1969.
24. D. Kajfez, Q -factor measurement with a scalar network analyzer, *IEE Proc.—Microw. Antennas Propag.*, **142** (5): 369–372, 1995.
25. M. C. Sanchez, E. Martin, and J. M. Zamarro, Unified and simplified treatment of techniques for characterizing transmission, reflection or absorption resonators, *IEE Proc.*, **137** (4): 209–212, 1990.
26. A. Khanna and Y. Garault, Determination of Loaded, Unloaded and External Quality Factors of a Dielectric Resonator Coupled to a Microstrip Line, *IEEE Trans. Microw. Theory Tech.*, **MTT-31**: 261–264, 1983.
27. M. E. Tobar et al., Low Noise Microwave Oscillators Based On High- Q Temperature Stabilized Resonators, *Proc. 48th IEEE Int. Freq. Control Symp.*, 433–440, 1994.
28. A. E. Siegman, *Microwave Solid State Masers*, New York: McGraw-Hill, 1964.
29. J. Petykiewicz, Chapter 10.14, *Wave Optics*, Dordrecht: Kluwer Academic, 1992.
30. A. Kocherzhin et al., Device for Measurement of High Q values by Attenuation, *Instrum. Exp. Tech.*, **31** (2): 494–496, 1988.
31. H. J. Schmitt and H. Zimmer, Fast Sweep Measurements of Relaxation Times in Superconducting Cavities, *IEEE Trans. Microw. Theory Tech.*, **MTT-14**: 206–207, 1966.
32. A. G. Mann, Physics Department, University of Western Australia, unpublished communication.
33. K. Watanabe and I. Takao, A bridge method for simultaneous measurements of coupling coefficient and loaded Q of a single ended cavity, *Rev. Sci. Instrum.*, **44** (11): 1625–1627, 1973.
34. K. D. McKinsty and C. E. Patton, Methods for determination of microwave cavity quality factors from equivalent circuit models, *Rev. Sci. Instrum.*, **60** (3): 439–443, 1989.
35. Z. Frait and C. E. Patton, Simple Analytic method for microwave cavity Q determination, *Rev. Sci. Instrum.*, **51** (8): 1092–1094, 1980.
36. R. K. Mongia and R. K. Arora, Accurate measurement of the Q factor of an open resonator in the W frequency band range, *Rev. Sci. Instrum.*, **63**: 3877–3880, 1992.
37. R. K. Arora, S. Aditya, and X. Xu, Computer Aided Measurement of Q -factor with Application to Quasi-Optical Resonators, *IEEE Trans. Instrum. Meas.*, **IM-40**: 863–866, 1991.
38. R. K. Mongia et al., Accurate Measurement of Q -factors of Isolated Dielectric Resonators, *IEEE Trans. Microw. Theory Tech.*, **MTT-42**: 1463–1467, 1994.
39. V. Subramanian and J. Sobhanadri, New Approach to measuring the Q -factor of a microwave cavity using the cavity perturbation technique, *Rev. Sci. Instrum.*, **65** (2): 453–455, 1994.
40. B. Tian and W. R. Tinga, Single-Frequency Relative Q Measurements Using Perturbation Theory, *IEEE Trans. Microw. Theory Tech.*, **MTT-41**: 1922–1927, 1993.
41. G. Magerl and K. R. Richter, A Novel Method for Cavity Parameter Measurement, *IEEE Trans. Instrum. Meas.*, **IM-25**: 145–151, 1976.
42. G. Magerl, Fast Measurement of Cavity Coupling Coefficients, *IEEE Trans. Microw. Theory Tech.*, **MTT-26**: 223–224, 1978.
43. I. Kneppo, Comparison Method of Measuring Q of Microwave Resonators, *IEEE Trans. Microw. Theory Tech.*, **MTT-25**: 423–426, 1977.
44. I. Kneppo, Integration Method of Measuring Q of the Microwave Resonators, *IEEE Trans. Microw. Theory Tech.*, **MTT-26**: 131, 1978.
45. J. C. Liu et al., Accurate Measurement of the Q factor for the Cavity Perturbation Method, *Microw. Opt. Lett.*, **13**: 87–90, 1996.
46. R. K. Pandrangi, S. S. Stuchly, and M. Barski, A Digital System for Measurement of Resonant Frequency and Q Factor, *IEEE Trans. Instrum. Meas.*, **IM-31**: 18–21, 1982.
47. M. Linzer and D. P. Stokesberry, A Frequency Lock Method for the Measurement of Q Factors of Reflection and Transmission resonators, *IEEE Trans. Instrum. Meas.*, **IM-22**: 61–77, 1973.
48. G. Ghosh et al., Method of measuring loaded Q -factor of single ended cavity resonators using reflection bridge, *Rev. Sci. Instrum.*, **49** (3): 378–379, 1978.
49. M. C. Sanchez, E. Martin, and J. M. Zamarro, New Vectorial automatic technique for characterization of resonators, *IEE Proc.*, **136** (2): 147–150, 1989.
50. A. N. Luiten, A. G. Mann, and D. G. Blair, High Resolution measurement of the temperature-dependence of the Q , coupling and resonant frequency of a microwave resonator, *Meas. Sci. Tech.*, **7**: 949–953, 1996.
51. C. P. Hearn, P. G. Bartley, and E. S. Bradshaw, A modified Q -circle Measurement procedure for Greater Accuracy, *Microw. J.*, **36** (10): 108–111, 1993.
52. D. Kajfez, Linear Fractional Curve Fitting for Measurement of High Q Factors, *IEEE Trans. Microw. Theory Tech.*, **MTT-42**: 1149–1153, 1994.
53. D. Kajfez, *Q Factor*, Oxford, MS: Vector Forum, 1994.
54. E. Y. Sun and S. H. Chao, Unloaded Q measurement, *IEEE Trans. Microw. Theory Tech.*, **MTT-43**: 1983–1986, 1995.
55. D. Kajfez and E. J. Hwan, Q -Factor measurement with network analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-32**: 666–670, 1984.
56. B. Sauer et al., Precise calibration of a microwave cavity with a non-ideal waveguide system, *Rev. Sci. Instrum.*, **62** (1): 189–197, 1991.
57. W. P. Wheless, A resonator de-embedding procedure, *IEEE Trans. Microw. Theory Tech.*, **MTT-38**: 864–869, 1989.
58. D. Kajfez and W. P. Wheless, Invariant definitions of the unloaded Q factor, *IEEE Trans. Microw. Theory Tech.*, **MTT-34**: 840–841, 1986.
59. M. Fuchs, Simple and accurate measurement of loaded Q factor of microwave oscillators using thermal noise injection, *Proc. 20th Eur. Microw. Conf.*, **1**: 512–516, 1990.

60. D. B. Leeson, A simple model of feedback oscillator noise spectrum, *Proc. IEEE*, **54**: 329–330, 1966.
61. J. Krupka, Computations of frequencies and intrinsic Q factors of TE_{0nm} modes of dielectric resonators, *IEEE Trans. Microw. Theory Tech.*, **MTT-33**: 1985.
62. C. Wang, B. Q. Gao, and C. P. Deng, Accurate study of Q factor of resonator by a finite difference time-domain method, *IEEE Trans. Microw. Theory Tech.*, **MTT-43**: 1524–1529, 1995.

ANDRE LUITEN
University of Western Australia

Q-METERS

The concept of resonance is one that is well known to all students of science. There is hardly a field of science that does not have a prominent example of the resonant phenomenon. Resonance is used in popular examples for solving differential equations in mathematics texts, linear systems, and electrical circuits. Resonance is found from the most microscopic atomic structure to the orbits of celestial bodies and orbiting satellites.

Resonance requires at least two energy storage elements that are coupled so that the energy stored in one element may be transferred to the other. An example of this is the pendulum. The pendulum swings from left to right at a constant frequency. When the pendulum has reached the highest point of its swing, the velocity of the pendulum bob reaches zero and the velocity reverses. Thus, the pendulum begins to swing in the reverse direction and will continue to swing until it reaches the opposite side and the process is repeated.

What causes this periodic activity to take place? A pendulum hanging straight down will not move (as Newton showed). A mass at rest will remain at rest unless acted upon by an external force. The pendulum bob hanging at its lowest position is in the lowest energy state, which is a stable condition for a mechanical system. If the pendulum were pulled to one side by a small amount, the pendulum would be raised slightly in the vertical direction. This slight raising of the pendulum mass would constitute an additional energy in the system. Certainly, pulling the pendulum mass to one side requires a force to be exerted through a distance, which is energy.

At the point when the pendulum is pulled to one side and before the bob is released, the pendulum system has a certain amount of potential energy. The actual amount of potential energy is not important at this time.

To keep the pendulum mass at the displaced position, it is necessary to retain a hold of the mass. Once the mass is released, the natural force of gravity will tend to pull the pendulum bob to its original position, hanging straight down. This is the lowest energy position, where the potential energy is zero. Mechanical systems attempt to assume the lowest energy level if there are no restraints, such as holding the pendulum mass.

The natural forces of nature—gravity in this example—tug at the pendulum bob; and when the bob has reached the lowest point of its swing, because forces have been acting on a mass, there has been acceleration of the bob and it now has a velocity. The pendulum will not stop at the lowest kinetic energy state because there is nothing to stop the swinging mass. The system is now an isolated system.

This situation can be further understood if the energy content of the system is investigated. One of the most powerful tools for analyzing physical systems is the concept of conservation of energy. All energy is conserved: mechanical, electrical, heat, and so on. Energy is a scalar so the complexity of vector calculations may be avoided. In the process of conservation of energy, the form of energy may be changed. Thus electric energy may become heat energy, and mechanical energy may be converted to electric energy. Even mass can be converted into energy and vice versa.

In the pendulum example, the energy conversion is simple; it is the conversion from potential energy to kinetic energy.

When the pendulum mass was pulled to its maximum deflection, the added energy was potential. Since the bob was not moving and thus had no velocity, the entire energy of the system was potential. When the mass was released and accelerated to the lowest point of its travel, the energy was all kinetic because this was the original point where the bob hung before energy was added to the system by pulling the bob to one side.

Therefore, as the pendulum swings, the energy is transferred from potential to kinetic but the total energy remains constant.

The pendulum would swing for all eternity if the energy did, in fact, remain constant. However, a swinging pendulum eventually comes to rest. The pendulum must give up some of its energy in moving the air to make way for the swinging bob. This energy expended in moving the air is not replaced and must come from the total energy of the pendulum. The bearings that the pendulum swings about will have some friction; and this, too, will subtract from the total energy content of the pendulum.

A pendulum can be made to have the least amount of friction so that it will swing for a long time. The pendulum clock is enclosed in a cabinet to reduce the effects of air friction due to moving air, and very low friction knife edge supports are used. The typical “grandfather” clock can run more than a week with just the energy from a weight.

In addition to reducing friction, a pendulum can be made to swing for a long time by increasing the mass of the bob. Foucault pendulums, found in a number of science museums, will swing for days with no external energy sources due to their bob mass of thousands of kilograms.

The pendulum swings with a constant frequency regardless of amplitude, provided that it is not swinging at a very great amplitude. Even though the amplitude of the pendulum’s oscillations is constantly decreasing, the period of oscillation remains the same until the very last oscillation.

Electric resonant circuits operate in a fashion similar to that of the pendulum. Two energy storage elements are required to create an electric resonant circuit. This would appear, at first, to be quite different from the example of the mechanical pendulum which had only one storage element, the bob at the end of the pendulum. Although there is, physically, only one bob, energy is stored two ways, kinetic and potential. Therefore, there are two energy storage “means,” to borrow a term from the patent attorneys. In the electrical circuit, two different physical devices are required to store energy two different ways.

A capacitor stores energy in an electric field that resides in the dielectric between the conducting plates of a capacitor. This energy storage means could be likened to potential energy

because the charges that create the electric field can be static, with no motion—that is, no current. The energy stored in a capacitor is

$$E(t) = \frac{1}{2}CV^2(t) \quad (1)$$

where $E(t)$ is the energy contained in the capacitor as a function of time in joules, $V(t)$ is the voltage across the capacitor, and C is a parameter called capacitance, with the units of farads which could actually be defined by this very equation. C defines how much energy is stored per volt of electric potential difference.

Opposite the capacitor, a magnetic field is the storage means for energy in an inductor. A magnetic field is created by current, or moving charges. Thus, the energy storage in an inductor could be viewed as kinetic energy. The energy stored in an inductor is

$$E(t) = \frac{1}{2}LI^2(t) \quad (2)$$

where $E(t)$ is energy, as before, $I(t)$ is the current within the inductor, and L is a parameter called inductance, with the units of henrys, which could, as in the case of the capacitor, be defined by this equation.

In the case of the pendulum, since the bob stored both the potential and kinetic energies there was no need to connect the two energy means. With two separate energy storage devices in the electric resonant circuit it is necessary to connect the two together. This can be done in two ways: a parallel connection and a series connection. The circuit shown in Fig. 1 is a resonant circuit that could be either parallel or series, because there is only one way to connect the inductor and the capacitor. Later in the discussion of electric circuit resonance the differences between series and parallel will become evident.

The circuit shown in the figure will oscillate if there is any energy in the circuit, and the oscillations will continue infinitely because the circuit shows no method of converting the internal energy of the circuit to any other form. The frequency of the oscillations will be

$$F_0 = \frac{1}{2\pi(LC)^{1/2}} \quad (3)$$

where F_0 is the resonant frequency in Hz or 1/s.

No resonant circuit can oscillate for eternity and the circuit in Fig. 1 must include a resistance, which is the circuit element that represents the conversion of energy to another form or represents a loss of energy from the circuit.

Now, it is clear that series and parallel resonant circuits are different. Figure 2 shows a series and a parallel resonant circuit with an energy loss element, a resistor.

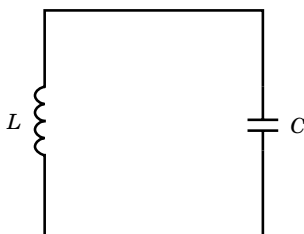


Figure 1. A simple resonant circuit with no energy dissipation.

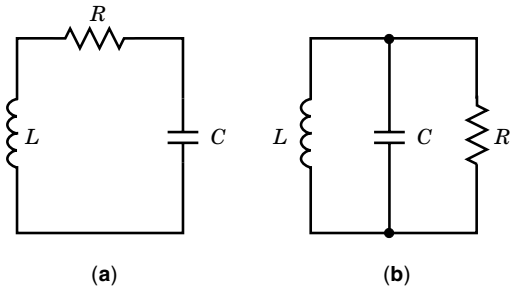


Figure 2. Series (a) and parallel (b) resonant circuits with energy dissipation.

As an aid to the understanding of resonance, a parameter is desired that would quantify how quickly or how slowly energy is lost from a resonant circuit. If the fraction of the energy that is lost for each cycle of oscillation were the parameter used, it would provide a good indicator of how quickly the energy was lost from the circuit. As an example, if one-tenth of the energy were lost each cycle, the energy would decay exponentially and could be easily predicted. Based on this premise, a factor called Q is defined as

$$Q = 2\pi \left(\frac{E_s}{E_d} \right) \quad (4)$$

where E_s is the total energy stored in a system and E_d is the energy dissipated per cycle of oscillation. Notice that Q has no units because it is a ratio of energies. Q factor can be used to describe resonance that involves electric circuits, mechanical devices, and combinations of electric and mechanical devices such as control systems.

In the discussion of the pendulum, the actual mechanisms responsible for energy loss were discussed and included air resistance and friction. It would be useful to investigate the loss mechanisms in the electric resonant circuit. First, consider a resonant circuit that receives no external energy and provides no energy to any external circuits. The resonant circuit has a finite initial energy and simply oscillates until the internal energy is totally exhausted. Of course an isolated resonant circuit has no use and must be eventually connected to other circuits to be of use, and this will be discussed further. A resonant circuit that has no external energy source or sink loses energy due to its inductor or capacitor.

The Q of a circuit that has no external energy paths, either source or sink, is called the unloaded Q . When other circuits are coupled to the resonant circuit, the Q for this situation is called the loaded Q . In the example of the clock pendulum, some energy is required to be removed from the swinging pendulum to drive the escapement of the clock, and energy must be replaced by the weights to keep the pendulum in motion. Both the extraction and addition of energy to the pendulum will reduce the Q .

In the case of the capacitor, there are several loss mechanisms. First is dielectric heating. Second are resistive losses due to the finite resistance of the conducting plates of the capacitor. Finally, there is radiation, which occurs when the electric field is not totally contained in the dielectric of the capacitor and can radiate. Generally, capacitors are not the major factor in resonant circuit losses. Even at the highest

frequencies, capacitors can be manufactured to have very low losses.

The inductor loses energy in three major ways. First is the resistive loss due to the wire used to make the inductor. At higher frequencies, this can become significant because of the increase of resistance due to the skin effect. Second are the losses due to energy dissipated in the core material of the inductor. When air or some other nonmagnetic material is used in the core, losses are virtually zero. However, when ferromagnetic material such as ferrite is used, the losses can be quite significant. The final major loss mechanism in an inductor is radiation. Unlike the capacitor, an inductor can have considerable radiation.

The solenoid inductor is the worst form factor when it comes to radiation. Solenoids are actually used as H -field antennas at low frequencies. By bending the solenoid around on itself to create a toroidal shape the radiation from the end of the solenoid can be significantly reduced. This type of inductor is expensive to make and difficult to work with and is avoided for many designs.

The Q of an inductor is a function of the frequency of measurement. The loss mechanisms for inductor loss are functions of frequency. Skin effect causes increased resistance with increasing frequency, and hence increased resistance losses. Radiation also increases with increasing frequency because the physical dimensions of the inductor become a more significant part of a wavelength. When ferromagnetic material is used in the core of an inductor, the frequency dependence of the inductor's Q can become quite complex. Shielded enclosures are often used to reduce the amount of radiation from an inductor, and this adds even more complexity to the Q factor calculation.

For most resonant circuits, the losses of the inductor are the major losses of the circuit, and the quality of the inductor will set the quality of the resonant circuit.

Because the inductor is the major loss element in a resonant circuit, it is common to attach a Q value to the inductor, alone. The very same definition of Q is applied except the resonant circuit is the inductor for which the Q value will be associated and a perfect capacitor. Therefore, the losses in the resonant circuit are those of only the inductor.

If the loss of a resonant circuit is considered to be that of the inductor alone, the resistance shown in the series resonant circuit of Fig. 2(a) or 2(b) is associated with the inductor. To find a relationship between Q , the equivalent resistance, and the inductance of an inductor, the current of the series circuit is given as

$$i(t) = I(t) \cos(2\pi F_0 t) \quad (5)$$

where $i(t)$ is the circuit current and $I(t)$ is the peak amplitude that is decreasing in some unspecified manner. A cosine function is assumed for the circuit current. A sine function could be employed as well as the inclusion of a phase angle. The solution to the circuit current of Fig. 1(a) is a sinusoid, which is either a sine or cosine function. The initial conditions will determine which of the two functions or a possible phase angle will solve the equation. For the purpose of understanding Q , the cosine function with a phase angle of zero will suffice.

Assume that a relatively high Q is involved such that $I(t)$ does not change much during the time of one cycle. The

root-mean-square (rms) value of $i(t)$ is $(0.707) I(t)$, where $I(t)$, is treated as a constant. The power during this cycle is $(I_{\text{rms}})^2 R$. The energy dissipated during the cycle is simply the time duration of one cycle times the power dissipated during the cycle, which is then

$$E_d = \frac{(I_{\text{rms}})^2 R}{F_0} = \frac{\frac{1}{2}(I^2(t)R)}{F_0} \quad (6)$$

Returning to the definition of Q and using the equations for E_d and E_s , the following relationship is obtained:

$$Q = \frac{2\pi E_s}{E_d} = \frac{2(\frac{1}{2}LI^2(t))}{\frac{1}{2}I^2(t)R/F_0} = \frac{2\pi F_0 L}{R} = \frac{X_L}{R} \quad (7)$$

where X_L is the inductive reactance.

The entire concept of Q is applied to circuits and systems where only a small fraction of the energy is dissipated per cycle or typically for Q values greater than about 10. Circuits with Q values significantly less than 10 do not oscillate but have a more exponential decay. Q values between 1 and 10 represent highly damped circuits where the normal assumptions concerning the calculation and measurement of Q may not produce accurate predictions of circuit behavior.

A similar analysis may be applied to the parallel resonant circuit as shown in Fig. 2(b). In this example, the voltage is the common parameter to the two circuit elements and may be represented by

$$v(t) = V(t) \cos(2\pi F_0 t) \quad (8)$$

where $v(t)$ is the circuit voltage and $V(t)$ is the peak amplitude. The same reasoning as was applied to the series circuit is used here to justify the form of the equation for $v(t)$. Likewise the same analysis will be applied to determine the relationship between Q and the circuit elements:

$$E_d = \frac{(V_{\text{rms}}(t))^2}{RF_0} = \frac{\frac{1}{2}V^2(t)}{RF_0} \quad (9)$$

For the stored energy to calculate Q , the energy stored in the capacitor will be used because this is a voltage equation and the energy stored in the capacitor is a function of voltage. Therefore,

$$E_s(t) = \frac{1}{2}CV(t)^2 \quad (10)$$

$$Q = \frac{2\pi E_s}{E_d} = \frac{2\pi(\frac{1}{2}CV^2(t))}{\frac{1}{2}V^2(t)/RF_0} = 2\pi F_0 CR = \frac{R}{X_C} = \frac{R}{X_L} \quad (11)$$

where X_C is the capacitive reactance.

MEASURING Q

The next consideration is measuring Q . From the discussion thus far, it would appear that if a resonant circuit were set into oscillation along with the number of cycles required to reach half-amplitude or other criterion, then Q could be determined in this fashion. This technique would work and produce the

value of Q . However, there are other methods of measuring Q that are more practical. In order to understand the measurement of Q , consider the impedance of a series resonant circuit at near resonance:

$$Z(f) = R + j(X_L - X_C) = R + j\left(2\pi fL - \frac{1}{2\pi fC}\right) \quad (12)$$

Using the substitutions

$$L = \frac{1}{4\pi^2 F_0^2 C} \quad \text{and} \quad C = \frac{1}{4\pi^2 F_0^2 L} \quad (13)$$

the impedance may be rewritten as

$$\begin{aligned} Z(f) &= R + j\left(\frac{2\pi f F_0 L}{F_0} - \frac{4\pi^2 F_0^2 L}{2\pi f}\right) \\ &= R + j(2\pi F_0 L)\left(\frac{f}{F_0} - \frac{F_0}{f}\right) \\ &= R\left[1 + j\left(\frac{2\pi F_0 L}{R}\right)\left(\frac{f}{F_0} - \frac{F_0}{f}\right)\right] \\ &= R\left[1 + jQ\left(\frac{f}{F_0} - \frac{F_0}{f}\right)\right] \end{aligned} \quad (14)$$

From this equation, the real part of the resonant circuit is a constant at R while the imaginary part is zero at the resonant frequency. Also, notice that if the frequency at which the impedance is measured is greater than the resonant frequency, then the sign of the imaginary part of the impedance is positive, indicating that the impedance becomes more inductive. Below the resonant frequency sign is negative, indicating that the impedance is more capacitive.

For the case of the parallel resonant circuit a similar equation is derived for the admittance of the circuit:

$$Y(f) = \frac{1}{R[1 + jQ(f/F_0 - F_0/f)]} \quad (15)$$

In this situation, when the frequency is greater than the resonant frequency the admittance becomes positive, indicating an inductive circuit. When the frequency is less than the resonant frequency, the admittance is capacitive.

Q was shown to be a parameter that described how the oscillations of a resonant circuit decayed in time. This was the time-domain picture of Q applied to a resonant circuit. When analyzed in the frequency domain, a circuit with a high Q has a narrow bandwidth. To determine the relationship between Q and the resonant circuit bandwidth the impedance of a series resonant circuit is considered:

$$Z(f) = R\left[1 + jQ\left(\frac{f}{F_0} - \frac{F_0}{f}\right)\right] \quad (16)$$

The series resonant circuit has a minimum impedance of, simply, R at the resonant frequency. The half-power level of a system is the usual way of specifying bandwidth of a circuit. In the case of a series circuit driven from a current source, half-power in the resonant circuit would occur when the magnitude of the impedance is $1.414R$.

Therefore,

$$|Z(F_{3dB})| = 1.414R \quad (17)$$

$$Z^2(F_{3dB}) = 2R^2 = R^2 \left[1 + Q^2 \left(\frac{F_{3dB}}{F_0} - \frac{F_0}{F_{3dB}}\right)^2\right] \quad (18)$$

Rearranging, we obtain

$$F_{3dB}^2 - \frac{F_0 F_{3dB}}{Q} - F_0^2 = 0 \quad (19)$$

Solving this equation for F_{3dB} , we obtain

$$F_{3dB} = \left(\frac{F_0}{2Q}\right)(1 \pm (1 + 4Q^2)^{1/2}) \quad (20)$$

For $Q \gg 1$ the equation above reduces to

$$F_{3dB} = \left(\frac{F_0}{2Q}\right)(1 \pm 2Q) \quad (21)$$

Again, by rearranging we obtain

$$F_{3dB} = F_{0-}^+ \left(\frac{F_0}{2Q}\right) \quad (22)$$

There are two 3 dB frequencies, one below the resonant frequency and a second above the resonant frequency. The difference between these two frequencies is the 3 dB bandwidth, BW, and is

$$BW = \frac{F_0}{Q} \quad (23)$$

Because a resonant circuit is highly frequency-selective, it is the major component of filters. The bandwidth of a resonant circuit is dependent on the Q of the circuit. When a resonant circuit is used in a filter the Q of the circuit is set by adding resistors to the basic resonant circuit to decrease the loaded Q and achieve the necessary bandwidth required for the filter design.

The unloaded Q will determine the bandwidth of the basic resonant circuit. To understand the relationship between Q and bandwidth, consider the driven resonant circuit in Fig. 3. In this example a 0 Ω generator is shown providing the input driving function to the circuit. Thus, the loaded Q and the unloaded Q are the same. If a real generator were used, some finite amount of impedance must be added to the voltage generator and the loaded Q will be less than the unloaded Q . In this circuit the maximum circuit current will be found at the resonant frequency and is V_{in}/R .

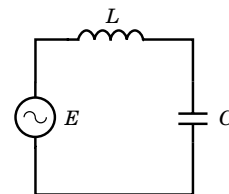


Figure 3. A driven resonant circuit.

Generally, the unloaded Q of a resonant circuit is significantly higher than the desired loaded Q . This is done so that the loaded Q can be controlled. Unloaded Q s are of a function of parameters that are not easily controlled.

The discussion thus far is of Q as associated with a resonant circuit. It was also shown that the Q of a resonant circuit is mainly set by the inductor because the inductor tends to have significantly more loss than capacitors. The Q of an inductor is the Q of a resonant circuit that is constructed with the inductor with a perfect capacitor. The Q of a capacitor is that of a resonant circuit with a perfect inductor. A resonator such as a quartz crystal or ceramic resonator is often described using a Q factor; this would be the Q of the resonator, alone.

For the case of Q -factor for inductors and capacitors, the Q of these components is a function of frequency. Generally, the higher the frequency the lower the Q because high frequencies exacerbate the loss mechanisms such as skin effect and radiation. Because of this frequency dependence, Q -factor must be measured and specified at a frequency near that which the inductor will operate.

Q -factor is also used for devices that are, in themselves, a resonant circuit. One example of this is the quartz crystal. This resonator has very high values of Q extending from tens of thousands to over 1 million. Ceramic resonators are similar to quartz crystals except the Q -factors are considerably lower and the ceramic devices do not have the superb frequency stability found in quartz crystals.

There are other devices such as ultrasonic transducers that are essentially resonators that are characterized by a Q -factor.

THE Q-METER

A method of measuring Q , primarily of inductors, is the appropriately named Q -meter. This device was invented in the mid-1930s by the Boonton Radio Company in Boonton, NJ. These instruments were manufactured for more than 40 years, mostly by Boonton Radio. There were a few other manufacturers such as The Marconi Corporation in England, and there was also a do-it-yourself kit by the venerable Heathkit Company that was a copy of the classic Boonton meter. When Boonton Radio was purchased by Hewlett-Packard in the 1960s a new, solid-state Q -meter was designed and was available for another two decades or so.

The Q -meter, as shown in Fig. 4, consists of an oscillator with a very low output impedance which was used to excite a resonant circuit which was made up of the unknown inductance and an internal calibrated capacitor. This instrument not only

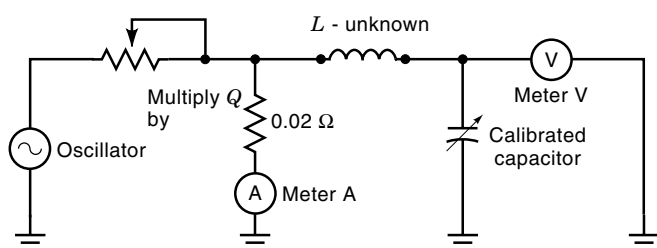


Figure 4. The basic schematic of a Q -meter.

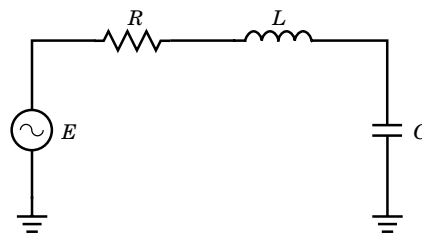


Figure 5. A driven resonant circuit with finite Q .

provides the ability to measure Q at a variety of frequencies, but could be used to measure inductance as well.

An oscillator feeds the resonant circuit which is in series resonance. A high-impedance alternating-current (ac) voltmeter is connected to the inductor and the actual inductor voltage is measured. The calibrated capacitor resonates the unknown inductor, and resonance is indicated by the maximum voltage indication of the voltmeter.

Historically, the output resistance of the generator was $0.02\ \Omega$ and the voltage was $0.02\ \text{V}$.

For a resonant circuit with an applied voltage from a $0\ \Omega$ generator, the voltage across the inductor or capacitor is Q times the applied voltage. The voltage-increasing phenomenon gives rise to the English name for a Q -meter, namely, "circuit magnification meter." To derive this relationship, consider the circuit shown in Fig. 5. In this example the generator has exactly $0\ \Omega$ internal resistance, which is, of course, only theoretically possible. The actual $0.02\ \Omega$ source resistance of the Q -meter causes only small errors for medium- or high- Q inductors.

At resonance the rms circuit current is

$$I = \frac{E_{\text{gen}}}{R} \quad (24)$$

where E_{gen} is the rms generator voltage.

The inductor rms voltage is

$$E_{\text{ind}} = IX_L = \frac{E_{\text{gen}}X_L}{R} = E_{\text{gen}}Q \quad (25)$$

The inductor voltage is Q times the generator, which shows how the input voltage is "magnified" by Q .

The voltage source in the Q -meter is not $0\ \Omega$ but $0.02\ \Omega$. The actual source voltage is determined by measuring the current in the $0.02\ \Omega$ resistor using a thermocouple ammeter. From the previous equation, the inductor voltage, which is the actual indication of Q , is directly proportional to the generator voltage. The accuracy of the meter depends on the precision of the generator voltage. The Q -meter provided a control to accurately set the generator voltage using a meter. Also, the generator voltage could be reduced to increase the scale factor of the meter. The meter was labeled "multiply Q by" with a prominent 1 on the scale along with other convenient values such as 2, 5, and 10.

The capacitor must not contribute significantly to the energy loss in the resonant circuit formed by the unknown inductor and the Q -meter's capacitor. Generally, it is easy to make a capacitor with significantly lower losses than the inductor under test, and the losses of the resonant circuit are those of the inductor. In the Q -meter, the resistance of the driving generator is very low and does not contribute to the total losses. Also, a high-impedance

voltmeter is used to measure the voltage across the inductor and is the indicator of Q .

The vacuum tube circuit used as a voltmeter in the early Q -meters was called the infinite impedance detector. Although the input impedance was not infinite, the actual impedance was quite high. This was due to the fact that vacuum tubes are inherently high-impedance devices and the infinite impedance detector used a high level of feedback in the circuit, which further increased the impedance. For a modern Q -meter, a field-effect transistor (FET) would be an acceptable replacement for the high-impedance vacuum tube. The input impedance of the vacuum tube voltmeter circuit is a high resistance in parallel with a small capacitor.

The capacitive part of the input impedance of the voltmeter can be compensated by including the fixed capacitance of the voltmeter in the calibration of the calibrated variable capacitor. The resistive part of the voltmeter would cause errors for extremely high Q . The typical Q -meter is capable of measuring Q values as high as 400 or so with accuracy. Very few conventional inductors would have Q values exceeding this range.

The resistive component of the variable capacitor is due to losses in the capacitor. In order to keep the losses low, the variable capacitor of the Q -meter is carefully constructed of the highest-quality materials. The capacitor is a conventional mechanically adjusted parallel-plate capacitor. The plates are either silver- or gold-plated, and the insulators supporting the fixed plates are a low-loss ceramic or glass. The shaft supporting the rotating plates has a large rotating contact to ensure a low-loss path to ground. The capacitor is mounted to the top of the Q -meter so that the "high" terminal of the inductor is directly connected to the stator plates of the calibrated capacitor. This reduces the amount of residual inductance in the tuned circuit.

In spite of all of the cautions taken to create a low-loss capacitor, the variable capacitor represents exposed plates that can radiate and lose energy. For measurements made at low frequencies, below about 1 MHz, the wavelengths involved are so long relative to the dimensions of the capacitor that no significant radiation occurs. This is not the case when the desired measuring frequency is in the very high frequency (VHF) range and higher. Because of this reason and other factors, Q -meters were made in several frequency ranges.

Special small geometry capacitors were made for the higher-frequency Q -meter to reduce the amount of radiation. Since the amount of capacitance required of a VHF range Q -meter was significantly less than the lower-frequency counterparts, the capacitor loss was mostly achieved by simply scaling down a conventional parallel-plate variable capacitor.

In addition to measuring inductance and Q of an inductor, the Q -meter can measure capacitors. The method of measuring capacitors is to resonate an inductor using the internal calibrated capacitor and adding the unknown capacitor in parallel with the calibrated capacitor. The additional capacitance of the external capacitor will detune the resonant circuit. The circuit can be returned to resonance by adjusting the calibrated capacitor. The amount of capacitance removed from the calibrated capacitor is equal to the amount of capacitance added by the external capacitor.

A set of inductors, called a "standard inductor set," sometimes called a "working inductor," was available from the Q -meter manufacturer to make these capacitor measurements. The actual value of the inductor was not important because it

was necessary to resonate the inductor at the frequency of interest. The inductance needs to be such that the Q -meter will resonate with the external capacitor and without it.

What is important, however, concerning the standard inductors, is that the Q of the working inductor is as high as practical. The inductors were typically large because it was made from large-gauge wire, often Litz wire. The inductor is shielded to prevent radiation.

When an inductor is resonated using the calibrated capacitor in a Q -meter, the inductance of the unknown may be calculated using the familiar resonance equation. This is making the assumption, however, that the unknown is a pure inductance. This is not exactly the case.

Although there are exceptions, inductors are made from wire wound into a solenoid or toroid shape. If a single layer of wire will not produce the necessary inductance, additional layers of wire can be added. In many cases, dozens of layers may be involved with hundreds of turns. When wires lie side by side with only the thin enamel insulation of the wire to separate the wires, capacitance is added to the inductor. This added capacitance is called distributed capacitance and can have an effect on the measurement of inductance.

When an unknown inductor is resonated in a Q -meter, the value of capacitance that is in parallel with the inductance is the external capacitance plus the distributed capacitance. Therefore, the capacitance indicated on the calibrated capacitor is that which is required to resonate the inductor at the desired frequency.

The typical inductor has an equivalent circuit as shown in Fig. 6. If this inductor was to be used in an oscillator in a parallel resonant circuit, the external capacitance required to resonate the inductor in the oscillator needs to be known. The actual inductance is not a factor.

The circuit shown in the figure also has a resonant frequency of its own, with no external components called the self-resonant frequency. When the series impedance of the inductor is an important factor such as the three-pole low-pass filter shown in Fig. 7, the self-resonant frequency of the inductor is important. In this example, the distributed capacitance causes two zeros to appear in the transfer function of the filter which will affect the performance of the filter in the stop band.

The Q -meter can measure the distributed capacitance of an inductor. The inductor is resonated at a frequency near the frequency where the inductor will be operated. The capacitance required for resonance is recorded. Notice that the inductance is not measured; only the capacitance required to resonate the inductor at the set frequency. The inductor is resonated at a second frequency and the capacitance is noted.

Historically, the two frequencies at which the inductor is resonated are a ratio of two to one. In other words, the inductor

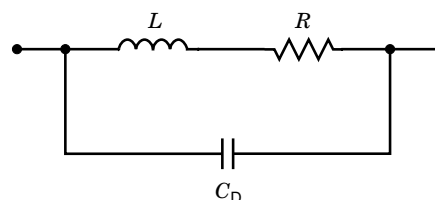


Figure 6. The equivalent circuit for an inductor showing the distributed capacitance.

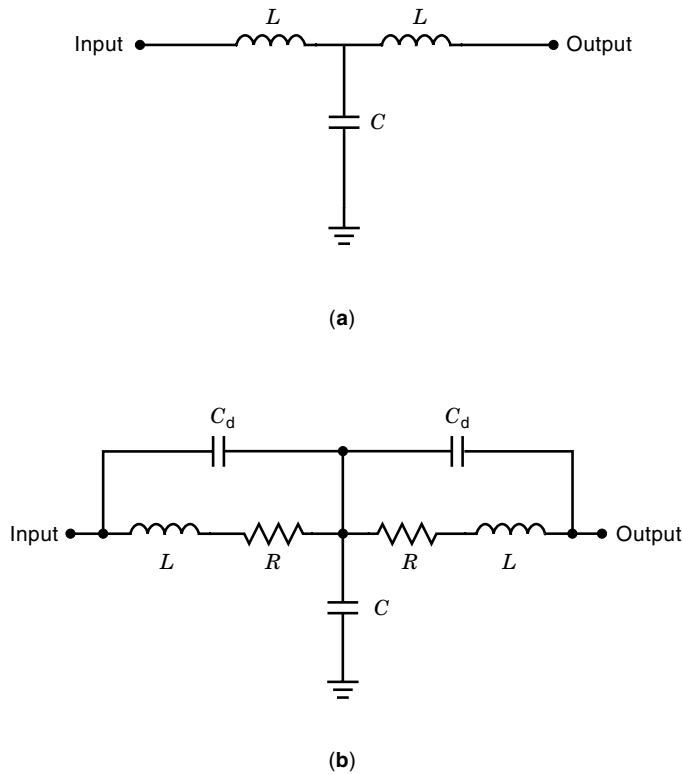


Figure 7. (a) The theoretical schematic of a three-pole low-pass filter. (b) The actual schematic showing indicators with finite Q and distributed capacitance.

is resonated at a frequency and then at twice that frequency. Using this convention the distributed capacitance, C_0 , can be calculated by the following reasoning:

For resonance at a frequency F_0 the resonance equation is

$$F_0 = \frac{1}{2\pi[L(C_0 + C_d)]^{1/2}} \quad (26)$$

where C_0 is the capacitance set on the Q -meter and C_d is the distributed capacitance. The frequency of the Q -meter oscillator is set to twice the previous frequency and the inductor is resonated again. In this case the resonance equation is

$$2F_0 = \frac{1}{2\pi[L(C_1 + C_d)]^{1/2}} \quad (27)$$

where C_1 is the capacitance set on the Q -meter for this higher frequency.

Taking the ratio of the two frequencies, the following result is obtained:

$$2 = (C_0 + C_d)^{1/2} / (C_1 + C_d)^{1/2} \quad (28)$$

Solving for C_d the result is

$$C_d = \frac{C_0 - 4C_1}{3} \quad (29)$$

In many cases it is not desirable to use a frequency ratio as high as two to one, particularly for a measuring frequency that is

near the upper frequency region of the meter. Almost any ratio can be used; but if the ratio is not sufficient, the difference in resonating capacitance for the two frequencies will be small, making it difficult to read an accurate difference. Therefore, the calculated distributed capacitance will be prone to error. Some models of Q -meters had a precise ΔC dial to increase the accuracy of these measurements. One ratio that produces good accuracy is 1.414, or the square root of two. Using this ratio the equation for the distributed capacitance is

$$C_d = (C_0 - 2C_1) \quad (30)$$

Although the Q -meter has been around for more than 60 years and is no longer available as a new instrument, there are thousands of the units in use. The older meters are doomed because their vacuum tubes are no longer available. There is lasting evidence of the Q -meter's long influence on inductors. The variable capacitor dial on the Q -meter was calibrated not only in microfarads and picofarads, but also in inductance. The inductance scale was only accurate if the frequency of the oscillator was set to one of several specific frequencies (250 kHz, 790 kHz, 2.5 MHz, etc.) which were etched on the front panel of the meter. These steps are in multiples of the square root of 10 so that the inductance scale on the variable capacitor may be used in multiples of 10. A new inductor of the latest modern design specifies a Q at a specific frequency which is, more often than not, one of the frequencies etched on the front panel of a 50-year-old Q -meter.

The Q -meter can measure reactances that are low Q , low values of inductance, or high values of capacitance. The difficulty in measuring low inductance is that large capacitance is required to resonate the inductance. For the case of a capacitor, the usual technique of measuring capacitance is to resonate a "working inductor," place the unknown capacitor across the calibrated capacitor, and note how much capacitance is to be removed from the calibrated capacitor to restore resonance. For capacitor values greater than the range of the calibrated capacitor, this could not be done because it would not be possible to remove an amount of capacitance equal to the unknown. To measure large capacitors and small inductors, the unknown circuit element is placed in series with a working inductor; the change in Q and capacitance is determined to find the capacitance, inductance, or Q of the unknown.

Assume that a working inductor is resonated with the Q -meter's calibrated capacitor, C_1 . This capacitance can be represented as

$$L = \frac{1}{4\pi^2 C_1 F_0^2} \quad (31)$$

The unknown is added in series as shown in Fig. 8. If the unknown is a capacitor, C_x , the calibrated capacitor will have to be increased in value because the unknown is in series with the calibrated capacitor and will reduce the total capacitance across the working inductor. The dial capacitance is changed to bring the inductor to resonance. The new capacitance is the combination of the dial and the unknown, and together they are equal to the original capacitance as resonance is reestablished. Therefore,

$$\frac{1}{C_1} = \frac{1}{C_x} + \frac{1}{C_2} \quad (32)$$

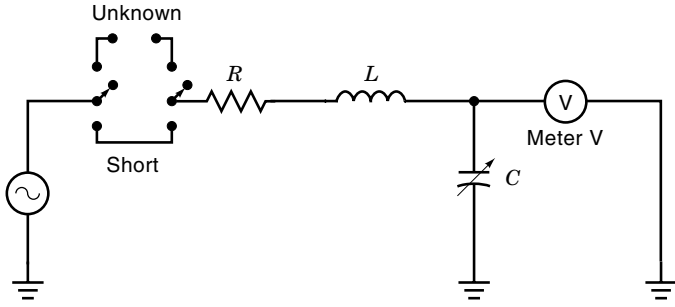


Figure 8. Simplified schematic of the method of measuring small inductors or large capacitors.

where C_x is the unknown capacitance and C_2 is the dial capacitance required to establish resonance with the unknown in place. Solving for C_x , we obtain

$$C_x = \frac{C_1 C_2}{C_2 - C_1} \quad (33)$$

If the unknown is an inductor, L_x , and there is no mutual inductance between the working inductor (this is the reason for shielding the working inductor), the circuit inductance will increase by an amount equal to the added inductance. Therefore, the calibrated capacitance must be reduced to account for the increase in inductance. The new inductance is

$$L_2 = L + L_x = \frac{1}{4\pi^2 C_1 F_0^2} + L_x = \frac{1}{4\pi^2 C_2 F_0^2} \quad (34)$$

Solving for L_x , we obtain

$$L_x = \frac{C_1 - C_2}{4\pi^2 C_1 C_2 F_0^2} \quad (35)$$

To derive the relationship of the measured Q and the Q of the unknown, consider the situation at resonance with just the working inductor. The indicated Q_1 is

$$Q_1 = 2\pi f L / R = \frac{1}{2\pi f C_1 R} \quad (36)$$

When an unknown is added to the circuit, the equivalent series resistances add but the additional reactance is removed by either increasing or decreasing the calibrated capacitance until resonance is restored. The Q is measured by the voltage across the calibrated capacitor and the measured Q is relative to that circuit element. With both the unknown and the working inductor present in the circuit the indicated Q will be Q_2 and the setting of the calibrated capacitor will be C_2 . Thus the total circuit resistance for this situation is

$$R + R_x = \frac{1}{2\pi f C_2 Q_2} \quad (37)$$

Solving for R_x we obtain

$$R_x = \left(\frac{1}{2\pi f} \right) \left(\frac{1}{C_2 Q_2} - \frac{1}{C_1 Q_1} \right) \quad (38)$$

The highest Q that can be measured by the Q -meter is under 1000. Conventional inductors are seldom manufactured with Q values as high as 1000. Ceramic resonators and other electromechanical resonators can have Q values as high as several thousand. Theoretically, the Q -meter could measure very high Q values if the “multiply Q by” control were reduced sufficiently to permit the Q to fall within the range of the meter. Even if this were possible, there is another problem with the use of the Q -meter for very high Q -circuits. The bandwidth of a circuit is inversely proportional to Q , and for very high Q -circuits the bandwidth can be so narrow that it becomes nearly impossible to adjust the frequency of the oscillator in the Q -meter to exact resonance.

An instrument (which is similar to the Q -meter) used for very-high- Q resonators is the crystal impedance meter or CI-meter. The first major difference between the CI-meter and the Q -meter is that the resonator is put in an oscillator circuit such that it is not necessary to adjust the signal generator with great accuracy.

The quality of a quartz or ceramic resonator is usually specified by the equivalent series resistance. Figure 9 shows the equivalent circuit of a quartz or ceramic resonator. The inductor L_0 is due to the motion of the quartz bar and is called the motional inductance. The capacitor C_0 is called the motional capacitance and is also due to the mechanical motion. The capacitance C_p is due to the electrodes plated on the quartz bar and the capacitance of the mechanical supports for the crystal and is called the parallel capacitance of the crystal. R is the equivalent series resistance and represents the losses in the crystal.

There are two resonant frequencies of the quartz resonator: the series and the parallel resonance. The series resonant frequency occurs when the circuit is at the minimum impedance, which is, essentially, the equivalent series resistance. There would also be a reactance due to the capacitance C_p ; but because this reactance is so much higher than the relatively low resistance of R , its effect can be discounted.

In the parallel mode of oscillation, the quartz crystal resonates with its motional capacitance in series with the capacitor C_p , with the two capacitors in parallel with the motional inductance.

In a quartz resonator, the equivalent inductance is in the tens and hundreds of millihenrys while the motional capacitance is very small, on the order of femtofarads. Although the capacitance C_p is small, it is large compared to C_0 . Therefore, the parallel resonant frequency of the quartz crystal is only slightly higher than the series resonant frequency, typically on the order of 0.1%.

The equivalent series resistance of a quartz crystal is measured in two ways. The first method is to place the quartz crystal in a resonant circuit and measure the input and output voltage

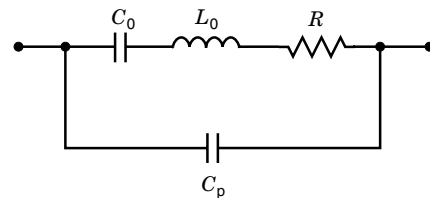


Figure 9. The equivalent circuit of quartz or ceramic resonator.

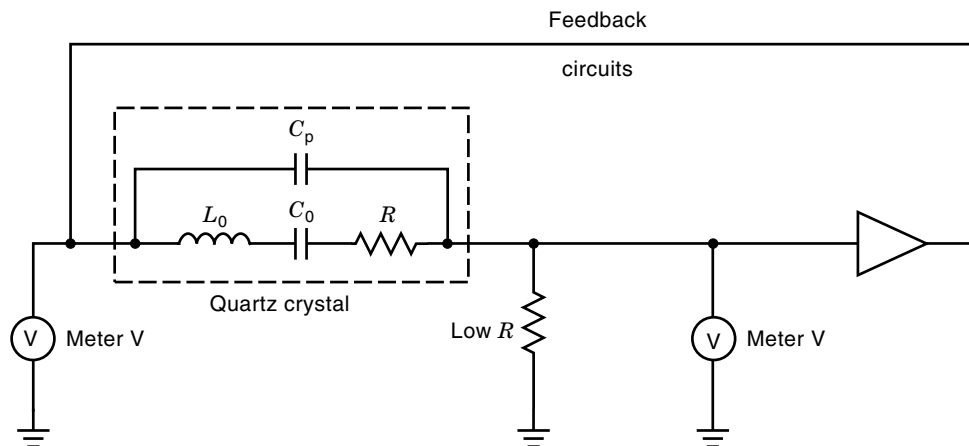


Figure 10. A simplified diagram of a crystal impedance meter.

of the crystal. Then the crystal is replaced with a variable resistor; and because of tuned circuits in the oscillator, the oscillator will continue to oscillate at nearly the resonant frequency of the crystal. The resistance value is adjusted until the input and output voltages and currents are the same as when the crystal was in the oscillator circuit. This method requires manual manipulation; and although it has been used for many years, it cannot be applied to an all-electronic method.

A method that can be adapted to an all-electronic method is similar to a Q -meter in reverse. The oscillator is adjusted to operate in the series mode of the crystal, and the crystal current passes through a low resistance (Fig. 10). Measuring the voltage across the low resistance provides an indication of

the crystal current while the voltage across the crystal provides the applied voltage. The resistance is simply the voltage divided by the current.

BIBLIOGRAPHY

1. Reference Data for Engineers, *Radio Electronics, Computer and Communications*, 8th ed., Indianapolis, IN: Sams, 1993.
2. A. Helfrick and W. Cooper, *Modern Instrumentation and Measurement Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

ALBERT HELFRICK
Embry-Riddle Aeronautical
University

ELECTROMAGNETIC VARIABLES

ELECTRIC STRENGTH

When a dielectric material is subjected to electric voltage, two different types of currents are induced. Applied voltage exerts force on the bound charges that give rise to displacement current. On the other hand, free charge carriers move under the influence of this force, which constitutes the conduction current. There are no free charge carriers in an ideal dielectric material. If the applied voltage is constant with time (dc), then the displacement current shows up only as a transient phenomenon that results in the polarization charge at steady-state condition. Displacement current will be nonzero at the steady state if applied voltage is time-dependent, and it is also known as the capacitive current or the polarization current. Conduction current (or leakage current) produces heat as a result of I^2R loss while the displacement current represents the energy stored. The two currents increase with a rise in applied voltage. This is true as long as the applied voltage is not sufficiently significant to introduce sharp or irreversible changes in the material.

In the foregoing description, it is assumed that the applied voltage remains below a certain limit such that the properties of dielectric material are preserved. However, if the applied voltage is high enough to exert a large force on bound charge carriers, then the material may break down. This change in characteristics of material may be irreversible. In this case, the electric resistance of the dielectric material decreases sharply; this gives rise to a large conduction current flowing through certain channels formed between the electrodes, practically producing a short-circuit. Maximum voltage V_{bd} applied at the moment of breakdown is known as the breakdown voltage of that material. The point of breakdown generates a spark or an electric arc that can fuse, burn or crack the material and the electrodes. When the applied voltage is removed, the solid dielectric may exhibit a trace of breakdown in the form of a punctured, fused or burnt-through hole, generally speaking, of an irregular shape. If this sample is again subjected to electric voltage, then the breakdown occurs in most cases at the already punctured spot even at a rather low voltage. Thus, a breakdown of solid insulation is a failure that puts the device out of operation and requires serious repairs. The liquid as well as gaseous dielectrics generally recover after the breakdown voltage is removed because of the mobility of constituent particles. However, if the power and duration of the electric arc are large, then the entire volume may change and the material may not recover.

Experiments indicate that the breakdown voltage of a material depends on the thickness between the two electrodes. The thicker the material, the more it withstands higher voltages; V_{bd} grows nonlinearly when thickness increases. On the other hand, layers of the same thickness but made of different dielectric material exhibit different breakdown voltage. This indicates that every insulating material has the unique ability to resist breakdown. This parameter is an electric field intensity value known as the electric strength E_{bs} .

For the simplest case of a uniform electric field produced by voltage V_{bd} in a dielectric of thickness d , one can write

$$E_{bs} = \frac{V_{bd}}{d} \quad (1)$$

Dielectric (or electric) strength may be regarded as the electric field intensity at a given point in a dielectric material that causes the breakdown at that location. Generally, the electric strength of an electrical insulating material is understood to mean average breakdown intensity. In Eq. (1), V_{bd} is in volts, d is in meters, and E_{bs} is in V/m. Other commonly used units of electric strength are kilovolts/millimeter, kilovolts/centimeter, and volts/micrometer. Another unit—volt/mil—is frequently used in the United States. Since a mil is one thousandth of an inch, 1 V/mil is equal to 39.37 kV/m. Alternately, 1 MV/m is equal to 25.4 V/mil.

The electric strength depends on various factors including temperature, humidity, frequency, and duration of the applied voltage. Its value varies over a broad range and therefore it is important to include the conditions under which a given datum is measured. For example, the electric strength for mica ranges from 100 to 300 MV/m, from 15 to 25 MV/m for transformer oil, while it is only 2 to 5 MV/m for air under normal conditions of pressure and temperature. Table 1 shows typical electric strength and the dielectric constant of selected materials. The table also shows that high-quality solid dielectrics possess higher electric strength, in general, in comparison with those of liquid and gaseous dielectrics. Therefore, the electric strength of the medium surrounding a specimen needs to be taken into account. For example, consider a comparatively thick solid test sample that is placed in air, as shown in Fig. 1. In this case, an increase in the applied voltage may cause first not a breakdown through the specimen, but a surface discharge or flashover through the surrounding medium.

The aforementioned characteristics lead to the conclusion that solid dense dielectrics should have higher electric strengths in comparison with those porous dielectrics containing many gaseous inclusions. The electric strength of porous dielectrics can be increased appreciably by impregnating it with high-quality liquid or solidifying insulating materials. The operating voltage V_{op} of insulating material in an electrical device (machine, cable, etc.) should be less than the breakdown voltage of that material. The ratio of the breakdown voltage to the operating voltage (V_{bd}/V_{op}) is generally called the safety factor of electric strength of insulation.

Two main kinds of breakdown occur largely in solid dielectrics: the electric and electrothermal breakdown. Electric

Table 1. Dielectric Constant and Electric Strength of Certain Materials (5)

Material	Dielectric Constant	Electric Strength (kV/m)
Air	1	30
Barium titanate	1200	75
Oil	2.3	150
Paper	3	200
Porcelain	7	200
Glass	6	300
Paraffin	2	300
Quartz (fused)	4	400
Polystyrene	2.6	500
Mica	6	2000

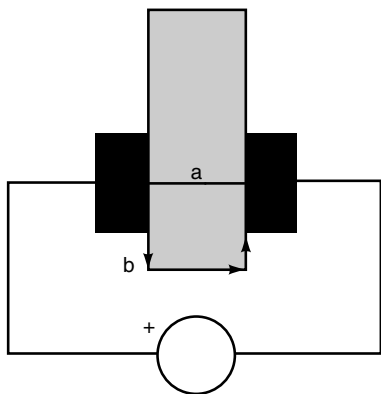


Figure 1. An arrangement of the electrodes with flashover path.

breakdown represents the destruction of a dielectric material due to the force of an applied electric field. It occurs either because accelerating free charged particles (electrons, ions) interact with the material structure or because of inelastic displacement of bound charges in a dielectric under the action of an external electric field. The secondary processes (heating, chemical reactions, and so forth), which may occur in a dielectric under the action of an electric field, never take place in this intrinsic electric breakdown.

Theoretical analysis of electric breakdown is extremely complex. Experimental results frequently differ from those calculated on the basis of the structure and various parameters of materials. Estimated values of intrinsic electric strength are generally higher than the experimental results. This discrepancy is sometimes attributed to minute cracks and other defects present in the material.

Electrothermal breakdown occurs due to loss of electrical energy in a material. The energy loss raises the temperature of insulating material and that, in turn, increases the energy loss even further. This process continues up to the point when the dielectric is fused, burnt, ruptured, or develops cracks.

BREAKDOWN IN LIQUIDS

Petroleum oils are important liquid dielectrics because of their frequent use in transformers, capacitors and cables that operate at high voltage. These are mixtures of various hydrocarbons and serve as neutral or weakly polarized dielectrics. In order to be employed in electrical insulation they must be free of water, oxidation products, aging and other contaminants. Even a small amount of water can influence the electric strength of petroleum oils significantly. This may be attributed to the fact that the dielectric constant of water (around 80) is much higher than that of oil (in the range of 2.2 to 2.4). Water droplets present in oil as emulsion, become elongated by the applied electric force and are broken into still finer droplets. These fine droplets form chains and are drawn into the sites where field intensity is especially high (i.e., towards the edges of the electrodes). It is precisely these sites that form the foci for the development of oil breakdown. The fibrous impurities reduce the electric strength of oil even more sharply.

Electric strength of the liquid dielectric material is found to increase noticeably when its thickness d between the electrodes is reduced. The following empirical relation can be used to

estimate the electric strength of petroleum oil with d ranging from 0.01 to 1.0 mm (1):

$$E_{bd} = 31.3 d^{-0.2085} \quad (2)$$

In Eq. (2), E_{bd} is in MV/m (acting value) and d is in millimeters.

BREAKDOWN IN GASEOUS DIELECTRIC MATERIAL

Besides chemical composition, pressure, and temperature of the gas, the design of electrodes and their separation affect the breakdown voltage and its pattern. In case of a comparatively uniform field, a gradual increase in applied voltage produces a breakdown of the entire gap between the electrodes. Generally, a spark is produced instantaneously and becomes an electric arc if the source of current has sufficient power.

If the field distribution is inhomogeneous, then an increase in voltage may first result in a discharge in the gas that appears only at the points where electric field intensity is strongest (for example, at sharp spots) without expanding over the entire gap between the electrodes. This discharge is commonly known as corona discharge (or for the sake of brevity, simply as corona) and it has a bluish luminescence. It is associated with a characteristic sound, that is, buzzing or crackling. The corona appears mainly because of chemical transformation of gas in that volume and the rapid growth of energy expenditure as voltage is increased. Peek's law may be used to estimate the latter as follows:

$$P_c = Af(V - V_c)^2 \quad \text{for } V \geq V_c \quad (3)$$

where P_c is power expressed in watts that is liberated in corona, and A is a constant that depends on the electrode geometry, gas composition, pressure and temperature. For example, it will be smaller if the electrode surface is made smoother; A has unit of watts/square volt, f represents frequency in hertz; V in volts is the voltage between electrodes, and V_c is voltage at which the corona appears initially.

As voltage is increased further, the corona occupies more and more space, causing a spark or an arc discharge between the electrodes puncturing the gas gap. Ordinarily, it develops during a very short time—of the order of several microseconds or less. In view of this, the breakdown voltage is practically the same for a direct current as well as for alternating current with a frequency up to a few kilohertz. If the breakdown occurs due to voltage pulses of very short duration, then the magnitude of breakdown voltage becomes larger than that of a low-frequency alternating voltage. The ratio of breakdown voltage with a pulse of given duration and form to breakdown voltage under alternating current of low frequency is known as the pulse coefficient β . When an electric field has a fairly high frequency (of the order of 100 MHz or so) the value of β reaches 1.5 to 2.

The influence of pressure and temperature of a gas on its electric strength is of great importance. The electric strength of air rises with increase in pressure above the normal atmospheric value. When it drops below the normal atmospheric pressure, the electric strength of the gas passes through a minimum and then increases appreciably in the region of reduced pressure. This is why gases are employed as an insulating medium at a high or markedly reduced pressure. This

Table 2. Comparative Values of Electric Strength of Gaseous Dielectrics (1)

Gas	Chemical Composition	Molecular Mass	Electric Strength Relative to Air
Helium	He	4	0.06
Hydrogen	H ₂	2	0.6
Carbon dioxide	CO ₂	44	0.9
Nitrogen	N ₂	28	1.0
Freon-12	CCl ₂ F ₂	120	2.6
Perfluoromethylcyclohexane (vapor)	C ₇ F ₁₄	350	6.0
Perfluorodibutylether (vapor)	C ₈ F ₁₈ O	454	7.5
Perfluorodimethylcyclohexane (vapor)	C ₈ F ₁₆	400	8.5
Perfluorophenatrene (vapor)	C ₁₄ F ₂₄	624	10.0

phenomenon can be explained via the theory of collision ionization as a cause of gas puncture. A gas always contains some electrically charged particles (electrons, negative and positive ions) generated by cosmic rays, radiation of radioactive matter, ultraviolet rays and other factors. Similar to the molecules of a gas, these particles are in a state of chaotic thermal motion. These charged particles are accelerated by the applied electric field. Hence, they acquire a velocity that results in higher kinetic energy until they collide with a molecule. If λ denotes the mean distance traversed by a charged particle before colliding with a molecule and the charge on that particle is q , then its kinetic energy in a uniform field with an intensity E at the moment of collision will be given as follows:

$$W = Eq\lambda \quad (4)$$

If this energy is sufficient to ionize the molecule, that is, to separate it into a positive ion and an electron, then the gas may be punctured. This occurs because the new particles generated upon collision ionization will, in turn, be accelerated by the electric field and ionize the other molecules. This process will continue until a complete breakdown occurs in the gap. A rise in pressure (with temperature remaining the same) increases the density of gas, and, in turn, reduces the mean molecular distance and diminishes the value of λ . It is true that λ increases in the region of high vacuum, but there is much less probability of collision between the charged particles and the molecules of rarefied gas. The latter effect dominates the former (the effect of increase in λ). At high vacuum, the value of λ can even exceed the dimensions of the container.

In practical calculations, the effect on the electric strength of air to slight deviations in normal pressure p and temperature T ($p = 760$ mm Hg, $T = 293$ K) is estimated as follows. Assume that V_{bdo} represents the breakdown voltage at normal condition and V_{bd} is its new value. Then

$$v_{bd} = V_{bdo}d \quad (5)$$

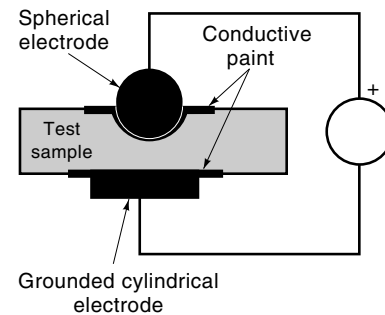
where d is the relative density of air. It is the ratio of the density of air at a pressure of p and a temperature T to the density of air in normal condition, and given by

$$d = 0.385526 \frac{p}{T} \quad (6)$$

In this equation p is measured in mm Hg and T in Kelvin.

Detailed study of the breakdown in gases with various pressure p and the electrode separation t indicates that the magnitude of breakdown voltage in a comparatively homogeneous field depends not on the value of p and t taken separately, but a product of these two parameters; this is known as Paschen's law. A definite minimum breakdown voltage V_{bdm} describes each gas. When the voltage is less than this value, the gas-filled gap of any length and under any pressure can not be punctured. The value of V_{bdm} is 326 V for air for $p \cdot t = 5.67$ mm Hg · mm. In the case of inert gases, V_{bdm} also depends on the material of the electrode (cathode). If the cathode is made of an alkali or alkaline-earth metal that possesses a low work function of electrons (or is at least coated with such metal) then V_{bdm} is reduced. This fact is utilized to make gas-discharge devices.

It may be inferred that gases varying in their chemical composition have different electric strength under the same conditions. Thus, the electric strengths of hydrogen and inert gases, such as argon, neon, and helium, are lower compared with air. There are also gases that possess an electric strength appreciably larger than that of air. The gases with high electric strengths that can be employed as electrical insulation in high-voltage devices (especially at high pressure) have rather high molecular mass and density. These are primarily the gases containing strongly electronegative elements—fluorine, chlorine, and so forth. The electronegativity is an arithmetic sum of the energies of ionization and affinity to electron. The metalloid properties in a given element with higher electronegativity are more pronounced, while the metallic properties manifest themselves more at a lower electronegativity. The high electric strength of electronegative gases is due to the ability of the molecules of these gases to combine easily with free electrons

**Figure 2.** An assembly of spherical and cylindrical electrode geometry.

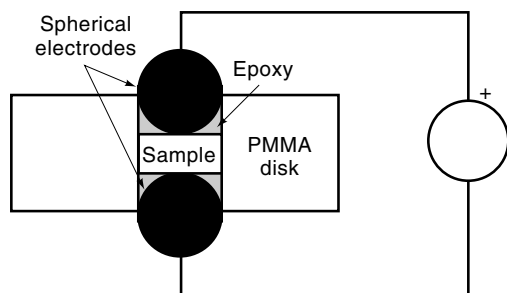


Figure 3. A two-spherical electrode assembly.

or to absorb part of the energy of the electrons colliding with them. Table 2 illustrates some of these properties.

ELECTRODE GEOMETRY USED FOR MEASUREMENTS

The value of electric strength is, in general, influenced by the shape of the electrode. If the measurement conditions are unknown, the specified value of electric strength will have very little practical importance. In that case, the designer must measure it before proceeding with the design. A commonly used electrode system is depicted in Fig. 2. Here, a cylindrical electrode is used as ground while the high voltage is applied through a spherical electrode that is placed over a recessed section of the sample-under-test. Conducting paint or an evaporated metallic film that extends over the normal thickness facilitates the contacts between the electrode and the specimen. Depending on the chemical compatibility, mineral or synthetic oils are used as the immersing medium to control the flashover and corona discharge along an external path. The spherical electrode can be rotated after each breakdown to avoid subsequent breakdown originating from the previously pitted electrode surface. If both electrodes are spherical, then their alignment may become a problem. This can be circumvented by the use of a translucent polymethyl methacrylate (PMMA) disk that has a cylindrical hole with a diameter equal to that of the spherical electrodes, as shown in Fig. 3. The sample-under-test is in the form of a disk without recessed surfaces. It is in contact with the spherical electrodes that together with the specimen are embedded in epoxy resin. This kind of system is preferred for precise measurement of electric strength. Various kinds of cylindrical disk

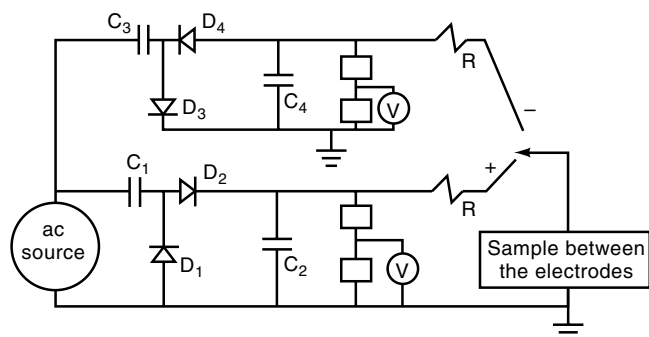


Figure 4. A direct voltage test arrangement for specimen.

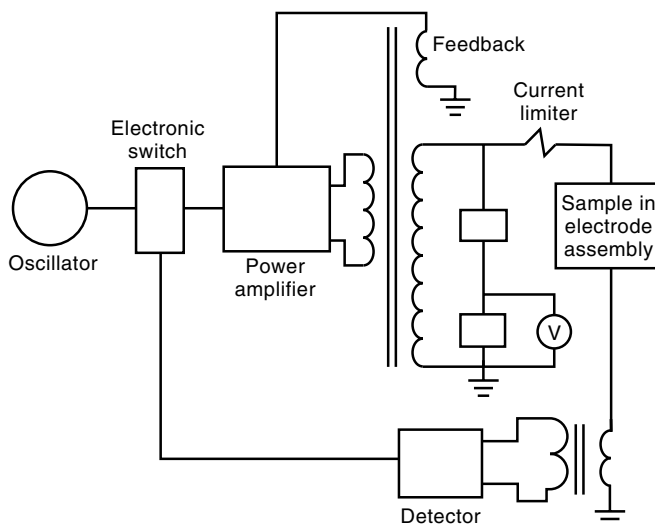


Figure 5. An alternating voltage test arrangement.

electrodes have also been used for routine measurement (2,3). However, those are not described here for the sake of brevity.

TEST APPARATUS

The test circuit requires a means to generate a controlled high voltage. In case of a direct voltage testing, the polarity of applied voltage may be another parameter that needs to be taken into account. Figure 4 shows a test circuit that employs the voltage multiplication to generate high dc from an ac source. A voltmeter connected via a voltage divider circuit monitors the voltage applied to electrodes. The purpose of resistor R is to limit the current at breakdown. It should not have too large a value that can produce large voltage drop. Separate circuits generate the positive and negative voltages.

A low-frequency (up to a few kilohertz) alternating voltage test circuit is illustrated in Fig. 5. An oscillator that feeds the primary side of a transformer via the power amplifier generates the desired frequency signal. A feedback from the transformer is used to regulate the power amplifier. The transformer output is applied to electrodes via a resistor that limits the current and protects the circuit. The detector block senses the breakdown of specimen and signals the electronic switch to interrupt the input.

When the alternating voltage has a fairly high frequency (in microwave range), the power capacity is a more significant specification than the breakdown voltage (4). However, in case of dielectric materials it is perhaps useful to convert the results to a corresponding voltage to permit correlation with data available at the lower frequency. The specimen of a specific geometry is placed inside the microwave cavity and the breakdown power is monitored.

BIBLIOGRAPHY

1. B. Tareev, *Physics of Dielectric Materials*, Moscow: Mir Publishers, 1975.
2. R. Bartnikas, *Engineering Dielectrics*, Special Technical Publication 926, Philadelphia: American Society for Testing and Materials, 1979.

3. W. P. Baker, *Electrical Insulation Measurements*, New York: Chemical Publishing, 1966.
4. M. Sucher and J. Fox (ed.), *Handbook of Microwave Measurements*, Vol. II, Brooklyn, NY: Polytechnic Press, 1963.
5. M. A. Plonus, *Applied Electromagnetics*, New York: McGraw-Hill, 1978.

DEVENDRA K. MISRA
University of Wisconsin, Milwaukee

ELECTROMAGNETIC FIELD MEASUREMENT

To establish standards for conducting electromagnetic (EM) field measurements, measurements must be made (1) in anechoic chambers, (2) at open area test sites, and (3) within guided-wave structures; and a means to transfer these measurements from one situation to another must be developed. The underlying principles of these measurement and transfer standards are: (1) measurements and (2) theoretical modeling. That is, a parameter or a set of parameters is measured, or a parameter is calculated by established physical and mathematical principles.

Various electromagnetic field sensors for measuring radio-frequency (RF) electric and magnetic fields are discussed below. For electric field measurements, electrically short dipole antennas with a high input impedance load, such as a field effect transistor (FET) and a high-frequency diode detector, are discussed. Since the input impedance of an electrically short dipole antenna is predominantly a capacitive reactance, very broadband frequency responses can be achieved with a high impedance capacitive load. However, because conventional dipole antennas support a standing wave current distribution, the useful frequency range of these dipole antennas is usually limited by their natural resonant frequencies. In order to suppress these resonances, a resistively loaded dipole antenna has been developed. To obtain a standard antenna with increased sensitivity at a specific frequency, a half-wave tuned dipole antenna with a diode is used, to measure the induced open circuit voltage. Also used was a tuned receiver with a half-wave tuned dipole antenna, to further improve antenna sensitivity.

For magnetic field measurements, this article discusses an electrically small, resistively loaded loop antenna to achieve a broadband response. Resistive loading is achieved either with the loading resistance at the loop terminal, or by uniform resistive loading along the loop antenna. This short-circuit current loop configuration gives a very flat frequency response over a wide frequency range.

In the region near a transmitting antenna or a scatterer, the electric and magnetic field vectors are not necessarily (spatially) orthogonal or in phase. For time-harmonic fields, the end points of the field vectors trace out polarization ellipses, and the Poynting vectors lie on the surface of a cone with its end point on an ellipse. In these cases, the electric and magnetic fields may be measured separately or, using the single-loop antenna element described in this article, they may be measured simultaneously.

Photonic sensors are also discussed; they provide the wide bandwidth and low dispersion necessary to maintain the fidelity of time-domain signals. Since they consist of electro-optic modulators and optical fibers, they are free from

electromagnetic interference, and there is minimal perturbation of the field being measured.

Throughout the discussion, the interplay between measured quantities and predicted (modeled) quantities is emphasized. The ability of measurements and the restrictions imposed by rigorous theoretical analysis of given models are discussed for the frequencies from 10 kHz to 40 GHz and upward.

ELECTRIC FIELD SENSORS

An Electrically Short Dipole Antenna with a Capacitive Load

Most electric field sensors consist of dipole antennas. The induced open circuit voltage V_{oc} at the dipole antenna terminal is given by

$$V_{oc} = E_{inc} L_{eff} \quad (1)$$

where E_{inc} is the normal incident electric field strength and L_{eff} is the effective length of the dipole antenna. For an electrically short dipole antenna whose physical length is much shorter than the wavelength, the effective length L_{eff} and driving point capacitance C_a are approximately (1)

$$L_{eff} = \frac{L(\Omega - 1)}{4(\Omega - 2 - \ln 4)} \quad (2)$$

and

$$C_a = \frac{4\pi\epsilon_0 L}{2(\Omega - 2 - \ln 4)} \quad (3)$$

where L is the physical length of the dipole antenna, ϵ_0 is the free space permittivity, Ω is the antenna thickness factor $\Omega = 2 \ln (L/a)$, and a is the antenna radius.

For an electrically short dipole antenna with a capacitive load C , the transfer function is given by (1)

$$S(f) = \frac{V_o(f)}{E_{inc}(f)} = \frac{h\kappa/2}{1 + C/C_a} \quad (4)$$

where

$$C_a = \frac{4\pi h}{c\zeta_0(\Omega - 2 - \ln 4)} \quad (5)$$

$$\kappa = \frac{\Omega - 1}{\Omega - 2 + \ln 4} \quad (6)$$

where

- a = antenna radius
- C = capacitance of load
- C_a = capacitance of antenna
- E_{inc} = incident electric field
- h = half of the physical length of the dipole antenna
- V_o = output voltage of the antenna
- ζ_0 = free space impedance
- c = speed of light in free space
- Ω = antenna thickness factor, $\Omega = 2 \ln (2 h/a)$

Since the input impedance of an electrically short dipole antenna is almost purely capacitive, it is possible to achieve

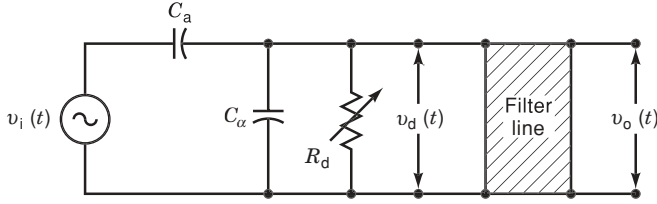


Figure 1. Thevenin's equivalent circuit of an electrically short dipole antenna with a shunt diode detector.

a frequency independent transfer function with a capacitive load, as indicated in Eq. (4). In practice, the load impedance is seldom purely capacitive, but also may have a resistive component. This resistive component will cause a 6 dB per octave roll-off at the low end of the frequency range.

An Electrically Short Dipole Antenna with a Diode

A common standard sensor used for EM field measurements is an electrically short dipole antenna with a diode load connected to a dc metering unit by an RF filter transmission line. The main advantage of including a diode is to make the frequency response of the sensor very flat, even at higher frequencies where an FET amplifier is not readily available. When used with a high-input impedance detector, the sensor can serve as a portable and compact transfer standard.

Figure 1 represents a dipole antenna, which can be used to determine the signal from the transmission line for a given incident electric field (2). This circuit is valid for the frequency range to be considered (dc and $f > 1$ MHz). In Fig. 1, the small antenna resistance has been neglected. The stray gap capacitance of the filter line from the antenna terminal connections has been included in the effective shunt capacitance C_d . The antenna capacitance is given by C_a .

When a first-order, nonlinear differential equation associated with the Thevenin's equivalent nonlinear circuit (shown in Fig. 1) is solved for the detected dc voltage V_o , for a small induced RF voltage V_i (2)

$$V_o = -\frac{\alpha}{4} \left(\frac{V_i}{1 + C_d/C_a} \right)^2 \quad (7)$$

where α ($\simeq 38 \text{ V}^{-1}$) is derived from the diode characteristics, while for large V_i

$$V_o \simeq -\frac{V_i}{1 + C_d/C_a} \quad (8)$$

Equation (7) indicates that for a small induced RF voltage V_i , the output dc voltage V_o is a square law function of the induced voltage. On the other hand, Eq. (8) indicates that, for a large induced voltage V_i , the output dc voltage V_o is directly proportional to the induced voltage.

A Resistively Loaded Dipole Antenna with a Diode

A conventional dipole antenna essentially supports a standing wave current distribution and is, therefore, highly frequency sensitive. For the antenna to have a flat frequency response beyond any natural resonant frequency, a traveling wave dipole antenna was realized by use of continuously tapered resistive

loading (3,4). If the internal impedance per unit length $Z^i(z)$ as a function of the axial coordinate z is expressed as

$$Z^i(z) = \frac{60\Psi}{h - |z|} \quad (9)$$

then the current distribution $I_z(z)$ along the linear antenna is that of a traveling wave

$$I_z(z) = \frac{V_i}{60\Psi(1 - j/kh)} \left[1 - \frac{|z|}{h} \right] e^{-jk|z|} \quad (10)$$

where $2h$ is the dipole antenna's total physical length, k is the wavenumber, V_i is the driving voltage, and Ψ is given by

$$\Psi = 2 \left[\sinh^{-1} \frac{h}{a} - C(2ka, 2kh) - jS(2ka, 2kh) \right] + \frac{j}{kh} (1 - e^{-j2kh}) \quad (11)$$

where a is the radius of the dipole, and $C(x, y)$ and $S(x, y)$ are the generalized cosine and sine integrals. The main advantage of a resistively loaded dipole antenna with a diode is the very flat frequency response of the sensor system. A shortcoming of this sensor system is the relatively low sensitivity. To overcome this problem, a standard half-wave tuned dipole antenna should be used.

A Tuned Half-Wave Dipole Antenna

The magnitude of the electric field component at a given point in an electromagnetic field is determined from the open circuit voltage V_{oc} induced in a standard half-wave receiving dipole antenna as obtained from Eq. (1). The induced voltage is measured across the center gap of the dipole antenna, which is oriented parallel to the electric field vector of the incident field.

The RF voltage induced in the half-wave standard dipole antenna is rectified by a high impedance Schottky barrier diode connected in shunt across the center gap of the antenna. The diode output is filtered by a balanced RC network, and this dc voltage is measured with a high impedance dc voltmeter.

The effective length h_e of a thin dipole antenna near resonance and the required total length L for resonance are given by (1)

$$h_e = \frac{\lambda}{\pi} \tan \left(\frac{\pi h}{\lambda} \right) \quad (12)$$

and

$$L = (\lambda/2) \left[1 - \frac{0.2257}{\ln(\lambda/D) - 1} \right] \quad (13)$$

where D is the diameter of the standard dipole antenna.

To further increase the sensitivity of a standard antenna, a half-wave tuned dipole antenna with a narrow band receiver should be used. In this case, the transfer function $S(f)$ is given by

$$S(f) = \frac{V_L(f)}{E_{inc}(f)} = \frac{h_e Z_a}{Z_a + Z_r} \quad (14)$$

where V_L is the voltage across the receiver load, E_{inc} is the

incident electric field, h_e is the effective length of the tuned dipole, Z_a is the antenna impedance, and Z_r is the receiver input impedance ($\sim 50 \Omega$). The input impedance of the antenna Z_a is complicated and is given in (1). When the cable losses are significant, they should be included in Z_a .

MAGNETIC FIELD SENSORS

An Electrically Small Loop Antenna with a Loading Resistance

A magnetic field sensor consists of an electrically small, balanced loop antenna. The voltage V_i induced in an electrically small loop antenna by an electromagnetic wave incident on the loop antenna is determined from Maxwell's equations and Stokes's theorem, and is given by (1)

$$V_i = \int E_i dl = j\omega\mu H_i NS \quad (15)$$

where E_i is the tangential electric field induced around the loop antenna, l is the circumference of the loop antenna, ω is the angular frequency of E_i , μ is the permeability of the loop antenna core, H_i is the component of the magnetic field normal to the plane of the loop antenna, N is the number of loop antenna turns, and S is the area of the loop antenna. The induced voltage V_i of an electrically small loop antenna is proportional to frequency, the number of loop turns, and the area of the loop antenna.

To make the response of a loop antenna flat over the frequency range of interest, the Q of the antenna has to be reduced through a loading resistance. The resonance of a loop antenna is the result of the combined effect of the distributed capacitance of the loop antenna, the gap capacitance, and the capacitance of the amplifier along with the inductance of the loop antenna. The equivalent circuit for an electrically small loop antenna is shown in Fig. 2. Here V_i is the induced voltage, L is the loop inductance, C is the capacitance, R is the loading resistance, and V_o is the voltage across the loading resistance. Then the response of an electrically small loop antenna is given by (1)

$$\frac{V_o}{V_i} = \frac{-j\frac{1}{\delta}}{\frac{1}{Q} + j\left(\delta - \frac{1}{\delta}\right)} \quad (16)$$

where

$$Q = \frac{R}{X_0}, \quad X_0 = \omega_0 L = \frac{1}{\omega_0 C}, \quad \delta = \frac{\omega}{\omega_0}, \quad \omega_0 = \frac{1}{\sqrt{LC}} \quad (17)$$

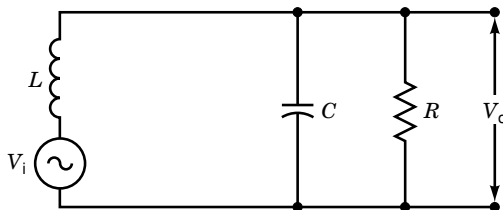


Figure 2. Thevenin's equivalent circuit of an electrically small loop antenna with a loading resistance.

The inductance L and the capacitance C of a loop antenna can be given by

$$L = \mu b \ln\left(\frac{b}{a}\right) \quad (18)$$

and

$$C = \frac{2\epsilon b}{\ln\left(\frac{b}{a}\right)} \quad (19)$$

where μ is the permeability of the medium, b is the loop antenna radius, and a is the radius of the loop wire.

The transfer function of $S(f)$ of an electrically small loop antenna can be obtained by combining Eqs. (15) and (16).

$$S(f) = \frac{V_o}{H_i} = \omega_0 \mu NS \frac{1}{\frac{1}{Q} + j\left(\delta - \frac{1}{\delta}\right)} \quad (20)$$

The normalized transfer function $S_n(f)$ of a loop antenna with a loading resistor is given by

$$S_n(f) = \frac{1}{\frac{1}{Q} + j\left(\delta - \frac{1}{\delta}\right)} \quad (21)$$

as a function of the normalized frequency $\delta = \omega/\omega_0$ and is given in Fig. 3 for various $Q < 1$. Figure 3 shows that the upper frequency end ω_h of the 3 dB roll-off point is given by $\delta_h Q = 1$ and, similarly, the corresponding low frequency ω_l of the 3 dB roll-off point is given by $\delta_l/Q = 1$. Thus, from these conditions,

$$\delta_h \delta_l = \frac{\omega_h \omega_l}{\omega_0^2} = 1 \quad (22)$$

or

$$\omega_0 = \sqrt{\omega_h \omega_l} \quad (23)$$

The self-resonant frequency of a loop antenna is therefore the geometrical mean of the highest and lowest cutoff frequencies.

AN ELECTROMAGNETIC FIELD SENSOR FOR SIMULTANEOUS ELECTRIC AND MAGNETIC FIELD MEASUREMENTS

The electric and magnetic field sensors discussed above measure either the electric or magnetic field only and, therefore, cannot measure complicated EM fields such as those with reactive near field components and multipath reflections. For this reason, a single sensor capable of performing simultaneous electric and magnetic field measurements was developed (5,6). In this case, a loop antenna is loaded at diametrically opposite points with equal impedances. Across one load, the magnetic loop response adds to the electric dipole response, whereas across the other load, the magnetic loop response subtracts from the electric dipole response. Thus, by taking the sum and difference of currents across loads at diametrically opposite points, the magnetic loop response and electric dipole response can be separated. That is, the sum current gives a

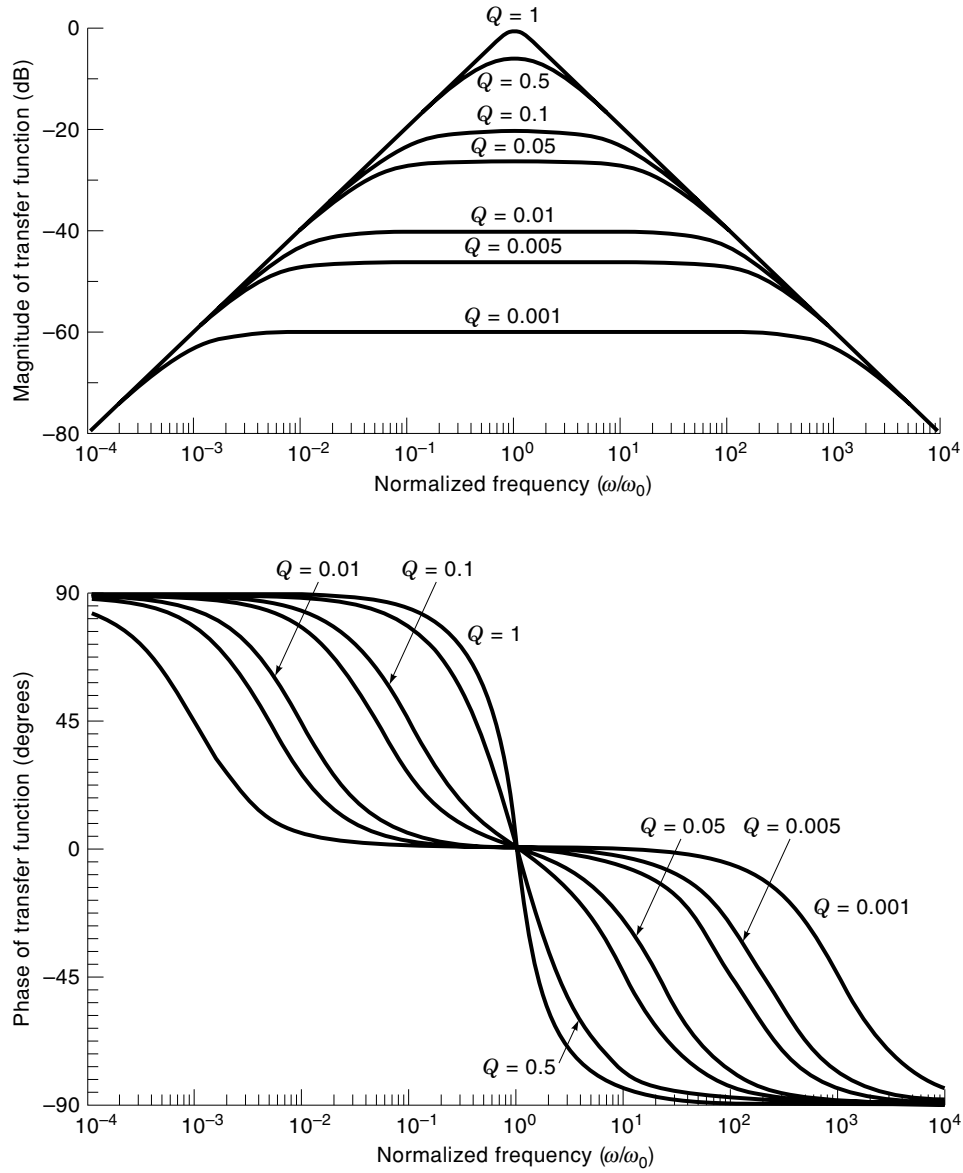


Figure 3. The normalized transfer function of a loop antenna as a function of normalized frequency for different values of Q .

measure of the magnetic field, whereas the difference current gives a measure of the electric field.

To explain the basic characteristics of a doubly loaded loop antenna, the currents I_1 and I_2 at each load are given by (5)

$$I_1 = 2\pi b E_{\text{inc}} \left(\frac{f_0 Y_0}{1 + 2Y_0 Z_L} + \frac{f_1 Y_1}{1 + 2Y_1 Z_L} \right) \quad (24)$$

and

$$I_2 = 2\pi b E_{\text{inc}} \left(\frac{f_0 Y_0}{1 + 2Y_0 Z_L} - \frac{f_1 Y_1}{1 + 2Y_1 Z_L} \right) \quad (25)$$

where b is the radius of the loop, E_{inc} is the incident electric field, Z_L is the load impedance, Y_0 is the admittance for the magnetic loop response, and Y_1 is the admittance for the electric dipole antenna response of a loop. In general, Y_0 is much larger than

Y_1 . f_0 and f_1 are Fourier coefficients of the incident wave. For a loop antenna orientation of maximum electric and magnetic field response, $f_0 = j\beta b/2$ and $f_1 = 1/2$. Taking the sum and difference of these currents yields (5)

$$I_{\Sigma} = \frac{1}{2} (I_1 + I_2) = 2\pi b E_{\text{inc}} \frac{f_0 Y_0}{1 + 2Y_0 Z_L} \quad (26)$$

and

$$I_{\Delta} = \frac{1}{2} (I_1 - I_2) = 2\pi b E_{\text{inc}} \frac{f_1 Y_1}{1 + 2Y_1 Z_L} \quad (27)$$

This indicates that the sum current can be used to measure the magnetic field and the difference current can be used to measure the electric field. In general, $2Y_0 Z_L > 1$ for the magnetic field loop antenna current. Therefore, when the antenna

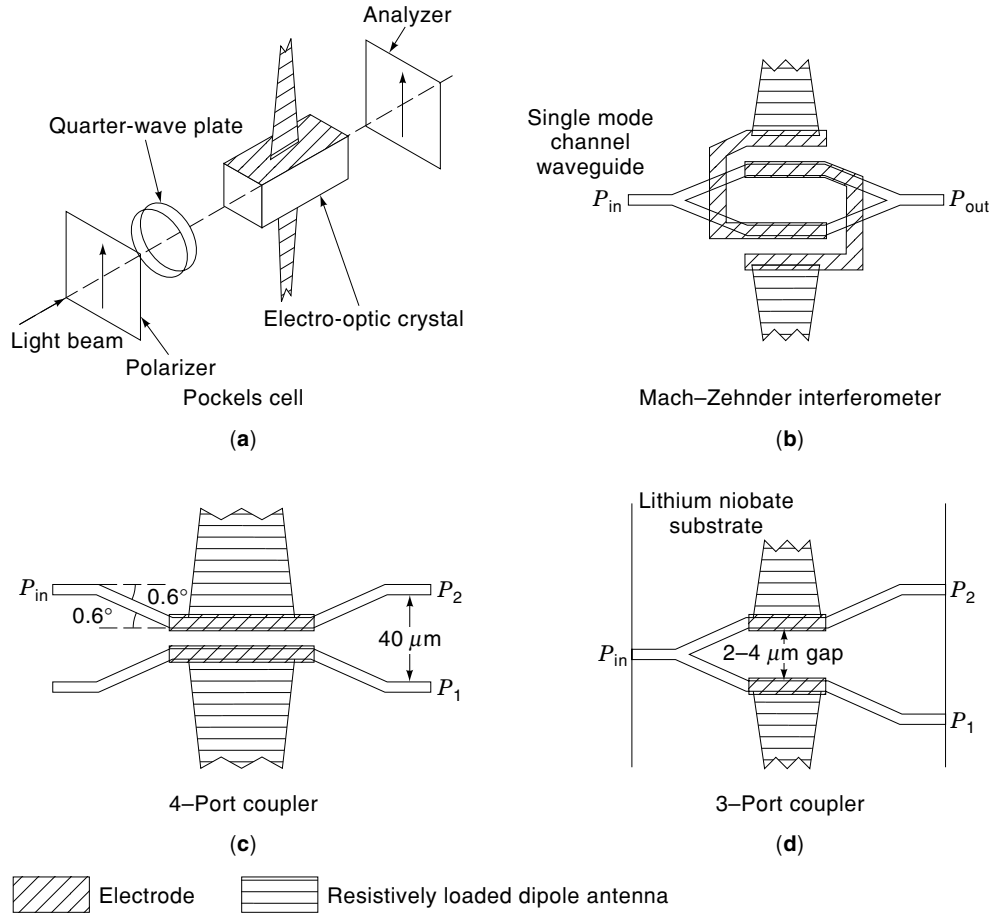


Figure 4. Measurement system of a photonic EM field sensor.

is oriented for maximum response, I_{Σ} can be approximated as

$$I_{\Sigma} \cong j \frac{E_{\text{inc}}}{2Z_L} \pi b^2 \beta \quad (28)$$

This indicates that the magnetic loop current is approximately proportional to the product of frequency and the area of the loop antenna, and is inversely proportional to the load impedance. Similarly, for the electric field dipole current, assuming that $2Y_1 Z_L \ll 1$,

$$I_{\Delta} \cong \pi b E_{\text{inc}} Y_1 \quad (29)$$

which is approximately proportional to the product of the circumference of the loop antenna and frequency, since Y_1 has a capacitive susceptance (positive) and increases with frequency. This device is intended not only to measure the polarization ellipses of the electric and magnetic field vectors in the near field region, but also to measure the time-dependent Poynting vector and thus describe the energy flow.

PHOTONIC ELECTROMAGNETIC FIELD SENSORS

Properly designed photonic EM field sensors provide the wide bandwidth and low dispersion necessary to maintain the fidelity of time-domain signals so that both amplitude and phase information can be retrieved (7). They are free from electro-

magnetic interference, and there is minimal perturbation of the field being measured.

A number of photonic passive sensors for EM field measurements have been reported in the literature (7). These are systems in which the sensor head contains no active electronics or power supplies. Optical measurement systems of the typical photonic EM field sensors is shown schematically in Fig. 4. Light from a laser is launched into an optical fiber link and serves as an optical signal carrier. At the sensor head, the EM field induces a voltage across the modulator crystal and changes its index of refraction. The crystal index changes occur at the frequency of the impressed EM field and result in a modulation in the amplitude of the optical carrier. At the receiver end of the fiber the light is converted to an electrical signal by a photodiode and is suitably amplified for analysis with a spectrum analyzer, oscilloscope, or other signal processor. The electro-optic interaction is weak, and, except for very high fields, the gain of a small antenna is usually required to obtain adequate modulation.

For the measurement of a pulsed electric field, an antenna with a flat broadband response is most desirable. A resistively loaded dipole antenna gives a nonresonant frequency response and is, therefore, ideal for use in the time-domain measurement of electromagnetic fields (3,4). Electro-optical modulators that are driven by antenna feeds are characterized as a function of the voltage applied to their electrodes. A convenient parameter for characterizing a modulator's performance is the voltage V_{π} that is needed across the electrode to drive a modulator

Table 1. Transfer Functions for Electro-Optic Modulators

Modulator	General	Small Signal
Pockels Cell	$\cos^2\left(\frac{\pi}{2}V_n + \frac{\phi_0}{2}\right)$	$\frac{\pi}{2V_\pi}$
Mach-Zehnder	$\cos^2\left(\frac{\pi}{2}V_n + \frac{\phi_0}{2}\right)$	$\frac{\pi}{2V_\pi}$
4-Port Coupler	$\frac{1}{1+3V_n^2}\sin^2\left(\frac{\pi}{2}\sqrt{1+3V_n^2}\right)$	$\approx \frac{1.6}{V_\pi}$
3-Port Coupler	$\frac{1}{2} - \frac{2V_n}{1+4V_n^2}\sin\left(\frac{\pi}{2\sqrt{2}}\sqrt{1+4V_n^2}\right)$	$\frac{2}{V_\pi}\sin\frac{\pi}{2\sqrt{2}} = \frac{1.79}{V_\pi}$

between maximum and minimum optical power levels. In general, a modulator transfer function is nonlinear. However, in most cases of interest, the field-induced RF voltages on the modulator's electrodes are small compared to V_π , and a small signal linear transfer function evaluated at the operating bias point is appropriate. The most desirable operating bias point is where the modulator's sensitivity and linear range are at maximum. The physical characteristics of four modulators used in photonic EM field sensors are given in Table 1.

Pockels Cell

Pockels cell modulators often are used in EM field sensors. The small signal transfer function takes the form given in Table 1, when the constant phase shift Φ_0 is set equal to $-\pi/2$ by introducing a quarter wave retardation plate into the cell. This biases the modulator at its point of maximum sensitivity and linearity. The characteristic voltage of a bulk crystal modulator is limited by (1) the magnitude of the electro-optic coefficients for available materials, (2) the minimum dimensions of crystals that can be handled as discrete optical elements, and (3) the maximum crystals lengths acceptable for the highest desired operating frequency. For operation to 5 GHz, the lowest obtainable values for V_π are on the order of 100 V for LiNbO_3 . It is relatively easy to increase V_π and reduce the modulator's sensitivity in order to measure high fields. In one application of this technology for measuring fields from electromagnetic pulses, a crystal of $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ (BGO), which has cubic symmetry and reduced temperature sensitivity, is used in a Pockels cell modulator and has a V_π of approximately 2100 V.

Mach-Zehnder Interferometer

Modulators based on Mach-Zehnder (MZ) interferometers and fabricated using optical guided wave (OGW) technologies have found widespread use in the fiber optics communications industry. Their use as EM field sensors also has been investigated (7). Photolithography is used to fabricate OGW modulators by defining the waveguide channels and then diffusing titanium into the surface of a LiNbO_3 crystal along these channels. The transfer function for an MZ interferometer is identical to that of the Pockels cell. The static phase shift Φ_0 is due to differences in the optical path length in the interferometer arms. Significantly, V_π for this device is typically two orders of magnitude smaller than that for a Pockels cell modulator of the same material and frequency response. The higher sensitivity is typical of most OGW modulators and arises from the higher fields created in the crystal with the closely spaced electrodes that are

achievable using photo lithography. The principal problems to date with the MZ interferometer have been the difficulty in obtaining the correct value for Φ_0 during fabrication, and its temperature and wavelength dependence.

Directional Coupler

Directional couplers are also OGW devices that have been investigated for use as EM field sensors (7). In the interaction region, the two waveguides lie close enough together that the evanescent field of the lightwave in one guide couples into the other guide. In such a coupled-mode system, the energy in one lightguide can be switched to the other guide by changing the relative propagation constants in the channels with the electro-optic effect. If the fabrication parameters are chosen correctly, the transfer function given in the literature (7) for the directional coupler can be written in the form given in Table 1.

ELECTROMAGNETIC FIELD STANDARDS

Transverse Electromagnetic Cell

Transverse electromagnetic (TEM) transmission line cells are devices used for establishing standard EM fields in a shielded environment (8). Their application is becoming increasingly widespread because of their versatility, measurement accuracy, and ease of operation.

A TEM cell is essentially a 50 Ω triplate transmission line with the sides closed in, to prevent radiation of RF energy into the environment and to provide electrical isolation. A properly designed cell, terminated in its characteristic impedance, is capable of producing a calculable electric and magnetic field for calibrating an electrically small antenna or RF sensor. The cell consists of a section of rectangular coaxial transmission line tapered at each end to adapt to standard coaxial connectors. The line and tapered transitions are designed to have a nominal characteristic impedance of 50 Ω along their length, to ensure minimum voltage standing wave ratio. A fairly uniform EM field is established between the plates inside the cell when RF energy is conducted in the line from a transmitter connected to the cell's input port. A 50 Ω termination is connected to the cell's output port. The expression for determining the electric field E in the cell is given by

$$E = \frac{V}{b} = \frac{\sqrt{PZ_0}}{b} \quad (30)$$

where V is the rms voltage on the septum (center conductor), b is the separation distance between the septum and lower or

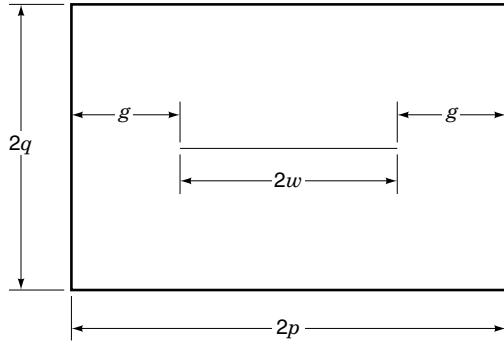


Figure 5. Cross-sectional view of a TEM cell.

upper walls, P is the net power flow to the cell and Z_0 is the real part of the cell's characteristic impedance $\cong 50 \Omega$.

A wave traveling through a cell has essentially the free space impedance ($\cong 120\pi \Omega$), thus providing a close approximation to a far-field plane wave propagating in free space. The design of TEM cells can be based on an approximate equation for the characteristic impedance of a rectangular transmission line (8)

$$Z_0 \cong \frac{377}{4} \left\{ \left[\frac{p}{q} - \frac{2}{\pi} \ln \left(\sinh \frac{\pi g}{2q} \right) \right] - \frac{\Delta c}{\epsilon_0} \right\}^{-1} \quad (31)$$

where p , q , and g are shown in Fig. 5, and $\Delta c/\epsilon_0$ is related to the fringing capacitance between the edges of the septum and the side walls. For large gaps ($g/p > 0.2$) this fringing term approaches zero (8).

The upper useful frequency for a cell is limited by distortion in the test field caused by multimoding and resonances that occur within the cell at frequencies above the cell's multimode cut off. Resonant frequencies associated with these modes can be found from the expression (8)

$$F_{\text{res}} = \sqrt{f_{mn}^2 + \left(\frac{c\ell}{2L} \right)^2} \quad (32)$$

where f_{mn} are the frequencies of the higher-order mode(s) excited inside the cell, c is the wave propagation velocity ($\cong 3.0 \times 10^8$ m/s), L is the resonant length of the cell in meters, and ℓ , m , and n are integers corresponding to multiples of the resonant length and the particular waveguide mode. The influence of the first-order TE modes does not become significant until approaching their resonances. Since most cells are designed with the center plate (septum) centered symmetrically, the odd-order TE modes are not excited in an empty cell. The presence of a device placed in the cell will, however, excite these modes in varying degrees, depending on its size, shape, and placement.

Waveguide Chamber

For the frequency range of 300 MHz to 1000 MHz, one can use a waveguide section with a rectangular cross-section with a width-to-height (aspect) ratio of two to one. The length of a guide "cell" must exceed two wavelengths over the specified frequency band, in order to create a fairly uniform field within the guide. Electromagnetic power is transmitted through the

guide to a matched resistive load, and the maximum frequency is limited by the requirement that power propagates in the guide in the dominant TE_{10} mode. In this well-known case, the direction of the electric field vector is across the narrow face of the guide.

Assuming good conductivity of the waveguide walls, an air dielectric, and sinusoidal excitation, the lowest cutoff frequency f_{co} is

$$f_{\text{co}} = \frac{c}{2a} \quad (33)$$

where c is the wave propagation velocity ($\cong 3.0 \times 10^8$ m/s) and a is the guide width.

The longest or cutoff wavelength is given by $\lambda_{\text{co}} = 2a$. The wavelength λ_{wg} inside the guide for these operating conditions is

$$\frac{1}{\lambda_{\text{wg}}} = \sqrt{\left(\frac{1}{\lambda_0} \right)^2 - \left(\frac{1}{\lambda_{\text{co}}} \right)^2} \quad (34)$$

where λ_0 is the free space wavelength.

The transverse impedance Z_w of the wave traveling in the guide is

$$Z_w = \frac{-E_y}{H_z} = \zeta_0 \left(\frac{\lambda_{\text{wg}}}{\lambda_0} \right) = \frac{\zeta_0}{\sqrt{1 - (\lambda_0/2a)^2}} \quad (35)$$

where ζ_0 is the intrinsic impedance of free space ($\cong 120\pi$).

The direction of the electric field vector is across the narrow face of the guide, and its rms magnitude at the center of a rectangular waveguide is given by

$$E = \sqrt{\frac{2Z_w P_z}{ab}} \quad (36)$$

and, similarly, the rms magnitude of the magnetic field in the guide center is given by

$$H = \sqrt{\frac{2P_z}{Z_w ab}} \quad (37)$$

where P_z is the total power flow in the guide, and ab is the cross-sectional area of the waveguide ($= 0.5 a^2$).

Microwave Anechoic Chambers

Microwave anechoic chambers currently are used for a variety of indoor antenna measurements, electromagnetic field measurements, electromagnetic interference (EMI) measurements, and electromagnetic compatibility (EMC) measurements. The primary requirement is that a transmitting antenna at one location within a chamber or at a chamber wall generates a known field throughout a volume of the chamber, which has dimensions sufficient to perform EM field measurements. This volume is frequently called a *quiet zone*, and the level of reflected waves within it will determine the performance of the anechoic chamber.

Electromagnetic field measurements in an anechoic chamber usually are performed in the near-field region of a transmitting standard antenna. To establish the standard field, the

radiated field intensity in the near-field region of the transmitting antenna is calculated. The antennas typically used for the anechoic chamber measurements consist of a series of open-ended waveguides at frequencies below 450 MHz, and a series of rectangular pyramidal horn antennas at frequencies above 450 MHz.

The electric field strength at a specific distance from the radiating antenna is calculated from measurements of the power delivered to the transmitting antenna and a knowledge of the gain of the antenna as a function of frequency and distance to the field point. The equation used to calculate the electric field E on the boresight axis of the transmitting antenna is

$$E = \frac{1}{d} \sqrt{\frac{\zeta_0 P_{\text{net}} G}{4\pi}} \cong \frac{\sqrt{30 P_{\text{net}} G}}{d} \quad (38)$$

where P_{net} is the net power delivered to the transmitting antenna, ζ_0 is the free space impedance ($\cong 120\pi \Omega$), G is the near-field gain of the transmitting antenna at the given frequency and distance, and d is the distance from the center of the aperture of the transmitting antenna (horn or open-ended waveguide) to the on-axis field point.

The net power P_{net} delivered to the transmitting antenna is the difference between the incident P_{inc} and reflected P_{refl} powers as measured with a calibrated directional coupler (4 ports) with calibrated power meters. In order for measurements to be accurate, an anechoic chamber must provide a truly free space test environment. The performance of a rectangular RF anechoic chamber can be determined by measuring the relative insertion loss versus separation distance between a source antenna and a receiving antenna.

Antenna insertion loss is the ratio of power received by a receiving antenna or probe to the power accepted by the transmitting antenna. If the anechoic chamber is a perfect free-space simulator, the relative insertion loss between two polarization-matched antennas will vary with distance according to the Friis transmission formula (9)

$$P_r/P_t = G_r G_t (\lambda/4\pi d)^2 \quad (39)$$

where P_t is the net power delivered to the transmitting antenna, P_r is the power received by the receiving antenna, G_t is the near-field gain of the transmitting antenna, G_r is the near-field gain of the receiving antenna, d is the separation distance between the two antennas, and λ is the wavelength. Experimental data can be compared with the calculated free space transmission loss, using appropriate near-field transmitting antenna gains. The difference between the measured and calculated transmission loss is a measure of reflections from chamber surfaces.

Open Area Test Sites

An open area test site typically is used for the antenna calibration in the frequency range of 10 kHz to 1000 MHz. A calibration consists of determining the antenna factor that permits a receiver (RF voltmeter) to be used with the calibrated antenna to conduct measurements of field strength.

At frequencies below about 50 MHz, loop antennas are calibrated in a quasi-static, near-zone, magnetic field produced by a balanced single turn transmitting loop with a 10 cm radius. Above 25 MHz, dipole antennas are calibrated in a far-zone

electric field, which is evaluated in terms of the open circuit voltage induced in a self-resonant receiving dipole antenna. Between 30 kHz and 300 MHz, vertical monopole antennas and small probes are evaluated in an elliptically polarized electromagnetic field produced by a transmitting monopole antenna above a conducting ground screen.

Field strength can be evaluated using two independent techniques: (1) the standard field method, and (2) the standard antenna method. For the standard field method, a transmitted field is calculated in terms of the type and dimensions of a transmitting antenna, its current distribution or net delivered power, the frequency of the transmitted signal, the distance from the transmitting antenna to the field point, and the effect of ground reflections (if present). For the standard antenna method, an unknown field is measured with a calculable receiving antenna. The voltage or current induced in a standard antenna by the component of field being evaluated is measured. The field strength then is calculated in terms of this induced voltage, the dimensions and form of the receiving antenna, and its orientation with respect to the field vector.

All of the techniques described above for field strength standards are applicable only to steady state RF fields with sinusoidal time variation. They are not intended for use with pulsed fields or other broadband applications.

Magnetic Field Strength Standards for Loop Antennas at 10 kHz to 50 MHz (Standard Field Method)

The response of an electrically small receiving loop antenna is proportional to the average normal component of magnetic field strength incident on the antenna. A calculable quasistatic magnetic field can be produced to calibrate these antennas using a circular single-turn balanced transmitting loop. Up to 30 MHz, the current in a loop with a 10 cm radius is approximately constant in amplitude and phase around the loop. The receiving loop antenna being calibrated is positioned on the same axis as the transmitting loop at a distance of 1.5 m to 3 m. The normal component of the magnetic field, averaged over the area of the receiving loop, is given by (10)

$$H = \frac{\beta I r_1}{r_2} \sum_{m=0}^{\infty} \frac{1}{(2m+1)!} \cdot \frac{1 \cdot 3 \dots (2m+1)}{2 \cdot 4 \dots (2m+2)} \left[\frac{\beta r_1 r_2}{R_0} \right]^{m+1} h_{2m+1}^{(2)}(\beta R_0) \quad (40)$$

where

H = rms value of the magnetic field

I = rms current in the transmitting loop

r_1 = radius of the transmitting loop

r_2 = radius of the receiving loop

$R_0 = \sqrt{d^2 + r_1^2 + r_2^2}$

d = axial distance between the two loops

$\beta = 2\pi/\lambda_0$

λ_0 = free-space wavelength

$h_n^{(2)}$ = nth order spherical Hankel function of the second kind

The current in the transmitting loop antenna is measured with a vacuum thermocouple calibrated with direct current. The thermocouple is at the top of the loop winding.

While coaxial loop antennas normally are used for calibration purposes, the two loop antennas also can be positioned in the same plane. Coplanar loop antennas are advantageous under certain conditions (e.g., with some ferrite core antennas in which the core length is large). In the coplanar loop antenna set-up, the calibrating value of H would be half of that given by Eq. (41).

The calibration and subsequent measurement of magnetic field strength H often are expressed in terms of the electric field E that would exist if the measurement were made in free space, in which case $E/H \cong 120\pi \Omega$. When such a field strength meter is used to make measurements near the ground, the indicated value of the electric field is not necessarily valid. The same is true for measurements made in the near zone of a transmitting antenna. However, the value of the magnetic component H still can be measured correctly.

For calibrating loop antennas or magnetic field sensors at a higher field, it is possible to use the calculable magnetic field generated in a TEM cell, or a waveguide chamber, or at the center of a flat multiturn coil, or at the midpoint of a Helmholtz coil pair.

Electric Field Strength Standards for Dipole Antennas from 25 MHz to 1000 MHz (Standard Antenna Method)

The magnitude of the electric field component at a given point in a locally generated field is determined from the open circuit voltage V_{oc} induced in a standard (calculable) half-wave receiving dipole antenna. The induced voltage is measured across the center gap of the dipole antenna, which is oriented parallel to the electric field vector of the incident field. In using the standard antenna method, a plane wave field can be generated by a suitable transmitting antenna, such as a log periodic or half-wave dipole antenna. The magnitude of this incident field is measured with the standard dipole antenna by the relation

$$E_{inc} = \frac{V_{oc}}{L_{eff}} \quad (41)$$

where E_{inc} is the field strength of the locally generated field, V_{oc} is the open circuit voltage induced in the standard dipole antenna, and L_{eff} is the effective length of the standard dipole antenna.

The RF voltage V_{oc} picked up by the $\lambda/2$ standard dipole is detected by a high-impedance Schottky barrier diode connected in shunt across the center gap of the antenna. The diode output is filtered by a balanced RC network, and this dc voltage is measured with a high-impedance dc voltmeter. The RF-to-dc characteristic of the dipole antenna and its filter circuit is obtained experimentally. Assuming a cosinusoidal current distribution on an infinitesimally thin dipole, the effective length of a half-wave dipole antenna in free space is given by Eq. (12).

Electric Field Strength Standards for Vertical Monopole Antennas 30 kHz to 300 MHz (Standard Field Method)

Several approaches were considered for generating a standard (calculable) field to calibrate vertically polarized antennas. The system chosen for this measurement consists of a thin cylindrical transmitting monopole antenna over a metallic ground plane. The field strength is calculated in terms of the magnitude and distribution of the monopole antenna current, and other factors such as: (1) monopole height, (2) horizontal distance

from the transmitting antenna to the field point, (3) vertical height of this point above the ground plane, and (4) electrical conductivity of the ground plane.

The height of the transmitting monopole antenna is adjustable, with a maximum height of about 3 m. The electrical height of this antenna is $\lambda/4$ (resonant) at 25 MHz, but only 0.0003λ at 30 kHz. At frequencies above 25 MHz, the antenna height is reduced to a $\lambda/4$ value. The base diameter of the monopole antenna is about 1 cm. The monopole antenna is excited through a coaxial cable from a transmitting room located beneath a concrete ground slab which is covered by a conducting metal screen to form the electrical ground plane.

Equations (42), (43), and (44) give the magnitudes of the three field components E_z , E_ρ , and H_ϕ , respectively, of a transmitting $\lambda/4$ monopole antenna above a perfect ground plane of infinite extent:

$$E_z = 30I_0 \left(\frac{e^{-j\beta r_1}}{r_1} + \frac{e^{-j\beta r_2}}{r_2} \right) \quad (42)$$

$$E_\rho = \frac{30I_0}{r_0} \left[\left(\frac{e^{-j\beta r_1}}{r_1} \right) \left(z - \frac{\lambda}{4} \right) + \left(\frac{e^{-j\beta r_2}}{r_2} \right) \left(z + \frac{\lambda}{4} \right) \right] \quad (43)$$

$$H_\phi = \frac{I_0}{4\pi r_0} (e^{-j\beta r_1} + e^{-j\beta r_2}) \quad (44)$$

where

E_z = vertical electric field component

E_ρ = horizontal electric field component

H_ϕ = magnetic field, encircling the monopole antenna

I_0 = rms base current of the monopole antenna

$\beta = 2\pi/\lambda$ = the wavelength constant

$r_1 = [d^2 + (z - l)^2]^{1/2}$

$r_2 = [d^2 + (z + l)^2]^{1/2}$

$r_0 = [d^2 + z^2]^{1/2}$

l = monopole antenna length

d = horizontal distance between the monopole antenna and the field point

z = vertical distance from the ground plane to the field point

For frequencies near self-resonance, the monopole antenna base current is measured with an RF ammeter consisting of a thermo-converter that has been calibrated with known values of dc current. At lower frequencies, where the monopole antenna input impedance Z_{in} is a high-capacitive reactance, the base current is calculated from Ohm's law in terms of the base voltage measured with a high-input impedance voltmeter and the theoretical input impedance. At very low frequencies, Z_{in} may be calculated from the antenna capacitive reactance (11).

$$Z_{in} = 1/(j\omega C_a) \quad (45)$$

where

$$C_a = \frac{5.56 \times 10^{-11} h}{\ln(h/a) - 1} \quad (46)$$

where C_a is the monopole antenna input capacitance (F), h is the monopole antenna height (m), and a is the monopole antenna radius (m).

The standard field equations are relatively simple for a ground plane with infinite extent and infinite conductivity. In addition, the current on a vertical monopole antenna with finite diameter departs from the sinusoidal current distribution of a filamentary monopole antenna. This does not seriously affect the calculated values of current-related field components, such as the magnetic field or the far-zone electric field. However, the low-frequency near-zone quasi-static electric field components are more nearly charge-related and are given by the spatial derivative of the current distribution. Hence, there is greater uncertainty in calculating the electric field components at frequencies well below that of a $\lambda/4$ -resonant monopole antenna.

If a transmitting monopole antenna is electrically short; that is, if the height is less than $\lambda/4$ and the frequency is below resonance, the current distribution is triangular. The field equations are a little more complicated; only the vertical electric field is given below (11)

$$E_z = \frac{-j30I_0}{\sin(\beta\ell)} \left[\frac{e^{-j\beta r_1}}{r_1} + \frac{e^{-j\beta r_2}}{r_2} - 2\cos(\beta\ell)e^{-j\beta r_0} \right] \quad (47)$$

The EM field values in the half space above a perfect ground are the same as those in each half volume of a center fed $\lambda/2$ dipole antenna in free space. The input impedance of a monopole antenna above perfect ground is half that of a dipole antenna in free space. The power required to generate a given field strength is half that required for a dipole antenna, but the radiated power goes into half the volume, so the field is the same. Measurements of Z_{in} with a commercial impedance meter are performed to check the theoretical values from 0.5 MHz to 50 MHz. Measurements of the monopole antenna capacitance can be made at lower frequencies with a commercial Q meter.

FUTURE DIRECTIONS

Established techniques for EM field measurements will be probably extended to higher frequencies; however, such work would not involve fundamental changes in the instrumentation or measurement strategy. The measurement methods described earlier are suitable only for (1) measuring plane-wave sinusoidal fields of a given frequency and (2) calibrating the devices that measure such fields.

The challenge of the future resides in the development of standards for measurement of fields that are nonsinusoidal and/or nonplanar. The fundamental requirement for EM field sensors is that their outputs provide amplitude and phase information simultaneously over a broad spectrum for nonsinusoidal fields. For fields containing more than one frequency component, it is clear that such simultaneous measurement is necessary. Similarly, if phase information is preserved, then measurements of single-frequency nonplanar fields can be made in terms of true energy density or in terms of the Poynting vector.

The standard techniques of today rely upon very high-resistance transmission lines to convey dc voltage and current to regions external to the field under measurement. Future standards of measurement employ optically sensed EM field sensors whose optical-sensing signal paths will not perturb the EM field under measurement, but will also convey field

amplitude and phase information to a region external to the field for measurement. Optically sensed sensors are being built today. However, additional work is needed to transform these experimental devices into stable elements for standards applications. Also, standard fields need to be developed, in order to calibrate these probes of the future.

BIBLIOGRAPHY

1. M. Kanda, Standard probes for electromagnetic field measurement, *IEEE Trans. Antennas Propag.*, **AP-41**: 1349–1363, 1993.
2. M. Kanda, Analytical and numerical techniques for analyzing an electrically short dipole with a nonlinear load, *IEEE Trans. Antennas Propag.*, **AP-28**: 71–78, 1980.
3. M. Kanda, A relatively short cylindrical broadband antenna with tapered resistive loading for picosecond pulse measurements, *IEEE Trans. Antennas Propag.*, **AP-26**: 439–447, 1978.
4. M. Kanda and L. D. Driver, An isotropic electric-field probe with tapered resistive dipoles for broadband use, 100 kHz to 18 GHz, *IEEE Trans. Microw. Theory Tech.*, **MTT-35**: 124–130, 1987.
5. M. Kanda, An electromagnetic near-field sensor for simultaneous electric and magnetic field measurements, *IEEE Trans. Electromagn. Compat.*, **EMC-26**: 102–110, 1984.
6. L. D. Driver and M. Kanda, An optically linked electric and magnetic field sensor for Poynting vector measurements in the near fields of a radiating source, *IEEE Trans. Electromagn. Compat.*, **30**: 495–503, 1988.
7. M. Kanda and K. D. Masterson, Optically sensed EM-field probes for pulsed fields, *Proc. IEEE*, **80**: 209–215, 1992.
8. J. C. Tippet, *Model characteristics of rectangular coaxial transmission line*, Ph.D. Dissertation, E.E. Department, University of Colorado, Boulder, 1978.
9. S. A. Schelkunoff and H. T. Friis, *Antennas, Theory and Practice*, New York: Wiley, 1952.
10. F. M. Greene, The near-zone magnetic field of a small circular-loop antenna, *J. Res. Nat. Bur. Stand. U.S.C. Eng. and Instr.*, **71C** (4): 1967.
11. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1968.

MOTOHISA KANDA
National Institute of Standards
and Technology

MAGNETIC FIELD MEASUREMENT

RELEVANCE OF ELECTROMAGNETIC FIELD MEASUREMENTS

The measurement of electromagnetic (em) fields is relevant for various purposes: for scientific and technical applications, for radio propagation, for Electromagnetic Compatibility (EMC) tests (i.e. testing of the immunity of electronic equipment to electromagnetic fields and the measurement of radiated electromagnetic emissions aiming at the protection of radio reception from radio interference), and for safety reasons (i.e. the protection of persons from excessive field strengths). For radio propagation and EMC measurements, below about 30 MHz a distinction is made between electric and magnetic components of the em field to be measured. In the area of human safety, this distinction is continued to even higher frequencies.

QUANTITIES AND UNITS OF MAGNETIC FIELDS

Especially in the measurement of radio propagation and of radio interference, magnetic field measurements with loop antennas have traditionally been used to determine the received field intensity which was quantified in units of the electric field strength, i.e. in $\mu\text{V/m}$, respectively, in $\text{dB}(\mu\text{V/m})$. For radio propagation this can be justified for far field conditions where electric field strength E and magnetic field strength H are related via the impedance Z_0 of the free space; $E = HZ_0$ (see also antenna factor definition). Commercial EMC standards (1) and (2) specify radiated disturbance measurements below 30 MHz with a loop antenna; however, until 1990 measurement results and limits were expressed in $\text{dB}(\mu\text{V/m})$. Since this measurement is done at less than the far field distance from the equipment under test (EUT) over a wide frequency range, the use of units of the electric field strength was difficult to justify. Therefore, the CISPR (the International Special Committee on Radio Interference) decided in 1990 to use units of the magnetic field strength $\mu\text{A/m}$, respectively, $\text{dB}(\mu\text{A/m})$.

Guidelines and standards for human exposure to em fields specify the limits of electric and magnetic fields. In the low frequency range [i.e., below 1 MHz (3)], limits of the electric field strength are not proportional to limits of the magnetic field strength. Magnetic field limits in frequency ranges below 10 kHz are frequently expressed in units (T and G, for Tesla and Gauss) of the magnetic flux density B despite the absence of magnetic material in human tissue. Some standards specify magnetic field limits in A/m instead of T (see (4) in contrast to (5)). For easier comparison with other applications we therefore convert limits of the magnetic flux density to limits of the magnetic field strength using $H = B/\mu_0$ or $1 \text{ T} = 10^7/4\pi \text{ A/m} \approx 0.796 \cdot 10^6 \text{ A/m}$ and $1 \text{ G} = 79.6 \text{ A/m}$. At higher frequency ranges all standards specify limits of the magnetic field strength in A/m. Above 1 MHz the limits of the magnetic field strength are related to limits of the electric field strength via the impedance of the free space. Nevertheless both quantities, electric and magnetic fields, have to be measured, since in the near field the exposition to either magnetic or electric field may be dangerous.

RANGE OF MAGNETIC FIELD LEVELS TO BE CONSIDERED FOR MEASUREMENT

In order to show the extremely wide range of magnetic field levels to be measured, we give limits of some national or regional standards. In different frequency ranges and applications magnetic field strength limits vary from as much as 10 MA/m down to less than 1 nA/m (i.e. over 16 decades). This wide range of field-strength levels will normally not be covered by one magnetic field meter. Different applications require either broadband or narrowband equipment.

On the high level end there are safety levels and limits of the magnetic field strength for the protection of persons which vary from as much as 4 MA/m (i.e. $4 \times 10^6 \text{ A/m}$ corresponding to the specified magnetic flux density of 5 T in nonferrous material) at frequencies below 0.1 Hz, to less than 0.1 A/m at frequencies above 10 MHz (see Fig. 1) (3–6). These limits of the magnetic field strength are derived from basic limits of the induced body current density (up to 10 MHz), respectively, basic limits of the specific absorption rate (SAR, above 10 MHz). There are also

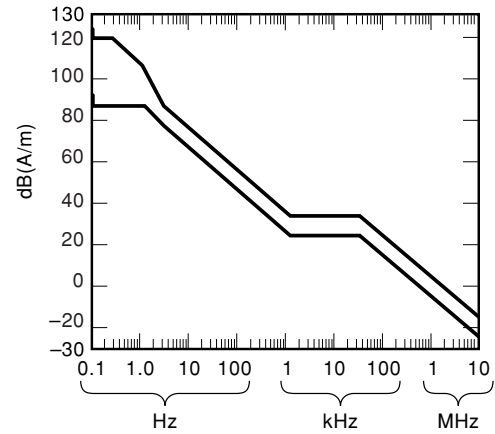


Figure 1. Safety limits of the magnetic field strength derived from the European Prestandard ENV 50166 Parts 1 and 2: 120 dB(A/m) are equivalent to 1 MA/m corresponding to 1.25 T, 0 dB(A/m) are equivalent to 1 A/m.

derived limits of the electric field strength which are however not of concern here.

By using an approach different from the one of the safety standards, the Swedish standard MPR II, which has become an international de-facto standard for video display units (VDU) without scientific proof, specifies limits of the magnetic flux density in two frequency ranges, which are bounded by filters: a limit of 40 nT ($\approx 0.032 \text{ A/m}$) in the range 5 Hz to 2 kHz and a limit of 5 nT ($\approx 0.004 \text{ A/m}$) in the range 2 kHz to 400 kHz.

On the low level end there are limits for the protection of radio reception and electromagnetic compatibility in some military standards (see Figs. 2 and 3).

International and national monitoring of radio signals and the measurement of propagation characteristics require the measurement of low-level magnetic fields down to the order of $-30 \text{ dB}(\mu\text{A/m})$; see also subsequent discussions and refs. 7–9. For the protection of radio reception, international, regional (e.g. European) and national radiated emission limits and

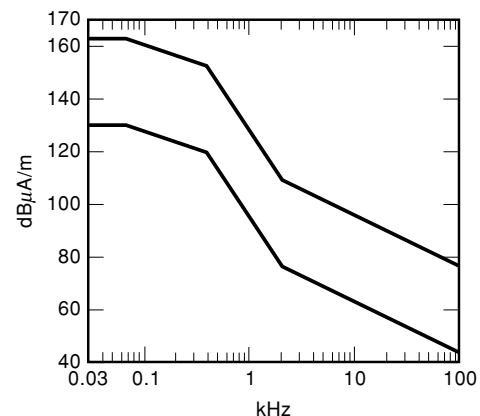


Figure 2. Magnetic field strength limits derived from US MIL-STD-461D RE101 (Navy only) (7). These limits are originally given in dB(pT) (decibels above 1 pT). The measurement procedure requires a 36 turn shielded loop antenna with a diameter of 13.3 cm. Measurement distance is 7 cm for the upper limit and 50 cm for the lower limit.

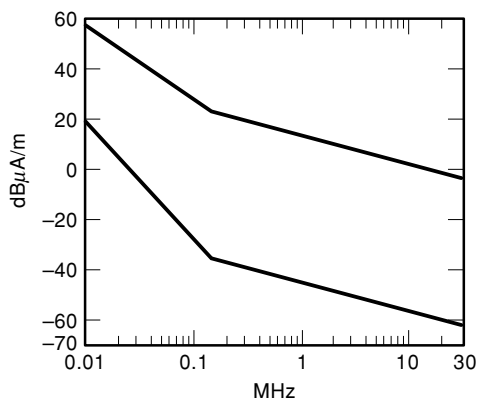


Figure 3. Narrowband emission limits of the magnetic field strength derived from the German military standard VG 95343 Part 22 (8). This standard gives the limits of $H \cdot Z_0$ in $\text{dB}(\mu\text{V}/\text{m})$ of four equipment classes, the emissions have to be measured with a loop antenna calibrated in $\text{dB}(\mu\text{V}/\text{m})$ in the near field of the equipment under test (EUT). Therefore, the limits have been converted into $\text{dB}(\mu\text{A}/\text{m})$. The lower limit is Class 1, the upper is Class 4.

measurement procedures have been standardized for industrial, scientific, medical (ISM) and other equipment (1,2,10–12). An example is given in Fig. 4.

Radiated emission limits of fluorescent lamps and luminaires are specified in a $\text{dB}(\mu\text{A})$ using a large-loop-antenna system (LAS) (10). For further information, see the text below.

EQUIPMENT FOR MAGNETIC FIELD MEASUREMENTS

Magnetic Field Sensors Other Than Loop Antennas

An excellent overview of magnetic field sensors other than loop antennas is given in Ref. 13. Table 1 lists the different types of

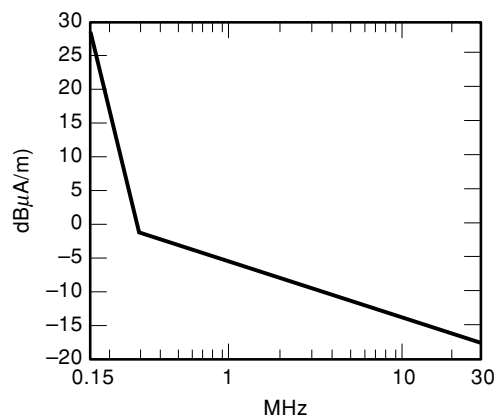


Figure 4. Radiated emission limits for navigational receivers according to draft revision IEC 945 (IEC 80/124/FDIS), originally given in $\text{dB}(\mu\text{V}/\text{m})$, for the purpose of this article converted into $\text{dB}(\mu\text{A}/\text{m})$.

field sensors which are exploiting different physical principles of operation.

Magnetic Field-Strength Meters With Loop Antennas

Especially for the measurement of radio wave propagation and radiated electromagnetic disturbance pick-up devices, the antennas become larger and therefore they are used separate from the indicating instrument (see Fig. 5). The instrument is a selective voltmeter, a measuring receiver or a spectrum analyzer. The sensitivity pattern of a loop antenna can be represented by the surface of two spheres (see Figs. 6 and 7). In order to determine the maximum field strength, the loop antenna has to be turned into the direction of maximum sensitivity.

To obtain an isotropic field sensor, three loops have to be combined in such a way that the three orthogonal components

Table 1. Overview over Different Magnetic Field Sensors, their Underlying Physical Effects, their Applicable Level, and Frequency Ranges from Ref. 73. For Easier Comparison with the Rest of the Text, the Values of Ref. 73 Have Been Converted from G into A/m

Type	Principle of Operation	Level of Operation	Frequency Range
Search-coil magnetometer	Faraday's law of induction	10^{-6} to 10^9 A/m	1 Hz to 1 MHz
Flux-gate magnetometer	Induction law with hysteresis of magnetic material	10^{-4} to 10^4 A/m	dc to 10 kHz
Optically pumped magnetometer	Zeeman effect: splitting of spectral lines of atoms	10^{-6} to 10^2 A/m	dc
Nuclear-precession magnetometer	Response of nuclei of atoms to a magnetic field	10^{-5} to 10^2 A/m	dc (upper frequency limited by gating frequency of hydrocarbon fluid)
SQUID magnetometer	Superconducting quantum interference device	10^{-8} to 10^{-2} A/m; speciality: differential field measurements	dc
Hall-effect sensor	Hall effect	10^{-1} to 10^5 A/m	dc to 1 MHz
Magnetoresistive magnetometer	Magnetoresistive effect	10^{-4} to 10^4 A/m	dc to 1 GHz
Magnetodiode	Semiconductor diode with undoped silicon	10^{-2} to 10^3 A/m	dc to 1 MHz
Magnetotransistor	Hall and Suhl effects	10^{-3} to 10^3 A/m	dc to 1 MHz
Fiberoptic magnetometer	Mach-Zehnder interferometer	10^{-7} to 10^3 A/m	dc to 60 kHz
Magneto-optical sensor	Faraday polarization effect	10^2 to 10^9 A/m	dc to 1 GHz

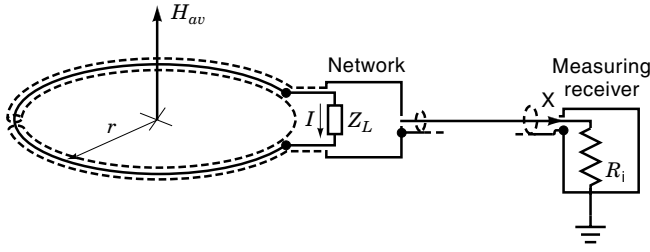


Figure 5. Magnetic field strength measuring loop. The network may consist of a passive or active circuit.

of the magnetic field H_x , H_y and H_z are combined to fulfill the equation

$$H = \sqrt{H_x^2 + H_y^2 + H_z^2}$$

Isotropic performance is however only a reality in broadband magnetic field sensors, where each component is detected with a square-law detector and combined subsequently. For the measurement and detection of radio signals isotropic antennas are not available. Hybrids may be used for limited frequency ranges to achieve an omnidirectional azimuthal (not isotropic) pickup.

Antenna-Factor Definition. The output voltage V of a loop antenna is proportional to the average magnetic field strength H perpendicular to the loop area. If the antenna output is connected to a measuring receiver or a spectrum analyzer, the set consisting of antenna and receiver forms a selective magnetometer.

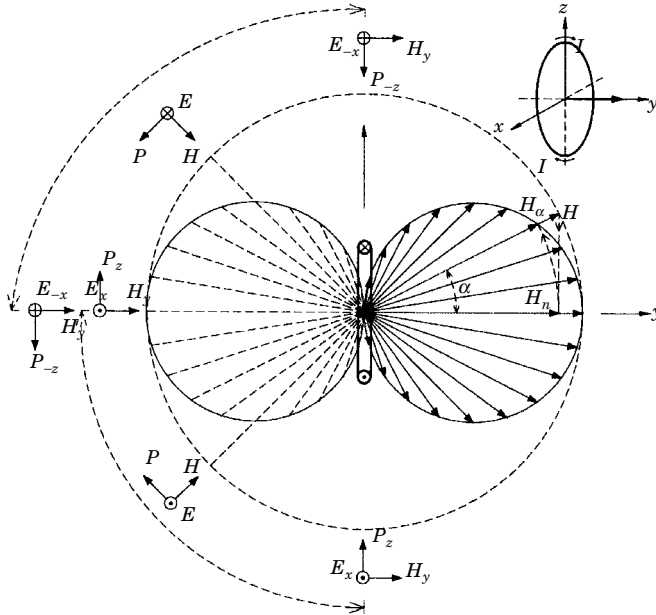


Figure 6. Cross section of a loop antenna sensitivity pattern. The arrow length H_α shows the indicated field strength at an angle α which is a fraction of the original field strength H , with $H_\alpha = H \cos \alpha$.

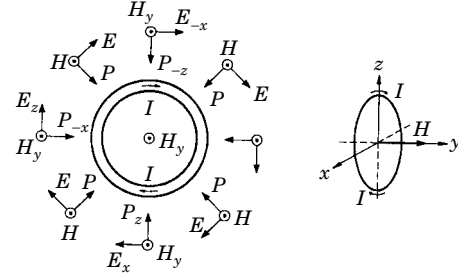


Figure 7. Direction of the field vectors (H , E , and P) under far-field conditions.

The proportionality constant is the antenna factor K_H for the average magnetic field strength H :

$$K_H = \frac{H}{V} \quad \text{in} \quad \frac{\text{A}}{\text{m}} \frac{1}{\text{V}} = \frac{1}{\Omega \text{m}} \quad (1a)$$

For the average magnetic flux density B the corresponding proportionality constant is

$$K_B = \frac{B}{V} = \frac{\mu_0 H}{V} = \mu_0 K_H \quad \text{in} \quad \frac{\text{Vs}}{\text{Am}} \frac{1}{\text{m}} \frac{1}{\text{V}} = \frac{\text{T}}{\text{m}^2 \text{V}} = \frac{\text{T}}{\text{V}} \quad (1b)$$

In the far field, where electric field and magnetic fields are related via the free-space wave impedance Z_0 , the loop antenna can be used to determine the electric field strength E . For this case the proportionality constant is:

$$K_E = \frac{E}{V} = \frac{Z_0 H}{V} = Z_0 K_H \quad \text{in} \quad \frac{\text{V}}{\text{Am}} \frac{1}{\text{m}} \frac{1}{\text{V}} = \frac{1}{\text{m}} \quad (1c)$$

In the area of radio wave propagation and radio disturbance measurement, quantities are expressed in logarithmic units. Therefore, the proportionality constants are converted into logarithmic values too:

$$k_H = 20 \log(K_H) \quad \text{in} \quad \text{dB} \left(\frac{1}{\Omega \text{m}} \right) \quad (2a)$$

$$k_B = 20 \log(K_B) \quad \text{in} \quad \text{dB} \left(\frac{\text{T}}{\text{V}} \right) \quad (2b)$$

$$k_E = 20 \log(K_E) \quad \text{in} \quad \text{dB} \left(\frac{1}{\text{m}} \right) \quad (2c)$$

By using logarithmic antenna factors, a field-strength level $20 \log(H)$ is obtained in $\text{dB}(\mu\text{A}/\text{m})$ from the measured output voltage level $20 \log(V)$ in $\text{dB}(\mu\text{V})$ by applying the equation: $20 \log(H) = 20 \log(V) + k_H$. The final section of this article describes a method to calibrate the antenna factors of circular loop antennas.

Concepts of Magnetic Field-Strength Meters. The loop antenna of a magnetic field-strength meter may be mounted on the measuring receiver or used as a separate unit, connected to the measuring receiver with a coaxial cable. CISPR 16-1, the basic standard for emission measurement instrumentation to commercial (i.e., nonmilitary) standards, requires a loop antenna in the frequency range of 9 kHz to 30 MHz which is completely enclosed by a square having sides 0.6 m in length. For

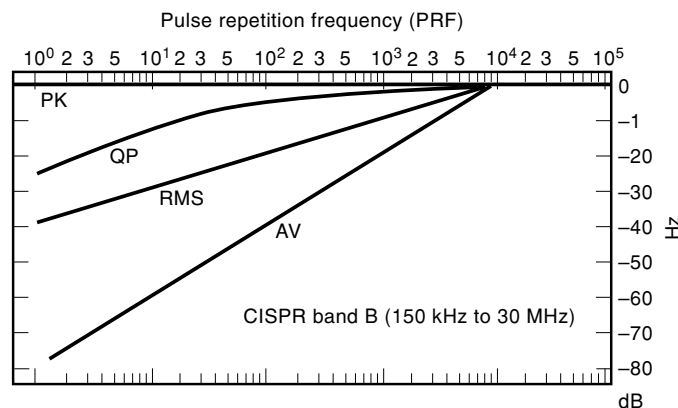


Figure 8. Detector response of a test receiver for impulsive interference as specified in Ref. 1.

protection against stray pick-up of electric fields, loop antennas employ a coaxial shielding structure. For optimum performance, the shielding structure may be arranged symmetrically in two half-circles around a circular loop with a slit between the two halves in order to avoid electric contact between the two shields.

For narrowband magnetic field measurements of radio disturbance, measuring receivers employ standardized bandwidths and weighting detectors in order to produce standardized measurement results for all types of perturbations including impulsive signals. For comparison with the emission limit, usually the quasi-peak (QP) detector is to be used.

To understand the function of a weighting curve in measuring receivers, the following interpretation is given. The test receiver has certain elements that determine a weighting curve (e.g. for the QP detector): the measurement bandwidth, the charge and discharge times of the detector circuit, and the time constant of the meter. When measured with a QP detector, for the frequency range given in Fig. 8, an impulsive signal with a constant impulse strength and a pulse repetition frequency of 100 Hz will cause a meter indication 10 dB above that of the indication when the pulse repetition frequency is 10 Hz. Or, to produce the same indication on the meter as a signal with 100-Hz repetition frequency, the level of the 10-Hz impulsive signal will have to be increased by an amount of 10 dB.

Earlier manually operated field-strength meters achieved high sensitivity by operating the loop at resonance (14). The sensitivity was raised by the amount of the Q -factor of the resonating circuit. One of the latest models which was used up to the 1980s, reached a sensitivity of -60 dB(μ A/m) with a measurement bandwidth of 200 Hz in the frequency range 100 kHz to 30 MHz (15).

For automated field-strength measurement systems, tuning of the loop circuit could no longer be afforded. A broadband active loop employs an output voltage proportional to the short-circuited loop current thus achieving a flat response of the antenna factor versus frequency (16).

A flat response of the system is also achieved using a current probe which measures the short-circuit current in the large-loop antenna system (LAS) described by Bergervoet and van Veen (9). It is essentially a magnetic-field-induced current measurement (see subsequent explanations). The highest sensitivity described in the literature for a wideband system

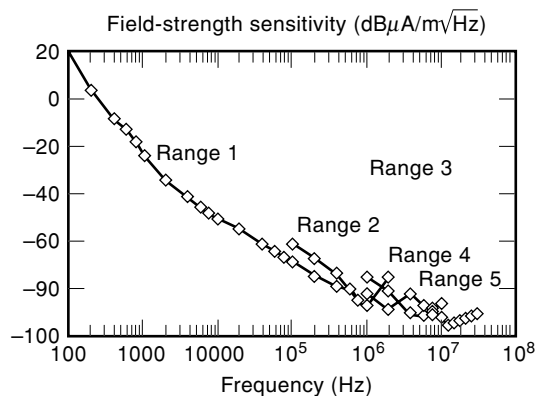


Figure 9. Sensitivity per hertz bandwidth of the active loop (16).

was achieved with a specially designed active loop. With additional frequency-dependant switching of elements (17) sensitivity is even better than that of manually operated field-strength meters with tuning of the loop circuit. Figure 9 shows the amplitude density of the minimum detectable magnetic field strength H_{Neq} in dB(μ A/m/ $\sqrt{\text{Hz}}$) equivalent to the internal electronic noise of the system consisting of antenna and measuring receiver.

MAGNETIC-FIELD-STRENGTH MEASUREMENT METHODS

Measurement of Magnetic Fields With Regard to Human Exposure to High em Fields

Usually, to measure magnetic fields with regard to human exposure to high fields, magnetic field-strength meters are using broadband detectors and apply an isotropic response. Modern concepts of low-frequency electric and magnetic field strength meters apply fast Fourier transform (FFT) for proper weighting of the total field with regard to frequency-dependent limits (18,19).

Use of Loop Antennas for Radio Wave Field-Strength Measurements Up to 30 MHz

ITU-R Recommendation PI.845-1 Annex 1 gives guidance to accurate measurement of radio wave field strengths. Rod antennas are the preferred receiving antennas since they provide omnidirectional azimuthal pickup. The positioning of vertical rod antennas is however important, since the result is very sensitive to field distortions by obstacles and sensitive to the effects of ground conductivity. It is a well-known fact that measurements with loop antennas are less sensitive to these effects and their calibration is not affected by ground conductivity apart from the fact that the polarization may deviate from horizontal if ground conductivity is poor. Therefore, many organizations use vertical monopoles for signal measurements but standardize results by means of calibration data involving comparisons for selected signals indicated by field-strength meters incorporating loop-receiving antennas. Accuracy requirements are given in Ref. 20, general information on equipment and methods of radio monitoring are given in Ref. 21.

Solutions to Problems With Ambients in Commercial EMI Standards. CISPR Class B radiated emission limits in the

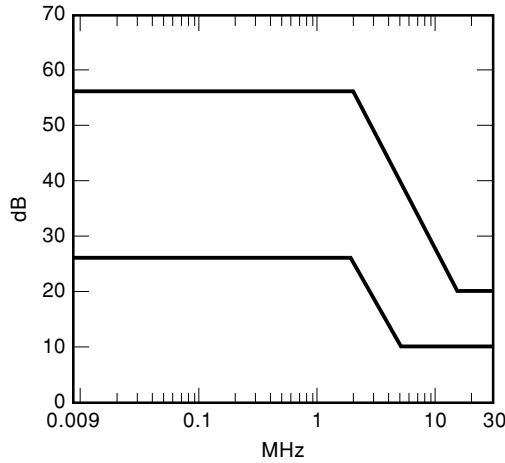


Figure 10. Conversion factors ΔH for the limit of the magnetic field strength from 30-m measurement distance to 10-m and 3-m measurement distances above a conducting ground plane according to Ref. 26. The upper curve is for 30 to 3 m, and the lower curve is for 30 to 10 m distances.

frequency range 9 kHz to 30 MHz have been at 34 dB(μ V/m) at a distance of 30 m from the EUT for a long time. Moreover, the test setup with EUT and vertical loop antenna required turning of both EUT and loop antenna in order to find the maximum emission. On most of the open area test sites the ambient noise level makes compliance testing almost impossible. This is due to the fact that ambient noise itself is near or above the emission limit. Two different approaches were proposed as a solution to that problem:

(1) To reduce the measurement distance from 30 m to 10 m or even 3 m. A German group proposed frequency-dependent conversion factors, justified by calculations and an extensive amount of measurements. The conversion factors are given in Fig. 10. In Fig. 10 the slopes between 1.8 MHz and 16 MHz show the transition region from near field, where H is inversely proportional with r^3 or $r^{2.6}$, to far field, where H is inversely proportional with r .

(2) To reduce the measurement distance to zero. A Dutch group proposed the large-loop antenna system mentioned previously (9). With this method the EUT is placed in the center of a loop antenna system, which consists of three mutually perpendicular large loop antennas (Fig. 11). The magnetic field emitted by the EUT induces currents in the large loop antennas. Since there are three orthogonal loops, there is no need to rotate either the EUT or the loop antenna system. The current induced in each loop is measured by means of a current probe, which is connected to a CISPR measuring receiver. Since the current is measured, emission limits are given in dB(μ A) instead of dB(μ A/m). Each loop antenna is constructed of a coaxial cable which contains two slits, positioned symmetrically with respect to the position of the current probe. Each slit is loaded by resistors in order to achieve a frequency response flat to within ± 2 dB in the frequency range from 9 kHz to 30 MHz (9,10). In order to verify and validate the function of each large loop, a specially designed folded dipole has been developed (9,10). It produces both a magnetic dipole moment m_H and an electric dipole moment m_E , when a signal is connected to the folded dipole. The folded dipole serves to test the large loop antenna for its sensitivity in 8 positions.

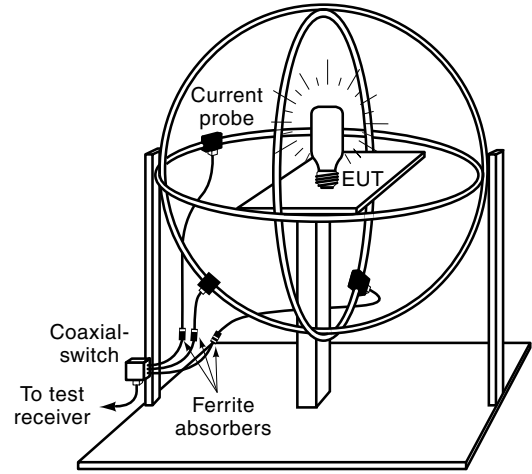


Figure 11. Simplified drawing of a large loop antenna system with position of the EUT.

Problems in the Near Field to Far Field Transition Zone. Problems with magnetic field strength measurements in the transition region between near field and far field are discussed in detail in Ref. 22. When a small magnetic dipole is located in the free space, the electromagnetic field in a point $P(r, \theta, \phi)$ is described by the following three relations (see Fig. 12):

$$H_r = \frac{jk}{2\pi} \frac{m_H \cos \theta}{r^2} \left(1 + \frac{1}{jkr}\right) e^{-jkr} \quad (3a)$$

$$H_\theta = \frac{-k^2}{4\pi} \frac{m_H \sin \theta}{r} \left(1 + \frac{1}{jkr} - \frac{1}{(kr)^2}\right) e^{-jkr} \quad (3b)$$

$$E_\phi = \frac{Z_0 k^2}{4\pi} \frac{m_H \sin \theta}{r} \left(1 + \frac{1}{jkr}\right) e^{-jkr} \quad (3c)$$

where $k = 2\pi/\lambda$, and $m_H = \pi R^2 I_0$ is the magnetic dipole moment, a vector perpendicular to the plane of the dipole. Equations (3a–3c) completely describe the electromagnetic field of the magnetic dipole.

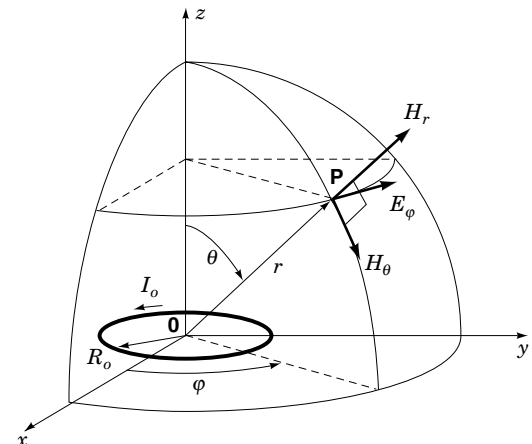


Figure 12. Field components H_r , H_θ , and E_ϕ in P at a distance r from the center of the magnetic dipole in the xy -plane.

Two situations are further discussed: (1) the near field, where r is much smaller than λ but larger than the maximum dimension of the source (i.e. $kr \ll 1$), and (2) the far field, where r is much larger than λ and much larger than the maximum dimension of the source (i.e. $kr \gg 1$).

For the near field case, where $kr \ll 1$ and using $e^{-jkr} = \cos(kr) - j\sin(kr)$, Eqs. (3a–3c) are simplified to

$$H_r = \frac{2m_H \cos \theta}{4\pi r^3} \quad (4a)$$

$$H_\theta = \frac{m_H \sin \theta}{4\pi r^3} \quad (4b)$$

$$E_\phi = \frac{kZ_0 m_H \sin \theta}{4\pi r^2} \quad (4c)$$

From Eqs. (4a–4c) we can see that H_r and H_θ are inversely proportional to r^3 , whereas E_ϕ is inversely proportional to r^2 .

For the far-field case where $kr \gg 1$, Eqs. (3a–3c) are reduced to

$$H_r = \frac{jk m_H \cos \theta}{2\pi r^2} e^{-jkr} \Rightarrow 0 \quad (5a)$$

$$H_\theta = \frac{-k^2 m_H \sin \theta}{4\pi r} e^{-jkr} \quad (5b)$$

$$E_\phi = \frac{k^2 Z_0 m_H \sin \theta}{4\pi r} e^{-ikr} \quad (5c)$$

From Eqs. (5a–5c) one can see that in the far field H_r vanishes in comparison to H_θ and that H_θ and E_ϕ are inversely proportional to r .

In the frequency range of 9 kHz to 30 MHz, where emission limits have been set, the corresponding wavelength is 33 km to 10 m. Since for compliance testing, ambient emissions on an open area test site require a reduction of the measurement distance to 10 m or even 3m, measurements are carried out in the near field zone over a wide frequency range. At the higher frequency range the transition zone and the beginning far field zone are reached. Goedbloed (22) investigated the transition zone and identified the critical condition where H_r and H_θ are equal in magnitude. It occurs where

$$\frac{2m_H}{4\pi r^3} \sqrt{1 + k^2 r^2} = \frac{m_H}{4\pi r^3} \sqrt{1 - k^2 r^2 + k^4 r^4} \quad (6)$$

or where

$$fr = 112.3 \text{ in MHz} \cdot \text{m} \quad (7)$$

For $r = 10$ m, $H_{\theta \max} > H_{r \max}$ at frequencies greater than 11 MHz.

The CISPR magnetic field measurement method is illustrated by Fig. 13, with the test setup on a metallic ground plane and the receiving antenna in the vertical plane. In Figs. 14 and 15, two different cases of radiating electrically small magnetic dipoles are illustrated: the first one with the dipole moment parallel to the ground plane and the second one with the dipole moment perpendicular to the ground plane. Because of the reflecting ground plane two sources are responsible for the field at the location of the receiving antenna: the original source and the mirror source. The points and crosses drawn in both

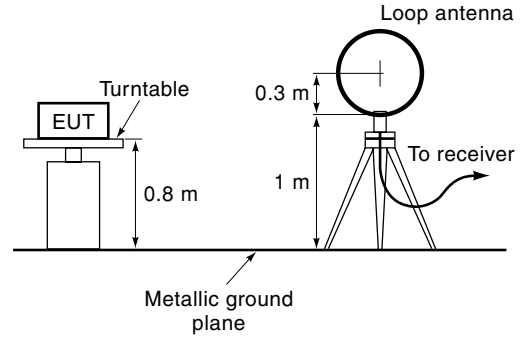


Figure 13. Basic CISPR setup for magnetic field measurements. Both EUT and loop antennas have to be turned round until the maximum indication on the receiver has been found.

sources show the direction of the current. In Fig. 14, the currents are equally oriented. In this case the loop antenna detects the radial component $H_{d,r}$ and the direct tangential component $H_{d,\theta} = 0$ since $\theta_d = 0$. Therefore, direct radiation will only contribute if $fd \ll 112 \text{ MHz} \cdot \text{m}$, see Eq. (7). In the case of $fd \gg 112 \text{ MHz} \cdot \text{m}$, the loop antenna will receive direct radiation if it is rotated by 90° . This may be observed frequently in practical measurements: at low frequencies the maximum radiation is found with the loop antenna in parallel to the EUT and at high frequencies with the loop antenna oriented perpendicular to the EUT. In addition to these direct components, the indirect radial and tangential components $H_{i,r}$ and $H_{i,\theta}$ are superpositioned in the loop antenna. Assuming near-field conditions it follows from Eqs. (4), that the magnitude of the magnetic field H_m is given by

$$H_m = H_{d,r} + H_{i,r} \cos \theta_i - H_{i,\theta} \sin \theta_i \quad (8)$$

$$= \frac{m_H}{4\pi d^3} \left(2 + \frac{d^3}{d_i^3} (2 \cos^2 \theta_i - \sin^2 \theta_i) \right)$$

where $d_i = \sqrt{(2h)^2 + d^2}$ is the distance between the mirror dipole and the loop antenna.

Goedbloed gives a numerical example with $m_H = 4\pi 10^3 \mu\text{Am}^2$ (e.g. 100 mA through a circular loop with a diameter of 0.40 m). Using Eq. (8) with $d = 3$ m and $h = 1.3$ m will give $H_m = 38.6 \text{ dB}(\mu\text{A/m})$ with the mirror source and $37.4 \text{ dB}(\mu\text{A/m})$

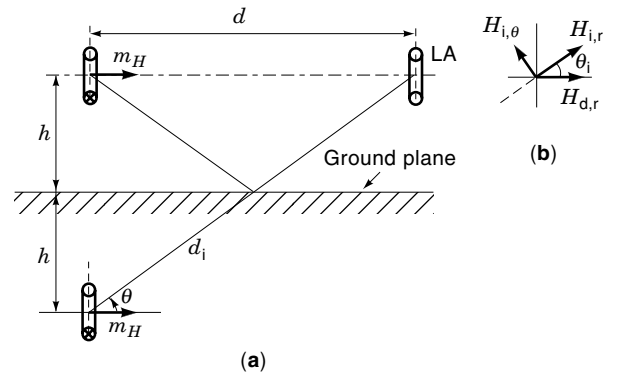


Figure 14. (a) Receiving conditions for a magnetic dipole with a horizontal dipole moment; (b) Vectors of the direct and indirect radiated H -field components.

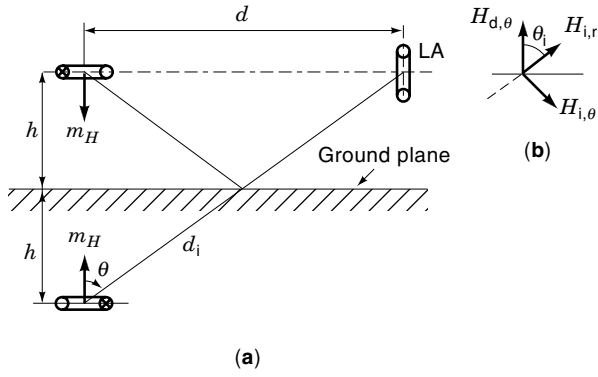


Figure 15. (a) Receiving conditions for a magnetic dipole with a vertical dipole moment, and the receiving loop-antenna in the vertical position as specified by the standard; (b) Vectors of the indirect radiated H -field components (no reception of direct radiation).

without the mirror source, which shows that in this case the reflecting ground plane has little influence. The influence of the ground plane is quite different in the case of a vertical dipole moment, i.e. a dipole moment perpendicular to the ground plane as illustrated in Fig. 15. In the case of Fig. 15 the loop antenna does not receive direct radiation at all, as $H_{d,r}(\theta_d = \pi/2) = 0$ and $H_{d,\theta}$ is parallel to the loop antenna. Hence, the received signal is completely determined by the radiation coming from the mirror source, which also means that the result is determined by the quality of the reflecting ground plane. With the reflecting ground plane $H_m = H_{i,r} \sin \theta_i + H_{i,\theta} \cos \theta_i = 27.2 \text{ dB}(\mu\text{A/m})$, whereas without the reflecting ground plane no field strength will be measured. If the loop antenna were positioned horizontally above the ground plane at $h = 1.3 \text{ m}$, $H_m = H_{d,\theta} + H_{i,r} \cos \theta_i - H_{i,\theta} \sin \theta_i = 32.4 \text{ dB}(\mu\text{A/m})$ and $H_m = 31.4 \text{ dB}(\mu\text{A/m})$ without the reflecting ground plane. Measurements in a shielded room would even be less predictable, since the result would be determined by mirror sources on each side including the ceiling of the shielded room. Absorbers are not very helpful in the low frequency ranges. From these results, Goedbloed concludes that in order to judge the interference capability of an EUT, the method proposed by Bergervoet and Van Veen (9), is an efficient method of magnetic field measurements.

CALIBRATION OF A CIRCULAR LOOP ANTENNA

A time-varying magnetic field at a defined area S can be determined with a calibrated circular loop. For narrow-band magnetic field measurements, a measuring loop consists of an output interface (point X on Fig. 5), which links the induced current to a measuring receiver. It may have a passive or an active network between loop terminals and output. The measuring loop can also include a shielding over the loop circumference against any perturbation of strong and unwanted electric fields. The shielding should be interrupted at a point on the loop circumference.

Generally in the far-field the streamlines of magnetic flux are uniform, but in the near-field, i.e. in the vicinity of the generator of a magnetic field, they depend on the source and its periphery. Figure 19 shows the streamlines of the electromagnetic vectors generated by the transmitting loop L1. In the near-field, the spatial distribution of the magnetic flux, $B = \mu_0 H$, over the

measuring loop area is not known. Only the normal components of the magnetic flux, averaged over the closed-loop area, can induce a current through the loop conductor.

The measuring loop must have a calibration (conversion) factor or set of factors, that, at each frequency, expresses the relationship between the field strength impinging on the loop and the indication of the measuring receiver. The calibration of a measuring loop requires the generation of a well-defined standard magnetic field on its effective receiving surface. Such a magnetic field is generated by a circular transmitting loop when a defined root mean square (rms) current is passed through its conductor. The unit of the generated or measured magnetic field H_{av} is A/m and therefore is also an rms value. The subscript, av, strictly indicates the average value of the spatial distribution, not the average over a period of T of a periodic function. This statement is important for near-field calibration and measuring purposes. For far-field measurements the result indicates the rms value of the magnitude of the uniform field. In the following we discuss the requirements for the near-zone calibration of a measuring loop.

CALCULATION OF STANDARD NEAR-ZONE MAGNETIC FIELDS

To generate a standard magnetic field, a transmitting loop L1 is positioned coaxial and plane-parallel at a separation distance d from the loop L2, like in Fig. 16. The analytical formula for the calculation of the average magnetic field strength H_{av} in A/m generated by a circular filamentary loop at an axial distance d including the retardation due to the finite propagation time was obtained earlier by Greene (23). The average value of field strength H_{av} was derived from the retarded vector potential A_φ as tangential component on the point P of the periphery of loop L2:

$$H_{av} = \frac{I r_1}{\pi r_2} \int_0^\pi \frac{e^{-j\beta R(\varphi)}}{R(\varphi)} \cos(\varphi) d\varphi \quad (9a)$$

$$R(\varphi) = \sqrt{d^2 + r_1^2 + r_2^2 - 2r_1 r_2 \cos(\varphi)} \quad (9b)$$

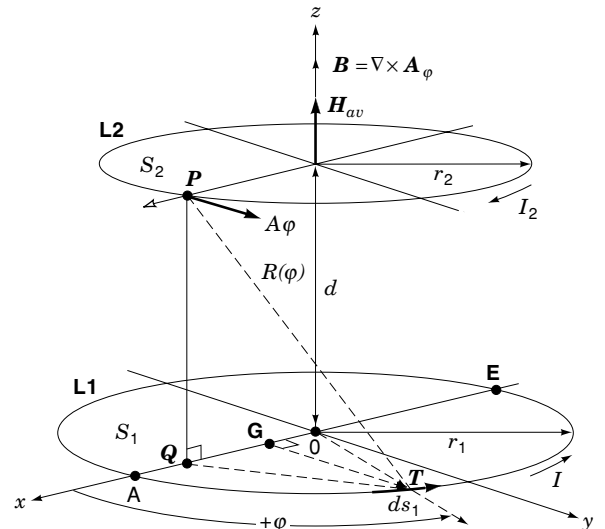


Figure 16. Configuration of two circular loops.

In these equations for the thin circular loop, I is transmitting loop rms current in A, d is distance between the planes of the two coaxial loop antennas in m, r_1 and r_2 are filamentary loop radii of transmitting and receiving loops in m, respectively, β is wavelength constant, $\beta = 2\pi/\lambda$, and λ is wavelength in m.

Equations (9a) and (9b) can be determined by numerical integration. To this end we separate the real and imaginary parts of the integrand using Euler's formula $e^{-j\phi} = \cos(\phi) - j \sin(\phi)$ and rewrite Eq. (9a) as

$$H_{av} = \frac{I r_1}{\pi r_2} (F - jG) \quad (10a)$$

where

$$F = \int_0^\pi \frac{\cos[\beta R(\varphi)]}{R(\varphi)} \cos(\varphi) d\varphi \quad (10b)$$

$$G = \int_0^\pi \frac{\sin[\beta R(\varphi)]}{R(\varphi)} \cos(\varphi) d\varphi \quad (10c)$$

and the magnitude of H_{av} is then obtained as

$$|H_{av}| = \frac{I r_1}{\pi r_2} \sqrt{F^2 + G^2} \quad (10d)$$

It is possible to evaluate the integrals in Eqs. (10) by numerical integration with an appropriate mathematics software on a personal computer. Some mathematics software can directly calculate the complex integral of Eqs. (9).

ELECTRICAL PROPERTIES OF CIRCULAR LOOPS

Current Distribution Around a Loop

The current distribution around the transmitting loop is not constant in amplitude and in phase. A standing wave of current exists on the circumference of the loop. This current distribution along the loop circumference is discussed by Greene on pp. 323–324 (23). He has assumed the loop circumference $2\pi r_1$ being electrically smaller than the wave length λ and the loop current being constant in phase around the loop and the loop being sufficiently loss-free. The single-turn thin loop was considered as a circular balanced transmission line fed at points A and D and short-circuited at the points E and F (Fig. 17).

In an actual calibration setup the loop current I_1 is specified at the terminals A and D. The average current was given as a function of input current I_1 of the loop (24):

$$I_{av} = I_1 \frac{\tan(\beta \pi r_1)}{\beta \pi r_1} \quad (11)$$

The fraction of I_{av}/I_1 from Eq. (11) expressed in dB gives the conditions for determining of the highest frequency f and the radius of the loop r_1 . The deviation of this fraction is plotted in Fig. 18.

The current I in Eqs. (9) must be substituted with I_{av} from Eq. (11). Since Eq. (11) is an approximate expression, it is recommended to keep the radius of the transmitting loop small enough for the highest frequency of calibration to minimize the errors. For the dimensioning of the radius of the receiving loop these conditions are not very important, until the receiving loop is calibrated with an accurately defined standard magnetic

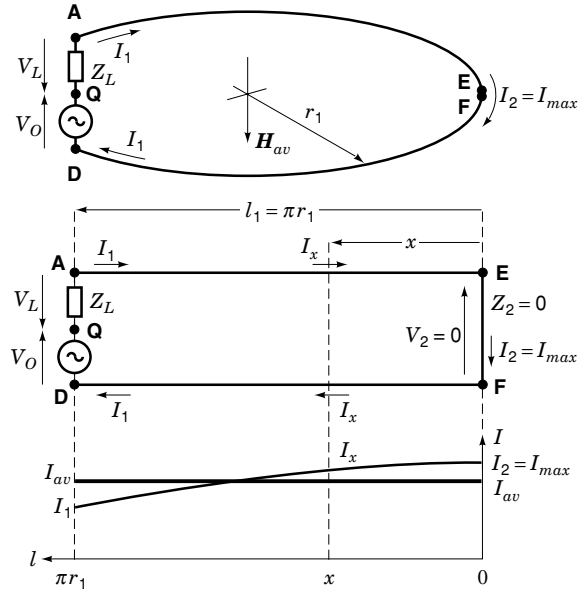


Figure 17. Current distribution on a circular loop.

field, but the resonance of the loop at higher frequencies must be taken into account.

Circular Loops With Finite Conductor Radii

A measuring loop can be constructed with one or more winding. The form of the loop is chosen as a circle, because of the simplicity of the theoretical calculation and calibration. The loop conductor has a finite radius. At high frequencies the loop current flows on the conductor surface and it shows the same proximity effect as two parallel, infinitely long cylindrical conductors. Figure 19 shows the cross-section of two loops intentionally in exaggerated dimensions. The streamlines of the electric field are orthogonal to the conductor surface of the transmitting loop L1 and they intersect at points A and A'. The total conductor current is assumed to flow through an equivalent thin filamentary loop with the radius $a_1 = \sqrt{r_1^2 - c_1^2}$, where $a_1 = \overline{OA} = \overline{OP} = \sqrt{\overline{OQ}^2 - \overline{QP}^2}$. The streamlines of the magnetic field are orthogonal to the streamlines of electric field. The receiving loop L2 with the finite conductor radius c_2 can encircle a part of magnetic field with its effective circular radius $b_2 = r_2 - c_2$.

The sum of the normal component of vectors H acting on the effective receiving area $S_2 = \pi b_2^2$ induces a current in the conductor of the receiving loop L2. This current flows through

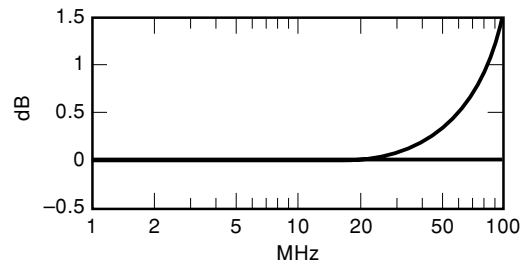


Figure 18. Deviation of I_{av}/I_1 for a loop radius, 0.1 m as $20 \log(I_{av}/I_1)$ in dB versus frequency.

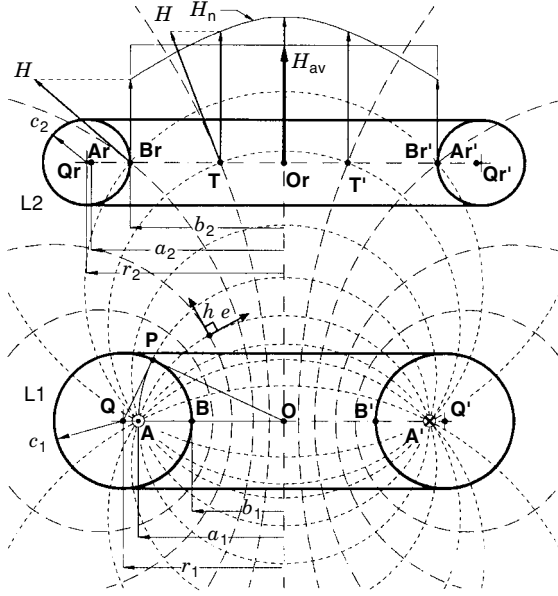


Figure 19. Filamentary loops of two loops with finite conductor radii and orthogonal streamlines of the electromagnetic vectors, produced from transmitting loop L1.

the filamentary loop with the radius a_2 . The average magnetic field vector H_{av} is defined as the integral of vectors H_n over effective receiving area S_2 , divided by S_2 . The magnetic streamlines, which flow through the conductor and outside of loop L2, cannot induce a current through the conductor along the filamentary loop Ar, Ar', of L2. The equivalent filamentary loop radii a_1 , a_2 and effective circular surface radii b_1 , b_2 can directly be seen from Fig. 19.

The equivalent thin current filament radius a_1 of the transmitting loop L1:

$$a_1 = \sqrt{r_1^2 - c_1^2} \quad (12a)$$

The equivalent thin current filament radius a_2 of the receiving loop L2:

$$a_2 = \sqrt{r_2^2 - c_2^2} \quad (12b)$$

The radius b_1 of the effective receiving circular area of the loop transmitting L1:

$$b_1 = r_1 - c_1 \quad (12c)$$

The radius b_2 of the effective receiving circular area of the receiving loop L2:

$$b_2 = r_2 - c_2 \quad (12d)$$

Impedance of a Circular Loop

The impedance of a loop can be defined at chosen terminals Q, D, as $Z = V/I_1$ (Fig. 17). Using Maxwell's equation with the Faraday's law $\text{curl} \mathbf{E} = -j\omega\Phi_m$ we can write the line integrals of the electric intensity E along the loop conductor through its cross section, along the path joining points D, Q, and the load

impedance Z_L between the terminals Q, A:

$$\int_{(AEFD)} E_s ds + \int_{(DQ)} E_s ds + \int_{(QA)} E_s ds = -j\omega\Phi_m \quad (13a)$$

Here, Φ_m is the magnetic flux. The impressed emf V acting along the path joining points D and Q is equal and opposite to the second term of Eq. (13a):

$$V = - \int_{(DQ)} E_s ds \quad (13b)$$

The impedance of the loop at the terminals D, Q can be written from Eqs. (13) dividing with I_1 as

$$Z = \frac{V}{I_1} = \frac{\int_{(AEFD)} E_s ds}{I_1} + \frac{\int_{(QA)} E_s ds}{I_1} + \frac{j\omega\Phi_m}{I_1} = Z_i + Z_L + Z_e \quad (14)$$

Z_i indicates the internal impedance of the loop conductor. Because of the skin effect, the internal impedance at high frequencies is not resistive. For the calculation of the Z_i we refer to Schelkunoff, p. 262 (25). Z_L is a known load or a source impedance on Fig. 17. Z_e is the external impedance of the loop:

$$Z_e = j\omega \frac{\Phi_m}{I_1} = j\omega \frac{\mu_0 H_{av} S}{I_1} \quad (15a)$$

We can consider that the loop consists of two coaxial and coplanar filamentary loops (i.e. separation distance $d = 0$). The radii a_1 and b_1 are defined in Eqs. (12). The average current I_{av} flows through the filamentary loop with the radius a_1 and generates an average magnetic field strength H_{av} on the effective circular surface $S_1 = \pi b_1^2$ of the filamentary loop with the radius b_1 . From the Eqs. (9) and (11) we can rewrite Eq. (15a), for the loop L1:

$$Z_e = j \frac{\tan(\beta\pi a_1)}{\beta\pi a_1} \mu_0 \omega a_1 b_1 \int_0^\pi \frac{e^{-j\beta R_0(\varphi)}}{R_0(\varphi)} \cos(\varphi) d\varphi \quad (15b)$$

$$R_0(\varphi) = \sqrt{a_1^2 + b_1^2 - 2a_1 b_1 \cos(\varphi)} \quad (15c)$$

The real and imaginary parts of Z_e are the radiation resistance and the external inductance of loops, respectively:

$$\text{Re}(Z_e) = \frac{\tan(\beta\pi a_1)}{\beta\pi a_1} \mu_0 \omega a_1 b_1 \int_0^\pi \frac{\sin(\beta R_0(\varphi))}{R_0(\varphi)} \cos(\varphi) d\varphi \quad (15d)$$

$$\text{Im}(Z_e) = \frac{\tan(\beta\pi a_1)}{\beta\pi a_1} \mu_0 \omega a_1 b_1 \int_0^\pi \frac{\cos(\beta R_0(\varphi))}{R_0(\varphi)} \cos(\varphi) d\varphi \quad (15e)$$

From Eq. (15e) we obtain the external self inductance:

$$L_e = \frac{\tan(\beta\pi a_1)}{\beta\pi a_1} \mu_0 a_1 b_1 \int_0^\pi \frac{\cos(\beta R_0(\varphi))}{R_0(\varphi)} \cos(\varphi) d\varphi \quad (15f)$$

Equations 15 include the effect of current distribution on the loop with finite conductor radii.

Mutual Impedance Between Two Circular Loops

The mutual impedance Z_{12} between two loops is defined as

$$Z_{12} = \frac{V_2}{I_1} = \frac{Z_2 I_2}{I_1} \quad (16)$$

The impedance of Z_2 in Eq. (16) can be defined like Eq. (14):

$$Z_2 = \frac{V_2}{I_2} = Z_{2i} + Z_L + Z_{2e} \quad (17)$$

here Z_{2i} is the internal impedance, Z_L is the load impedance, and Z_{2e} is the external impedance of the second loop L2.

The current ratio I_2 to I_1 in Eq. (16) can be calculated from Eqs. (9), (11), and (12). The current I_1 of the transmit loop with separation distance d :

$$I_1 = \frac{H_{av} \pi b_2}{\frac{\tan(\beta \pi a_1)}{\beta \pi a_1} a_1 \int_0^\pi \frac{e^{-j\beta R_d(\varphi)}}{R_d(\varphi)} \cos(\varphi) d\varphi} \quad (18a)$$

$$R_d(\varphi) = \sqrt{d^2 + a_1^2 + b_2^2 - 2a_1 b_2 \cos(\varphi)} \quad (18b)$$

and the current I_2 of the receive loop for the same H_{av} (here $d = 0$) is

$$I_2 = \frac{H_{av} \pi b_2}{\frac{\tan(\beta \pi a_2)}{\beta \pi a_2} a_2 \int_0^\pi \frac{e^{-j\beta R_0(\varphi)}}{R_0(\varphi)} \cos(\varphi) d\varphi} \quad (18c)$$

$$R_0(\varphi) = \sqrt{a_2^2 + b_2^2 - 2a_2 b_2 \cos(\varphi)} \quad (18d)$$

The general mutual impedance between two loops from Eqs. (16) and (17) is

$$Z_{12} = (Z_{2i} + Z_L + Z_{2e}) \frac{I_2}{I_1} = Z_{12i} + Z_{12L} + Z_{12e} \quad (19a)$$

where Z_{12i} is the mutual internal impedance, Z_{12L} denotes the mutual load impedance, and Z_{12e} is the external mutual impedance.

Arranging Eq. (15b) for Z_{2e} and the current ratio I_2/I_1 from Eqs. (18) the external mutual impedance yield

$$Z_{12e} = j \frac{\tan(\beta \pi a_1)}{\beta \pi a_1} \mu_0 \omega a_1 b_2 \int_0^\pi \frac{e^{-j\beta R_d(\varphi)}}{R_d(\varphi)} \cos(\varphi) d\varphi \quad (19b)$$

The real part of Z_{12e} may be described as mutual radiation resistance between two loops.

The imaginary part of Z_{12e} divided by ω gives the mutual inductance

$$M_{12e} = \frac{\tan(\beta \pi a_1)}{\beta \pi a_1} \mu_0 a_1 b_2 \int_0^\pi \frac{\cos(\beta R_d(\varphi))}{R_d(\varphi)} \cos(\varphi) d\varphi \quad (19c)$$

Equations (19b) and (19c) include the effect of current distribution on the loop with finite conductor radii.

DETERMINATION OF THE ANTENNA FACTOR

The antenna factor K is defined as a proportionality constant with necessary conversion of units. K is the ratio of the average magnetic field strength bounded by the loop to the measured output voltage V_L on the input impedance R_L of the measuring receiver. For the evaluation of the antenna factor there are two methods. The first is by calculation of the loop impedances, and the second is with the well-defined standard magnetic field calibration, which will also be needed for the verification of calculated antenna factors (24).

Determination of the Antenna Factor by Computing from the Loop Impedances

If a measurement loop (e.g. L2) has a simple geometric shape and a simple connection to a voltage measuring device with a known load R_L , we can determine the antenna factor by calculation. In the case of unloaded loop from Fig. 17 the open circuit voltage is

$$V_0 = j\omega\mu_0 H_{av} S_2 \quad (20a)$$

For the case of loaded loop the current is

$$I = \frac{V_0}{Z} = \frac{V_0}{R_L + Z_i + Z_e} \quad (20b)$$

The antenna factor from Eq. (9a) can be written with $V_L = Z_L I$ and Eqs. (20) as

$$K_H = \left| \frac{1}{j\omega\mu_0 S_2} \left(1 + \frac{Z_e}{R_L} + \frac{Z_i}{R_L} \right) \right| \quad \text{in} \quad \frac{\text{A}}{\text{m}} \frac{1}{\text{V}} \quad (21)$$

The effective loop area is $S_2 = \pi b_2^2$. The external loop impedance Z_e can be calculated with Eqs. (15). The internal impedance Z_i can be evaluated from Ref. 25.

Standard Magnetic Field Method

In the calibration setup in Fig. 20 we measure the voltages with standard laboratory measuring instrumentation with the 50 Ω impedance. The device to be calibrated consists at least of a loop and a cable with an output connector. Such a measuring loop can also include a passive or active network between the terminals C, D, and a coaxial shield on the circular loop conductor against unwanted electric fields, depending on its development and construction. The impedance Z_L at the terminals C, D is not accurately measurable. Such a complex loop must be calibrated with the standard magnetic field method. The antenna factor in Eqs. (1) can be defined through measuring of the voltage V_L and the uncertainties between loop terminals C, D and measuring receiver are fully calibrated. The attenuation ratio α of the voltages V_2 and V_L can be measured for each frequency:

$$\alpha = \frac{V_2}{V_L} \quad (22)$$

By using the Eqs. (22), (1), (11), and (12), with $V_2 = -I_1 R_2$, and $V_0 = \text{constant}$, Eq. (9a) can be rewritten:

$$K_H = \left| \alpha \frac{1}{R_2} \frac{\tan(\beta \pi a_1)}{\beta \pi a_1} \frac{a_1}{\pi b_2} \int_0^\pi \frac{e^{-j\beta R_d(\varphi)}}{R_d(\varphi)} \cos(\varphi) d\varphi \right| \quad (23)$$

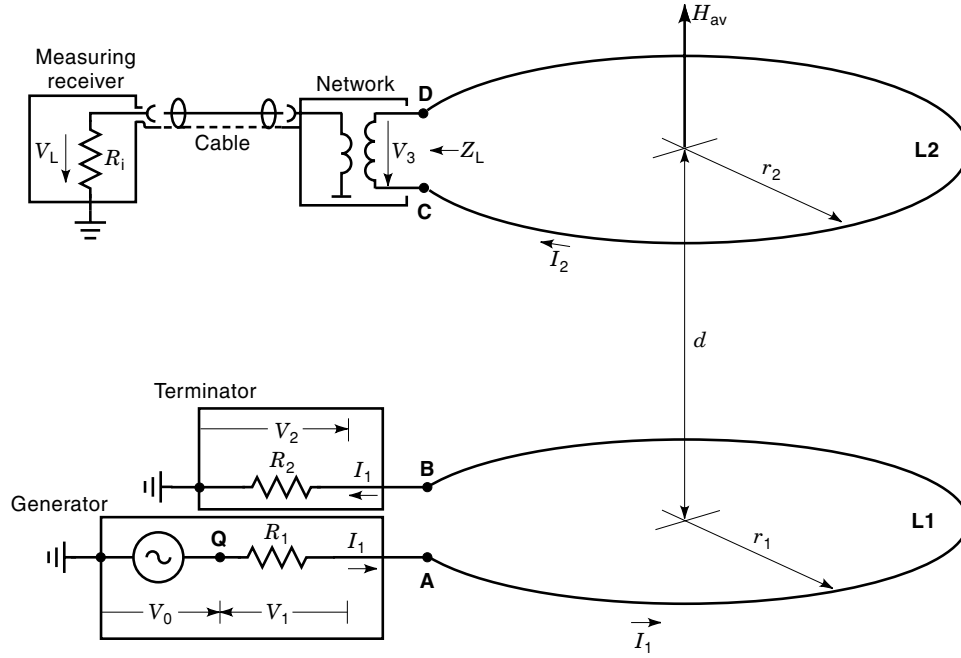


Figure 20. Calibration setup for circular loop antennas.

R_d is defined by Eq. (18b). Equation (23) can also be expressed logarithmically

$$k_H = 20 \log(K_H) \quad \text{in dB} \left(\frac{\text{A}}{\text{m}} \frac{1}{\text{V}} \right)$$

Equation (23) reduces the calibration of the loop to an accurate measurement of attenuation α for each frequency. The other terms of Eq. (23) can be calculated depending on the geometrical configuration of the calibration setup at the working frequency band of the measuring loop. The calibration uncertainties are also calculable with the given expressions. The uncertainty of the separation distance d between two loops must be taken into consideration as well. At a separation distance $d < r_1$ the change of the magnetic field is high.

For a calibration setup the separation distance d can be defined as small as possible. However, the effect of the mutual impedance must be taken into account in the calibration process and a condition to define the separation distance d must be given (Fig. 20). If the second loop is open circuited, that is the current $I_2 = 0$, the current I_1 is defined only from the impedances of the transmitting loop. In the case of a short-circuited second loop, I_2 is maximum and the value of I_1 will change depending on the supply circuit and loading of the transmitting loop. A current ratio q between these two cases can be defined as the condition of the separation distance d between the two loops.

It is assumed that the generator voltage V_0 is constant. The measuring loop L2 is terminated by Z_L . For $Z_L = 0$ and $V_{CD} = 0$, one obtains the current I_1 in the transmitting loop as

$$I_{1(Z_L=0)} = \frac{V_0}{R_1 + R_2 + Z_{AB} - \frac{Z_{12}^2}{Z_{CD}}} \quad (24a)$$

and for $Z_L = \infty$, i.e. $I_2 = 0$

$$I_{1(Z_L=\infty)} = \frac{V_0}{R_1 + R_2 + Z_{AB}} \quad (24b)$$

The ratio of Eq. (24a) to Eq. (24b) is

$$q \equiv \frac{I_{1(Z_L=0)}}{I_{1(Z_L=\infty)}} = \left| \frac{R_1 + R_2 + Z_{AB}}{R_1 + R_2 + Z_{AB} \left(1 - \frac{Z_{12}^2}{Z_{AB} Z_{CD}} \right)} \right| \quad (25a)$$

and here with the coupling factor $k = Z_{12} / \sqrt{Z_{AB} Z_{CD}}$ between two loops:

$$q = \left| \frac{R_1 + R_2 + Z_{AB}}{R_1 + R_2 + Z_{AB}(1 - k^2)} \right| \quad (25b)$$

where $R_1 = R_2 = 50 \Omega$, Z_{AB} , Z_{CD} , and Z_{12} can be calculated from Eqs. (15) and (19). For greater accuracy one must try to keep the ratio q close to unity (e.g., $q = 1.001$).

The influence of the loading of the second loop on the transmitting loop can also be found experimentally. The change of the voltage V_2 at R_2 in Fig. 20 must be considerably small, e.g. < 0.05 dB, while putting a short-circuited measuring loop at the chosen separation distance.

With the determining of K_H or k_H the loop can completely be calibrated up to its 50Ω output. A network analyzer is usually used for the attenuation measurement instead of a discrete measurement at each frequency with signal generator and measuring receiver. A network analyzer can normalize the frequency characteristic of the transmit loop and gives a quick overview on measured attenuation for the frequency band.

BIBLIOGRAPHY

1. CISPR 16 Specification for radio disturbance and immunity measuring apparatus and methods—Part 1: Radio disturbance and immunity measuring apparatus (8.1993); Part 2: Methods of measurement of disturbances and immunity (11.1996).
2. CISPR 11/2nd edition 1990-09 and EN 55011:07.1992: Limits and methods of measurement of electromagnetic disturbance characteristics of industrial, scientific, and medical (ISM) radio-frequency equipment.
3. *IRPA Guidelines on Protection against Non-Ionizing Radiation*, Oxford, UK: Pergamon Press, Inc., 1991.
4. ENV 50166 Part 1:1995—Human Exposure to electromagnetic fields—Low-frequency (0 Hz to 10 kHz) and Part 2:1995—Human exposure to electromagnetic fields—High frequency (10 kHz to 300 GHz).
5. VDE 0848 Part 4 A2:Draft 1992—Safety in electromagnetic fields. Limits for the protection of persons in the frequency range from 0 to 30 kHz and Part 2: Draft 1991—Safety in electromagnetic fields. Protection of persons in the frequency range from 30 kHz to 300 GHz.
6. IEEE standard C95.1-1991: IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz.
7. MIL-STD-461D, 11 January 1993: Requirements for the control of electromagnetic interference emissions and susceptibility, MIL-STD-462D, 11 January 1993: Measurement of electromagnetic interference characteristics, DOD, U.S.A.
8. VG 95373 Part 22, Cologne, Germany: Beuth Verlag, 1990.
9. J. R. Bergervoet and H. van Veen, A large loop antenna for magnetic field measurements, *Proc. Int. Symp. EMC*, 29–34, Zürich, 1989.
10. CISPR 15/5th edition 1996-03 and EN 55015:12.1993: Limits and methods of measurement of radio disturbance characteristics of electrical lighting and similar equipment.
11. Draft revision of IEC 945 (IEC 80/124/FDIS): Maritime navigation and radiocommunication equipment and systems—General requirements, methods of testing and required test results; identical requirements are given in Draft prETS 300 828/02.1997: EMC for radiotelephone transmitters and receivers for the maritime mobile service operating in the VHF bands, and Draft prETS 300 829/02.1997: EMC for Maritime mobile earth stations (MMES) operating in the 1.5/1.6 GHz bands; providing Low Bit Rate Data Communication (LBRDC) for the global distress and safety system (GMDSS).
12. U.S. FCC *Code of Federal Regulations* (CFR) 47 Part 18. Edition October 1, 1996.
13. J. E. Lenz, A review of magnetic sensors, *Proc. IEEE*, **78** (6): 973–989, 1990.
14. L. Rohde and F. Spies, Direkt zeigende Feldstärkemesser (Direct indicating field-strength meters), *Zeitschrift für technische Physik*, 10. Jahrg., Heft 11, 1938, pp. 439–444.
15. Data sheet edition 9.72 of Rohde & Schwarz Field-strength Meter HFH (0.1 to 30 MHz).
16. K. Danzeisen, Patentschrift DE 27 48 076 C2, 26.10.1977, Rohde & Schwarz GmbH & Co. KG, POB 801469, D-81614 München.
17. F. Demmel and A. Klein, Messung magnetischer Felder mit extrem hoher Dynamik im Bereich 100 Hz bis 30 MHz (Measurement of magnetic fields with an extremely high dynamic range in the frequency range 100 Hz to 30 MHz), *Proc. EMV '94*, Karlsruhe, 815–824, 1994.
18. CLC/TC111(Sec)61: Sept. 1995: Definitions and Methods of Measurement of Low Frequency Magnetic and Electric Fields with Particular Regard to Exposure of Human Beings (Draft 2: August 1995).
19. DKE 764/35-94: Entwurf DIN VDE 0848 Teil 1 “Sicherheit in elektrischen, magnetischen und elektromagnetischen Feldern; Meß- und Berechnungsverfahren” (Draft DIN VDE 0848 part 1 “Safety in electric, magnetic and electromagnetic fields; measurement and calculation methods”).
20. Recommendation ITU-R SM 378-5, *Field-strength measurements at monitoring stations*, SM Series Volume, ITU, Geneva 1994.
21. *Spectrum Monitoring Handbook*, ITU-R, Geneva 1995.
22. J. J. Goedbloed, “Magnetic field measurements in the frequency range 9 kHz to 30 MHz” EMC91, ERA Conference, Heathrow, UK: Feb. 1991.
23. F. M. Greene, The near-zone magnetic field of a small circular-loop antenna, *J. Res. Nat. Bur. Stand., Eng. and Inst.*, **71C** (4): 319–326, 1967.
24. A. Aykan, “Calibration of Circular Loop Antennas” to be published.
25. S. A. Schelkunoff, *Electromagnetic Waves*, New York: Van Nostrand, 1943.
26. J. Kaiser et al., Feldstärkeumrechnung von 30 m auf kürzere Messentfernungen (Conversion of field strength from 30 m to shorter distances), **110**: 820–825, 1989.

AYDIN AYKAN
MANFRED STECHER
Rohde & Schwarz GmbH & Co.KG

ATTENUATION MEASUREMENT

When a signal is sent along any transmission path, many different mechanisms degrade it. Because of finite conductivity, every cable shows a resistive loss. Furthermore, the dielectric loss and the skin effect may be significant at higher frequencies, and imperfect screening of cables leads to radiation losses which might be quite important at higher frequencies. Single-mode optical fibers have two main intrinsic loss mechanisms, scattering loss and absorption loss. Scattering loss is dominated by Rayleigh scattering, and below 1600 nm absorption losses are caused mainly by OH absorption. Fibers that are bent too tightly or cabled too poorly may have, in addition, bend losses because of nonguided modes. Connectors show losses because of nonideal contacts and imperfect impedance matching, which reflect part of the signal to the transmitter. Filters built in the ‘transmission path have losses in the passband caused by finite conductivity and probably dielectric losses. Wireless transmission paths, such as microwave links, satellite links, broadcast or mobile communication, are affected by scattering caused by rain, clouds, and multiple reflections.

The two examples following clearly show how additional attenuation or losses influence a system:

In a radar system a total loss of 2 dB in the feeder system and the duplexer wastes 37% of the transmitter power. During the development of the different parts, the losses have to be measured very carefully to minimize the total loss.

Satellite systems quite often operate with cooled front ends at the receiver because the signals to be picked up are extremely weak. Therefore the front ends often operate with noise temperatures of 5 K. An additional loss of 0.1 dB from an uncooled waveguide would raise the noise temperature to 7 K.

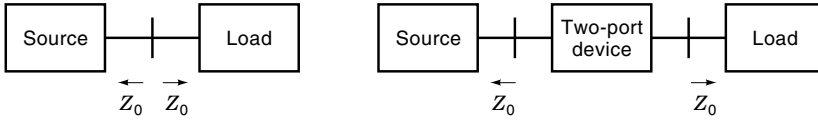


Figure 1. Attenuation measurement in a matched system.

These examples show how important it is to measure the losses of the different parts of a transmission system as accurately as possible to optimize system parameters.

Many different methods and a variety of systems for measuring attenuation have been developed. The most important techniques are described in the following sections.

ATTENUATION

Definition

In the field of loss measurement, the most important terms are attenuation, insertion loss, mismatch loss and voltage loss (1,2). These terms are discussed in the following sections.

Attenuation. According to R. W. Beatty (3), attenuation is defined as the decrease in power level at the load caused by inserting a device between a Z_0 source and load, where Z_0 is the characteristic impedance of the line. Figure 1 shows the basic idea of such an attenuation measurement.

Attenuation is mostly expressed by a logarithmic scale in decibels (dB) or in nepers. The attenuation of a two-port device is defined as follows (4):

in decibels,

$$A = 10 \log \left[\frac{\text{power delivered to a matched load by a matched source}}{\text{power delivered to the same load when the two-port device is inserted}} \right] \quad (1)$$

in nepers,

$$A = \frac{1}{2} \ln \left[\frac{\text{power delivered to a matched load by a matched source}}{\text{power delivered to the same load when the two-port device is inserted}} \right] \quad (2)$$

Because $\log(x) = \ln(x)/\ln(10)$, the following relationship between decibels and nepers is valid:

$$\text{Attenuation in decibels} = 8.6858 \times \text{attenuation in nepers}$$

Attenuation is a property only of a two-port device.

Insertion Loss. In practical applications neither the source nor the load has an impedance exactly equal to Z_0 the char-

acteristic impedance of a line. Therefore source and load have a reflection coefficient of r_s and r_L , respectively. Let P_1 be the power delivered from the source to the load and P_2 be the power absorbed by the same load when the two-port device is inserted (Fig. 2).

Then the loss is defined by

$$L_1 = 10 \log \frac{P_1}{P_2} \quad (3)$$

The insertion loss depends on the property of the device and the reflection coefficients of the source and the load.

Scattering Parameters. Two-port networks, especially at radio frequencies, are very well characterized by scattering parameters.

A two-port device inserted between a source and a load is shown in Fig. 3.

The complex wave amplitudes a_1 , a_2 , b_1 , and b_2 shown in Fig. 3 are related as follows:

$$b_1 = s_{11}a_1 + s_{12}a_2 \quad (4)$$

and

$$b_2 = s_{21}a_1 + s_{22}a_2 \quad (5)$$

Setting up the signal flow graph for the configuration in Fig. 3 and using the nontouching loop rules of Mason (5,6), the insertion loss is given by the following expression:

$$L_1 = 20 \log \frac{|(1 - r_s s_{11})(1 - r_L s_{22}) - r_s r_L s_{12} s_{21}|}{|s_{21}| \cdot |1 - r_s r_L|} \quad (6)$$

For a matched system where $r_s = r_L = 0$, Eq. (6) delivers the attenuation

$$A = 20 \log \frac{1}{|s_{21}|} \quad (7)$$

Equations (6) and (7) clearly show that the insertion loss L_1 depends on the property of the two-port device and the reflection coefficients of the source r_s and the load r_L . Otherwise the attenuation is the pure property of the two-port device.

Voltage Loss. Voltage loss is used only for applications in the dc to UHF part of the spectrum where voltage is well defined.

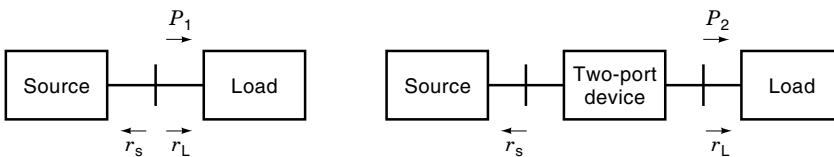


Figure 2. Insertion loss measurement in a non-matched system.

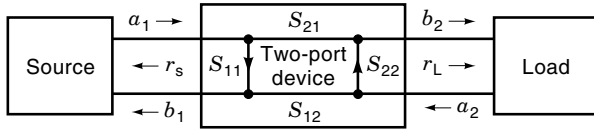


Figure 3. Scattering parameters for a two-port device inserted between source and load.

According to Warner (4), it is defined as follows:

$$L_V = 20 \log \left[\frac{\text{voltage at the input of the two-port device}}{\text{voltage at the output of the two-port device}} \right] \quad (8)$$

and using the scattering parameters

$$L_V = 20 \log \frac{|(1 + s_{11})(1 - r_L s_{22}) + r_L s_{12} s_{21}|}{|s_{21}(1 - r_L)|} \quad (9)$$

Note that L_V is independent of the source reflection coefficient r_s .

When $s_{11} = 0$ and $r_L = 0$, the voltage loss is equal to the attenuation.

Mismatch Loss. At higher frequencies every real two-port device has an input and output reflection coefficient that differs from zero. Therefore, there is always a mismatch between the source and the two-port device and between the two-port device and the load (Fig. 4). This reflects part of the incoming and outgoing wave toward the source and toward the two-port device, respectively, resulting in additional losses.

The mismatch loss between the source and the two-port device is expressed as:

$$L_{m1} = 10 \log \frac{\text{power absorbed at the input of the two-port device}}{\text{maximal available power from the source}} \quad (10)$$

According to Fig. 3 the mismatch loss is given by

$$L_{m1} = \frac{(1 - |r_s|^2)(1 - |r_1|^2)}{|1 - r_s r_1|^2} \quad (11)$$

Similarly, the mismatch loss between the two-port device and the load is given by

$$L_{m2} = 10 \log \frac{\text{power absorbed by the load}}{\text{maximal available power at the output of the two-port device}} \quad (12)$$

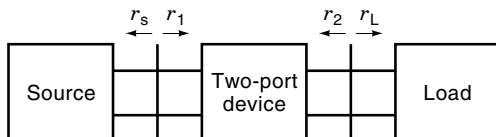


Figure 4. Mismatch loss of a two-port device between a source and a load.

and with the parameters of Fig. 4

$$L_{m2} = 10 \log \frac{(1 - |r_2|^2)(1 - |r_L|^2)}{|1 - r_2 r_L|^2} \quad (13)$$

If several two-port devices are cascaded, the mismatch loss between them has to be calculated similarly and taken into account.

ATTENUATOR

Apart from the natural losses in devices and transmission paths, manufactured devices have well-defined losses. These devices called attenuators are used for measurement and for adjusting power levels to a defined value in transmission systems. Attenuators are probably the most important devices in measurement systems and therefore exist in a large number of different forms (7–9), such as symmetrical, coaxial, waveguide, optical, fixed-value, and variable-loss. The important properties of attenuators are: frequency range, attenuation accuracy, attenuation variation versus frequency, input and output impedance match (reflection coefficient), power handling capacity, and phase linearity.

Balanced-Line Attenuator

Balanced lines are used especially in telecommunications and lately in local area networks and in-house communications systems. Therefore, there is a demand for balanced-line attenuators. Chains of symmetrical double-T or double-Π circuits, as shown in Fig. 5, are mostly used (10).

The reference handbooks (10) give formulas and tables to determine the circuit elements for a given line impedance and different element attenuations. The circuit has to be symmetrical to the ground plane and well matched to the line impedance. Special techniques are given to optimize the circuit for small frequency dependency.

Variable-value attenuators are commercially available for different impedances (150 Ω, 120 Ω) in the frequency range from dc to several megahertz and have attenuation ranges from 0 dB to 132 dB.

Coaxial-Line Attenuator

Fixed-Value Coaxial Attenuator. Coaxial attenuators generally have multi-octave bandwidth or frequently operate from dc to several gigahertz (GHz). Coaxial transmission lines are generally operated in the transverse electromagnetic (TEM) mode and therefore an obvious solution for attenuators is to use lumped elements that are small compared to the wavelength. There are four major constructions: T-circuit, Π-circuit, distributed lossy line and distributed thin-film technology.

T-Circuit Attenuator. The T-attenuator circuit is shown in Fig. 6. The values of R_1 and R_2 are calculated from the following formulas, where K is the transmission coefficient (7,10).

$$R_1 = Z_0 \frac{K - 1}{K + 1} \quad (14)$$

$$R_2 = \frac{2Z_0 K}{K^2 - 1} \quad (15)$$

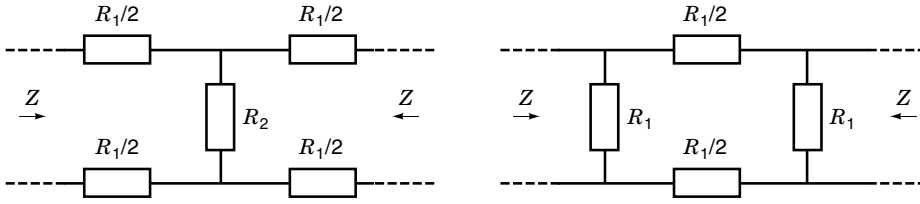


Figure 5. Basic symmetrical double-T and double-Π attenuation circuits.

The attenuation in decibels is given by

$$A(\text{dB}) = 20 \log K \quad (16)$$

Resistive rods for R_1 are often used, and disk resistors are used for R_2 , or film resistors are used for both R_1 and R_2 .

Π-Circuit Attenuator. Techniques similar to the T-circuit are used for the Π-circuit shown in Fig. 7, and the corresponding formulas are as follows (7,10):

$$R_1 = Z_0 \frac{K + 1}{K - 1} \quad (17)$$

$$R_2 = Z_0 \frac{K^2 - 1}{2K} \quad (18)$$

At higher microwave frequencies the elements of the T- and Π-attenuators have dimensions comparable to the wavelength, and therefore the reactance of the elements gets more important and therefore degrades the performance of the attenuator.

Lossy-Line Attenuator. Distributed lossy-line attenuators, explained by Weber (11), have very favorable performance. As long as the attenuation is not very high, they have a flat attenuation response and an excellent impedance match. The center conductor of the transmission line is mostly made of lossy material using thin film deposited on a substrate with a circular cross section. The disadvantage of lossy-line attenuators is that they have a lower frequency limit depending on their physical length.

Thin-Film Attenuator. Distributed thin-film attenuators (12) use a resistive material deposited on a dielectric substrate for the series and shunt losses (Fig. 8).

Strip transmission lines and narrow ground electrodes on the longitudinal sides are added to the input and output terminals. The characteristic impedance and the attenuation are constant and given as follows (7):

$$Z_0 = \rho \sqrt{\frac{D - a}{4a}} \quad (\Omega) \quad (19)$$

$$\alpha = \sqrt{\frac{4}{a(D - a)}} \quad (\text{in nepers per unit length}) \quad (20)$$

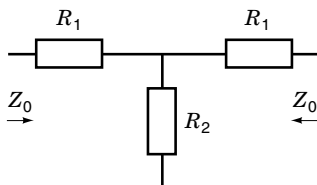


Figure 6. T-attenuator circuit.

Equation 20 shows that the attenuation is independent of the resistivity of the film as long as the film is homogenous. The attenuation depends only on the geometry and therefore is insensitive to temperature changes.

Variable-Value Coaxial Attenuator. There are two types of variable attenuators: continuous variable attenuators and step attenuators. Continuous variable attenuators have the advantage of being noninterruptive when changing attenuation. This feature is important for some measurements, for example, receiver sensitivity measurements. On the other hand, these attenuators may sometimes lack accuracy, settability, impedance match, high insertion loss and have limited bandwidth.

Step attenuators are very accurate and have the following qualities: good reproducibility, low insertion loss, excellent impedance match, and wide operating bandwidth. But most step attenuators are interruptive when changing the attenuation value.

Continuously Variable Attenuator

Piston Attenuator (Waveguide Beyond Cutoff Attenuator). The piston attenuator is one of the oldest continuously variable microwave attenuators (13). It is used especially as a precision attenuator or for handling high power levels. This type of attenuator uses a waveguide below its cutoff frequency. According to transmission line theory, the amplitude of a wave launched into such a waveguide decays exponentially. So the attenuation is calculable. Most constructions use a circular cross section and a magnetic coupling that generates the lowest cutoff higher order transverse electric mode (TE_{11}). Figure 9 shows the simplified construction of a piston attenuator. By sliding the two concentric cylinders into each other, the physical displacement of the coupling coil is changed and therefore also the attenuation. Special care must be taken to avoid unwanted higher order modes.

Attenuation as a function of wavelength is given per unit length as (10)

$$A = \frac{2\pi \cdot 20}{\lambda_c \cdot \ln 10} \sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \quad \text{in decibels per unit length} \quad (21)$$

where λ_c is the cutoff wavelength of the waveguide and λ is the free-space wavelength.

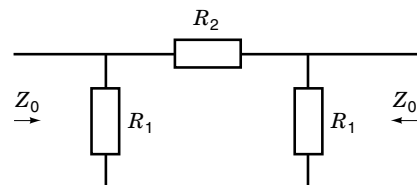


Figure 7. Π-attenuator circuit.

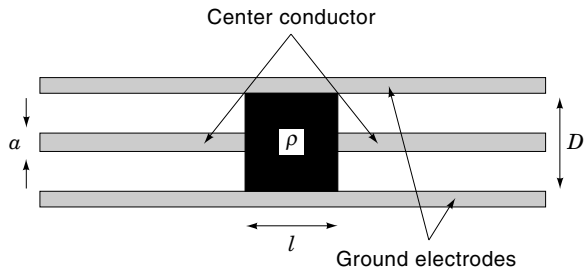


Figure 8. Distributed thin-film attenuator element.

If the operating frequency is chosen to be much lower than the cutoff frequency, the term λ_c/λ is negligibly small for the operating frequency band, and a flat attenuation frequency response is achieved. The advantages of cutoff attenuators are that they are calculable and that high accuracy is achieved (0.001 dB/10 dB for 60 dB attenuation). For this reason, cutoff attenuators are often used as standard attenuators. Their main disadvantage is the high insertion loss (15 to 20 dB) because tight coupling has to be avoided so as not to stimulate higher modes.

Resistant Card, T-, or Π -Type Attenuator. Several constructions of variable attenuators using resistive cards or resistive films are on the market (9). The variable card attenuator operates like a potentiometer. A resistive film is fixed on a substrate so that the resistance between the input and the output is varied with a movable coaxial wiper, and thus the amount of attenuation changes. This type of attenuator does not have good input and output impedance matches. More sophisticated constructions use T- or Π -type structures where the series and the shunt resistors are changed simultaneously. Therefore the input and output ports of these attenuators are quite well matched. The minimum insertion loss is on the order of 4 dB, and they operate up to several gigahertz.

Lossy-Line Attenuator. Lossy-line attenuators use a lossy center conductor partly covered with a thin sliding shield (7). This effectively changes the length of the resistive conductor and thus the attenuation. Some constructions use microstrip lines with variable lossy walls (9). This type of attenuator is limited in its maximal attenuation because of the length of the device.

Pin Attenuator. Pin attenuators change the loss in either a step or continuous mode. The series and shunt resistors are replaced by *pin* diodes, electronic devices that change their conductivity. The diodes are controlled by a bias current. Various types of circuits are available, such as series, shunt, bridged T, and Π (9). *Pin* attenuators are electronic circuits and therefore

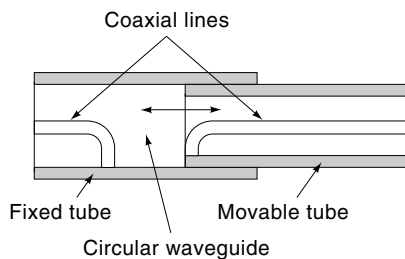


Figure 9. Principle of a piston attenuator.

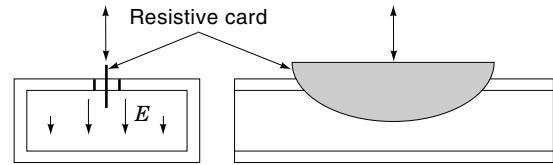


Figure 10. Principle of a waveguide flap attenuator.

may produce harmonics. Below about 10 MHz, the *pin* diode exhibits some rectifying behavior. At much higher frequencies, the *pin* diode behaves more like a variable resistor. Because of the matching circuits for input and output and the bias network, most devices operate in a limited frequency band. Attenuation from 2 dB to 80 dB is achieved.

Step Attenuator. Step attenuators always use a set of fixed attenuators that are switched into a line by different mechanisms. The steps are mostly 0.1 dB, 1 dB, and 10 dB. The 0.1 dB steps generally cover a range of 1.1 dB, the 1 dB steps a range of 11 dB, and the 10 dB steps cover a range up to 110 dB. The step attenuator has excellent repeatability, covers a wide frequency range (e.g., dc to 26.5 GHz), has a good impedance match, and mostly has a flat frequency response of the attenuation value.

In the turret-type coaxial attenuator a set of coaxial fixed attenuators is placed in a cylindrical arrangement. With a rotary motion the different elements (e.g., 0 dB, 10 dB, 20 dB, ...) are switched between the junctions of the transmission line.

Another type of step attenuator uses a variety of fixed attenuation elements. The different attenuation elements are cascaded by switches or bypassed. The switches may be activated manually or electrically.

Waveguide Attenuator

Waveguide attenuators work mostly in the entire usable waveguide bandwidth, which is not quite half an octave. To attenuate a wave propagated in a waveguide, either the electric or the magnetic field or even both are influenced. As an example, a resistive card inserted into the waveguide parallel to the *E*-field attenuates it. Another technique uses lossy walls that influence the current in the waveguide wall. Most of these attenuators are not phase-invariant.

Fixed-Value Waveguide Attenuator. The waveguide flap attenuator (Fig. 10) and the side-vane attenuator (Fig. 11) (7) are very popular.

The flap attenuator is based on a resistive card inserted in the center of the waveguide parallel to the *E*-field. The more the card dives into the waveguide, the more the *E*-field is weakened, and therefore attenuation increases. A smooth

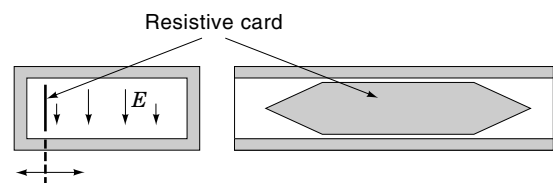


Figure 11. Principle of a side-vane attenuator.

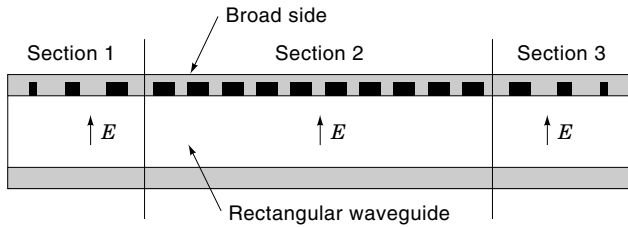


Figure 12. Lossy-wall attenuator.

card shape is chosen to minimize the reflection caused by the discontinuity.

The side-vane attenuator (Fig. 11) influences the E -field similarly. The vane is always completely inside the waveguide, but it uses the fact that the E -field varies along the broad side. For the most popular TE_{10} mode, the E -field is zero at the side wall and has its maximum in the center of the waveguide. Therefore the position of the resistive card defines the attenuation value. A smooth shape minimizes the reflection of the discontinuity.

Several constructions of lossy-wall attenuators exist. One version is shown in Fig. 12 (14).

Section 2 with equally spaced slots filled with a lossy material defines the attenuation. Sections 1 and 3 are configured to minimize the reflection due to the discontinuity in the wall. The lossy-wall attenuator withstands high power because the dissipated heat is transferred to any cooling system.

Variable-Value Waveguide Attenuator

Flap and Side-Vane Attenuator. By adding a mechanism that changes the position of the resistive card, the fixed-value-flap and side-vane attenuator are easily transformed into a variable-value attenuator. It is often used as a settable attenuator.

Rotary-Vane Attenuator. The rotary-vane attenuator was invented in the early 1950s by E. A. N. Whitebread (Elliot H. Brothers, London) and A. E. Bowen (Bell Telephone Laboratories). It was proposed and developed as a precision waveguide attenuator (15–19). The rotary-vane attenuator consists of three sections of waveguide that have a resistive film spanned across the waveguide, as shown in Fig. 13. The middle section has a circular cross section and can be rotated with respect to the two fixed-end sections. Figure 13 illustrates the principle

of a rotary-vane attenuator. For clarity, the transitions from a round to a rectangular waveguide at both ends are omitted.

The electric field is perpendicular to all resistive films whenever the films are aligned. In this case no current flows in the resistive film, and therefore no attenuation occurs. If the center part is rotated by an angle θ , the component $E_{\sin\theta}$ in θ produce a current flowing in the resistive film and are absorbed. Thus the resulting attenuation is given by

$$A(\text{dB}) = -40 \log(\cos \theta) + A_0 \quad (22)$$

The rotary vane attenuator has the following advantages:

The attenuation is almost independent of the frequency.

The phase shift is very small. The phase variations are smaller than 1° up to 40 dB attenuation.

The input and output VSWRs are very low under all conditions.

The attenuation is not sensitive to the resistive film as long as its attenuation is high enough.

The attenuation is not sensitive to temperature changes.

Rotary vane attenuators are commercially available for most of the known waveguide bands. They typically have an accuracy of 2% of the reading in decibels or 0.1 dB, whichever is greater.

Attenuator for Optical Fibers

Attenuators have to be adapted to an optical fiber system. Therefore different attenuators for single-mode or multimode applications are available and mostly operate within one or two wavelength windows. Various attenuation techniques are used to reduce the transmitting light, such as lateral or axial displacement of the fibers or optical prisms, and inserting absorbing filters.

Fixed- or Adjustable-Value Attenuator. A pair of lenses is used in most fiber optical attenuators to collimate the light of the input fiber and to refocus it to the output fiber. Any attenuation mechanism, such as absorbing filters or absorbing glass with variable thickness can be inserted into the optical beam path. Attention has to be given to the attenuation mechanism so that there is very little polarization-dependent loss. Fixed-value attenuators are commercially available for a range of 3 dB to 40 dB with an accuracy of 0.5 dB to 1 dB.

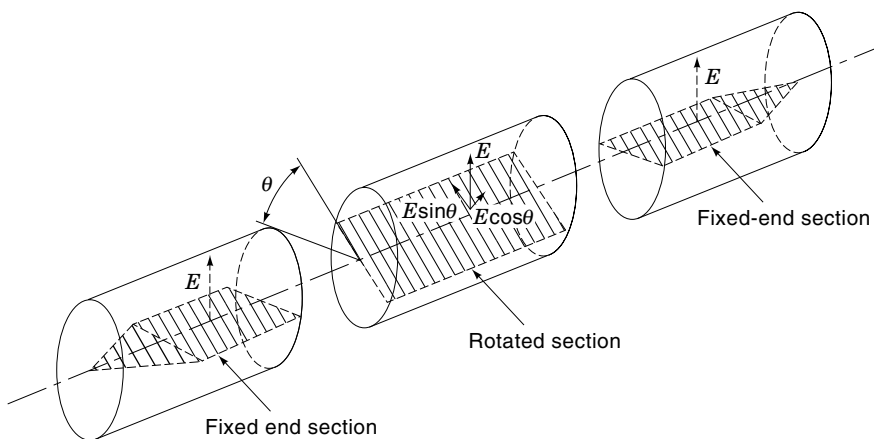


Figure 13. Principle of a rotary-vane attenuator.

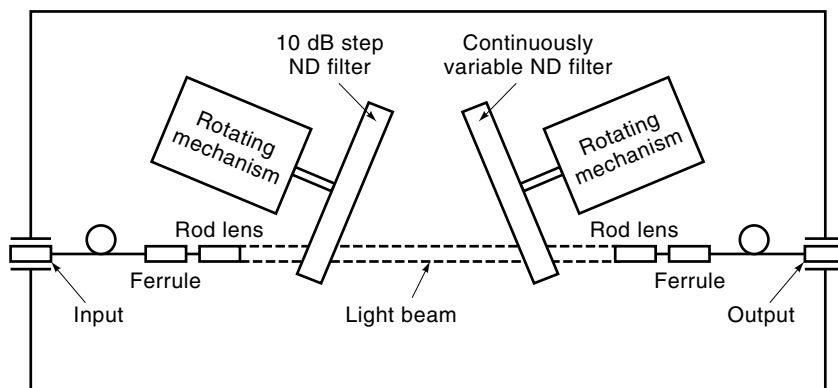


Figure 14. Basic configuration of an optical section.

Variable-Value Attenuator. The reflection type variable attenuator often combines a series of 10 dB steps for the high values and a continuous variable attenuation of up to 10 dB or 15 dB. Figure 14 shows the basic configuration of an optical section (20).

A rod lens collimates the light of the fiber into a spatial beam that passes two neutral density filters (ND-filter). The second rod lens focuses the spatial beam into the output fiber. The ND-filter must have high uniformity, little attenuation change with time, and flat spectral transmittance. This is achieved by vacuum-depositing Ni and Cr alloys onto a glass substrate. The ND-filters are the reflecting type. To avoid multiple reflections and to ensure that the light is not reflected into the input lead, the two filters are inclined at a small angle with respect to the optical axis.

Commercially available attenuators of this type have an insertion loss of about 3.5 dB, a range of up to 65 dB, and an accuracy of about ± 0.5 dB per 10 dB. They are available for wavelengths of 850 nm, 1300 nm, and 1550 nm.

Calculable Attenuation Standards

Many attenuators have been described in the previous sections, but only a few are suitable as attenuation standards. An attenuation standard is a device that can be traced to the SI

units by an unbroken chain. The application of the described standards depend on the frequency band and the technique used in the measurement system.

Kelvin–Varley Divider. The Kelvin–Varley divider (KVD) was described for the first time by Varley in 1866 (21). It is a resistive voltage divider and operates from dc up to several hundred kilohertz. Figure 15 is an example of a four-decade Kelvin–Varley divider.

Each decade consists of eleven equal resistors except for the last decade which has only ten. The decades are connected by switches that always span two resistors. The value of the resistors of each following decade is reduced by a factor of 5. The four-decade Kelvin–Varley divider allows varying the output-to-input voltage ratio from zero to one in steps of one part in 10^4 . The unloaded KVD has a constant input impedance that is independent of the switch setting whereas the output resistance varies with the setting. The original type of KVD (Fig. 15) requires either very large or very small resistance values because each decade needs five times larger values. Due to stray capacitance, the divider reacts similarly to an RC -filter. For example, a 100 k Ω input impedance limits the 3 dB bandwidth to about 100 kHz. To avoid large resistance values, modified constructions with a resistor shunting the decades have been developed.

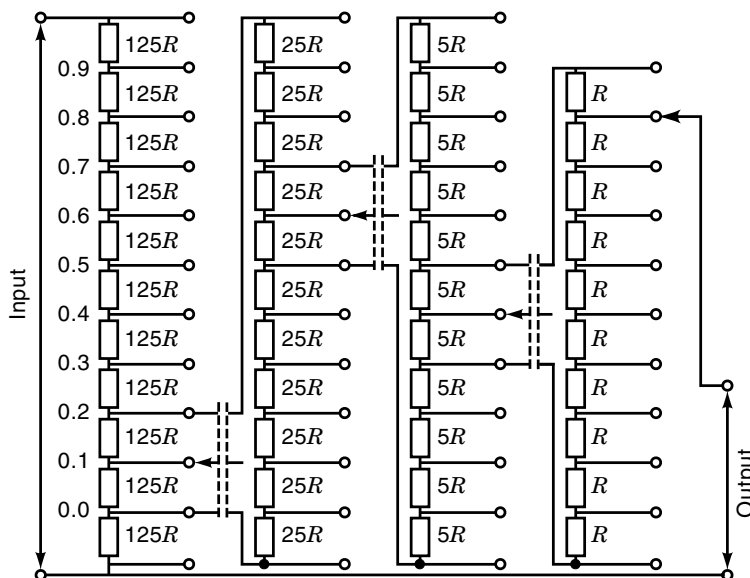


Figure 15. Principle of a four-decade Kelvin–Varley divider.

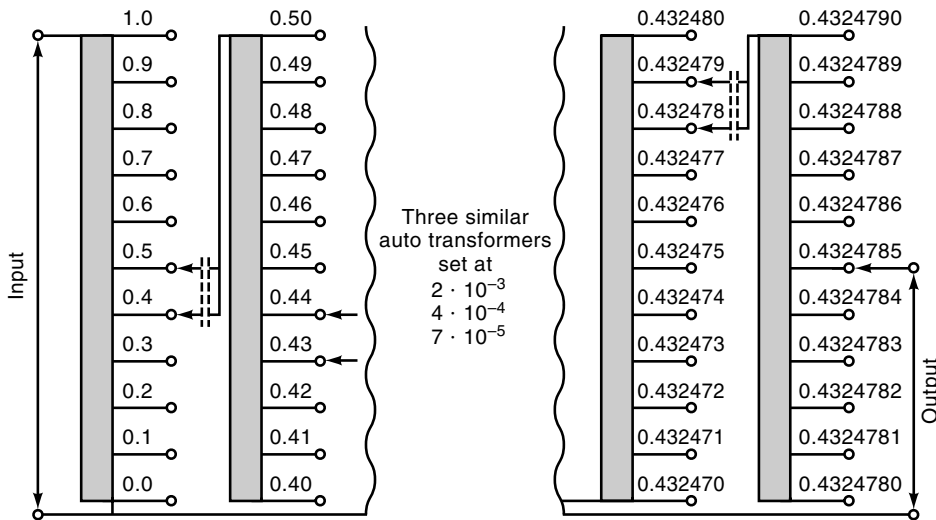


Figure 16. Principle of a seven-decade IVD.

Two major errors determine the accuracy: deviations of the resistors from nominal values and the resistances of the switch contacts and leads. The first three or four decades are the most sensitive, and therefore these resistors often are adjustable. Several calibration techniques have been developed and are described in the literature (22,23).

Today commercially available Kelvin–Varley dividers have up to seven decades, have an absolute linearity of ± 1 part in 10^7 , and long term stability of ± 1 part in 10^6 per year.

Inductive-Voltage Divider. The inductive-voltage divider (IVD), also called a ratio transformer, is an exceptionally accurate variable attenuation standard. It consists of a number of very accurately tapped autotransformers. The autotransformers are connected together by high quality switches. The IVD operates from 10 Hz to about 100 kHz, and the greatest accuracy is achieved at about 1 kHz. In 1962 Hill and Miller (24) described a multidecade IVD with seven decades and a resolution of 1 part in 10^7 . Figure 16 shows the principle of a seven-decade IVD with a setting of 0.4324785. The output-to-input voltage ratio can be set from zero to one.

The tapped autotransformers are constructed by winding exactly equal lengths of copper wire on a high permeability toroidal core. A superalloy having an extremely high permeability ($>100,000$) and low hysteresis loss is preferred as a core material. For an exact division of the tapped autotransformer, it is not necessary to have a 100% coupling between the ten inductors (4). But the 10 self-inductances and the 45 mutual inductances have to be exactly equal.

The following error sources limit the accuracy of IVDs:

- inequality in the series resistances and the leakage inductances of the sections in each autotransformer
- inhomogeneities in the magnetic cores
- distributed admittances between the windings
- internal loading caused by the later stages
- impedances of the connecting leads and switch contacts
- variations in the input voltage, frequency, and ambient temperature

With careful design the mentioned errors can be minimized. Recently, programmable binary IVDs with 30 bits, a resolution

of 1 part in 10^9 , and a linearity of 0.1 ppm have been developed (25).

IVDs with eight decades and an accuracy of four parts in 10^8 are commercially available.

Intermediate-Frequency Piston Attenuator. The IF piston attenuator is based on the same principle as the attenuator previously described for RF, but it is designed to operate at a specific, fixed frequency, mostly 30 MHz or 60 MHz. As Eq. 21 shows, attenuation depends on the cutoff wavelength λ_c , the free-space wavelength λ , and the displacement of the two coils. The waveguide dimensions, which can be determined, define the cutoff wavelength, and the displacement can be measured very precisely. Therefore the IF piston attenuator is used as a calculable standard. Figure 17 shows a simplified diagram of an IF piston attenuator.

The standard IF piston attenuator consists of a high-precision circular cylinder that has excellent conductivity, a fixed coil, and a movable coil mounted on a piston. The piston attenuator operates in the H_{11} (TE_{11}) mode that has the lowest attenuation. A well-designed metal strip filter in front of the fixed coil attenuates the higher modes. To allow smooth movement and to avoid any scratches, the plunger carrying the moving coil is insulated. Equation 23 expresses the attenuation per unit length more precisely:

$$A = \frac{s_{11}}{r} \sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \cdot \epsilon - \frac{\delta}{r} \quad (23)$$

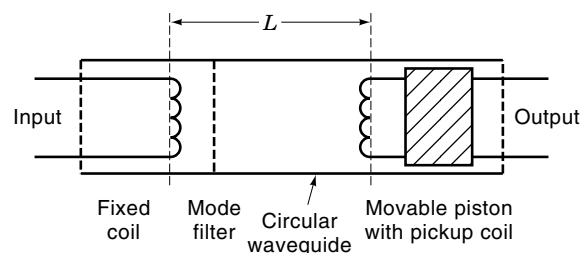


Figure 17. Simplified IF piston attenuator.

in nepers per unit length, where $\lambda_c = 2\pi r/s_{11}$, λ_c is the cutoff wavelength, λ is the free-space wavelength, s_{11} the first zero of the Bessel function $J_1 = 1.8411838$, r the radius of the cylinder, and δ is the skin depth.

Highly accurate standard attenuators use a laser interferometer to accurately determine the displacement of the coil. Yell (26,27) and Bayer (28) developed extremely accurate IF piston attenuators with a dynamic range of 120 dB, a resolution of 0.0001 dB, and an accuracy of 0.0002 dB/10 dB over the linear range of 90 dB.

Rotary-Vane Attenuator. The principle of the rotary-vane attenuator is described in the section Variable Waveguide Attenuator. Because the attenuation is given by the equation

$$A(\text{dB}) = -40 \log(\cos \theta) + A_0 \quad (24)$$

where A_0 is the insertion loss at a setting of $\theta = 0$, the device can be used as a calculable primary standard.

The rotating angle θ of the center vane has to be determined very precisely, especially for higher attenuation values. As an example, a rotational angle accuracy of $\pm 0.001^\circ$ results in an attenuation accuracy of ± 0.01 dB at a setting of 60 dB. Especially precise optical readouts have been developed by national standards laboratories (29–31) to allow an angular resolution of $\pm 0.001^\circ$. Following are the main error sources for precision rotary-vane attenuators:

- misalignment of the end vanes
- insufficient attenuation of the central vane
- incorrect readout of the rotational angle
- eccentricity of the rotor
- leakage of the rotating joints
- internal reflections at the ends of the three vanes

Careful design of the attenuator results in an accuracy of ± 0.0015 dB up to 16 dB at 10 GHz.

Comparison of Attenuation Standards. The attenuation standards mentioned previously are used in various precision measurement systems, such as RF-, IF- or LF-substitution. The standards have very different accuracy depending on the attenuation setting. Figure 18 shows a comparison of different precision attenuation standards used in national metrology laboratories (32).

Optical Attenuation Standards. Imamura (33) shows one solution of a calculable optical attenuation standard that is used to calibrate precision attenuators. The key element is a rotating disk with a well-defined opening. The device operates as an optical chopping system (Fig. 19).

As long as the response of the detector is slow compared to the rotational speed ω , the ratio P_1 to P_0 defines the attenuation. In this case the attenuation depends only on the opening angle θ (in radians) and is given by the following equation:

$$A = -10 \cdot \log \frac{P_1}{P_0} = -10 \cdot \log \frac{\theta}{2\pi} \quad (25)$$

An opening angle of 36° defines an attenuation of 10 dB. Special care has to be paid to the diffraction of light at the edges

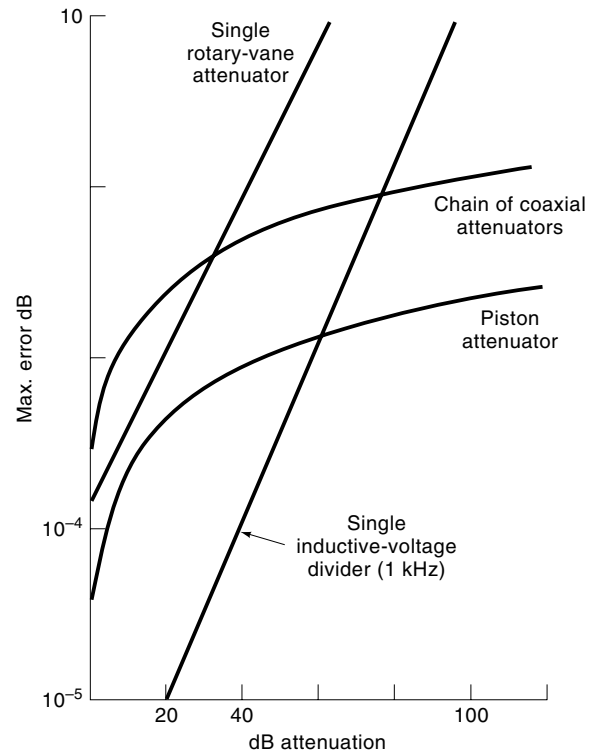


Figure 18. Comparison of attenuation standards.

of the disk openings and the stability of the light source and the sensor. The overall accuracy is estimated to be ± 0.008 dB for 10 dB attenuation.

MEASUREMENT OF ATTENUATION

Various kinds of measurement systems are used depending on the frequency range, the type of attenuator, the required accuracy and the available standards. Most of the modern attenuation measurement systems are computer- or microprocessor-controlled. They use the same principle as the manual systems but operate much faster. If maximum accuracy is required, manually controlled measurement systems are often preferred.

Low-Frequency Measurement Systems

Low frequency attenuation measurements are used mainly in communication systems where voice, video and data have to

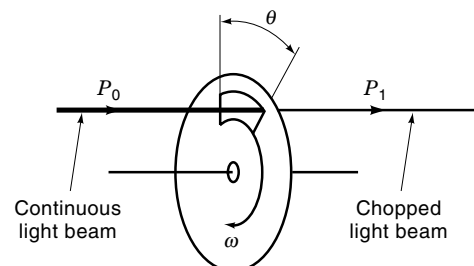


Figure 19. Principle of an optical chopping system.

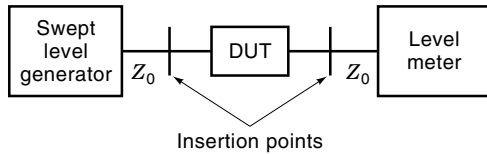


Figure 20. Principle of a direct measurement system.

be transmitted with as little distortion as possible. Dedicated test systems have been developed for testing and adjusting the communication systems working either with coaxial lines (50 Ω or 75 Ω) or balanced lines (124 Ω , 150 Ω , or 600 Ω).

Direct Measurement. Operation of communication systems requires a great number of test systems for which low-cost test systems that are easily handled were developed. The systems are based on current, voltage, or power measurement and operate in a frequency band from 200 Hz up to 30 MHz. The test system (Fig. 20) consists of a tuneable generator with a known constant output level and a wide-band voltmeter or a high-sensitivity, selective superheterodyne receiver.

Many test systems can be switched to operate in either coaxial-line (50 Ω or 75 Ω) or balanced line (124 Ω , 150 Ω , or 600 Ω) configuration. In the balanced-line mode the test systems have a limited frequency range of 200 Hz to several MHz depending on the impedance selected. In the selective-level meter mode, bandwidths of 25 Hz to 3.1 kHz are used, and the dynamic range achieved is on the order of 80 dB. The attenuation measurement accuracy in the frequency range indicated is about 1 dB.

Low-Frequency Substitution Method. The LF-substitution method is based on a precisely calibrated low-frequency reference attenuator used in a parallel or a serial configuration (Fig. 21).

Because the attenuation of the device under test (DUT) is compared with that of the reference attenuator, neither the output level of the generator nor the absolute level indication of the receiver have to be known. The only requirements are that generator and receiver remain stable during measurements. The accuracy is determined mainly by the calibration of the reference attenuator.

Radio Frequency and Microwave

In radio frequency and microwave many different measurement principles are known. They all have their own characteristics: one is meant to measure low values of attenuation, another to measure high values, a third to achieve highest accu-

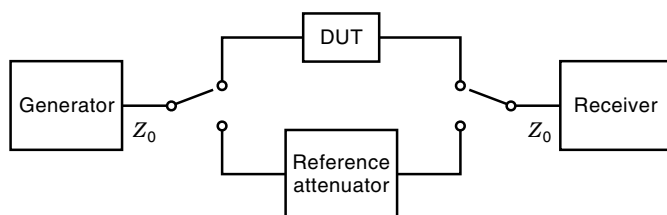


Figure 21. Principle of a LF-substitution method.

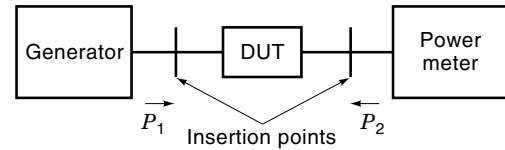


Figure 22. Principle of the power ratio method.

racy. The most popular measurement principles are discussed in the following sections.

Power Ratio Method. The power ratio method (4,40) (Fig. 22) is very simple for measuring attenuation. It is commonly used as long as maximum accuracy is not required.

The method is based on the linearity of the power meter or the receiver used. First the power P_1 of the generator is measured without the device under test (DUT) and then P_2 is measured with the DUT inserted. The attenuation of the DUT is calculated by the ratio of P_2 to P_1 :

$$A(\text{dB}) = 10 \log \frac{P_2}{P_1} \quad (26)$$

To measure attenuation, the insertion points have to be matched either by tuners or matching pads. The square law characteristic of the power sensors and the noise limit the dynamic range to about 40 dB. If a tuned receiver is used instead of a power meter, the measurement range is extended to about 100 dB.

Several error sources influence the accuracy:

- the stability of the generator and the detector system
- the frequency stability of the generator
- the matching at the insertion points
- the square law region of the detector system
- the crosstalk for high attenuation measurement

Commercially available systems using a tuned receiver achieve a measurement uncertainty of 0.1 dB at 50 dB attenuation. These systems are easily automated by controlling the instruments with a computer.

In national standards laboratories very sophisticated systems have been developed resulting in an accuracy of 0.06 dB at 50 dB attenuation (34,35).

Voltage Ratio Method. The voltage ratio method makes use of high resolution ac-digital voltmeters (ac-DVM) available now. Because the ac-DVMs work only up to several MHz, the RF signals have to be down-converted to low frequency. Figure 23 shows the principle of a voltage ratio system working at an audio frequency of 50 kHz.

If a synthesizer locked to the same reference frequency is used as a signal generator and local oscillator, a very stable audio frequency f_a (e.g., 50 kHz) is generated. The audio frequency signal is amplified and measured with an ac-DVM. If U_1 is the voltage measured with the two insertion points clamped together and U_2 is the voltage with the DUT inserted, the

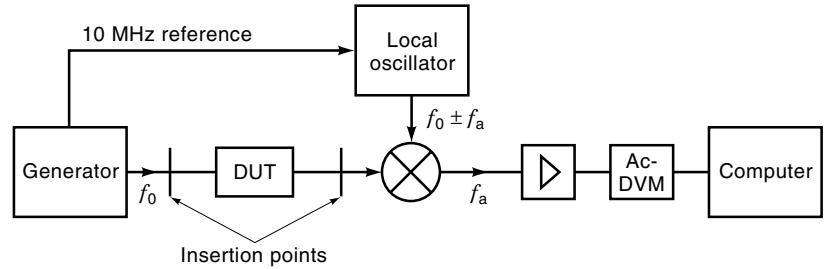


Figure 23. Principle of the voltage ratio method.

attenuation is given by

$$A(\text{dB}) = 20 \log \left(\frac{U_2}{U_1} \right) + C \quad (27)$$

where C is the correction factor in decibels for the nonlinearity of the amplifier and the DVM.

The dynamic range of the direct system is about 20 to 30 dB. More sophisticated systems achieve an uncertainty less than 0.001 dB for 20 dB attenuation. By adding a gauge block technique, for example, a calibrated step attenuator (10, 20, 30, 40, 50, 60, 70 dB) in series with the DUT in the RF path, the range is extended to 90 dB with excellent accuracy of 0.001 dB (32).

The error sources which limit the measurement uncertainty are

- the matching of the insertion points
- the generator output level stability
- the AF-amplifier stability
- the AF-amplifier and ac-DVM linearity
- the mixer linearity
- the gauge-block attenuator stability and reproducibility
- the crosstalk for high attenuation measurement

IF-Substitution Method. The IF-substitution method (4,40) (Fig. 24) gives good accuracy, operates over a large dynamic range, and is used up to very high frequencies. Most systems operate in a parallel substitution mode.

The signal passing through the DUT is mixed to an IF of 30 or 60 MHz. This signal is compared with the signal of the 30 MHz reference oscillator and the standard attenuator by a narrowband 30 MHz receiver (mostly with synchronous detection). In a first phase the insertion points are clamped together, and the standard attenuator is adjusted until there is no

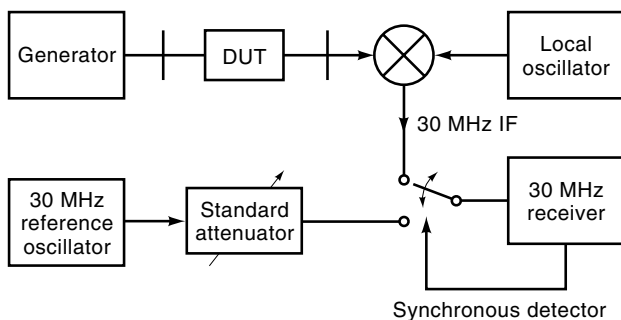


Figure 24. Principle of the IF-substitution method.

switching signal (i.e., 1 kHz) detectable any longer. The reading A_1 of the standard attenuator is taken. In a second phase the DUT is inserted, the standard attenuator is adjusted so that the signal of the standard attenuator equals the signal of the mixer, and the reading A_2 is taken. The attenuation of the DUT is given by the difference A_2 minus A_1 between readings.

A piston attenuator, an inductive-voltage divider, or a high-precision resistive attenuator can be used as a standard attenuator.

In national standards laboratories very high-precision piston attenuators with a resolution of 0.0001 dB over a 100 dB range have been used in a parallel substitution system. The accuracy achieved is better than 0.001 dB per 10 dB step (27,32,36).

Accuracy of about 0.002 dB and a dynamic range of 100 dB have been achieved by using a seven-decade 50 kHz inductive voltage divider in a parallel substitution system (37).

Weinert (38) proposed a parallel IF complex vector substitution system using a high precision IF attenuator. The system has a single-step dynamic range of 140 dB and a display resolution of 0.001 dB.

The following error sources limit the accuracy and the dynamic range:

- the matching at the insertion points
- the level stability of the signal source
- the mixer linearity
- the noise balance
- the level stability of the IF reference oscillator
- the standard attenuator resolution and stability
- the crosstalk for high attenuation measurement

RF-Substitution Method. In the RF-substitution method (4,40) (Fig. 25), the reference standard attenuator and the DUT operate at the same frequency. Because the attenuation of the reference standard is compared either in a series or in a parallel substitution system with the DUT, the results are independent of the receiver characteristics. A rotary-vane attenuator, a piston attenuator, or a chain of well-matched and precisely calibrated attenuators (e.g., step attenuator) is used as a reference standard.

In the first step the insertion points are clamped together, and the reference standard is adjusted to a value A_1 according to the estimated attenuation of the DUT. The receiver shows the reading U_1 . In the second step the DUT is inserted, and the reference standard is adjusted to get the same reading U_1 at the receiver A_2 . The attenuation of the DUT is calculated as the difference between the two decibel readings of the reference attenuator, A_1 minus A_2 .

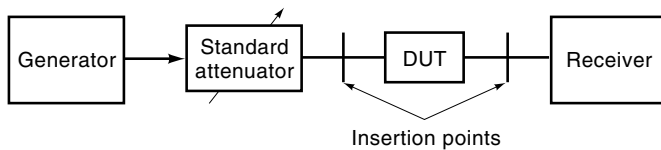


Figure 25. Principle of the series RF-substitution method.

Scalar Measurement. All of the attenuation measurement systems described in the previous sections provide scalar measurements. There are many commercial scalar network analyzers available (Fig. 26) (8,39). These analyzers measure the input reflection and the attenuation of the device under test. Because mostly wideband detectors are used, only the magnitude of two quantities can be determined.

The signal of the sweep generator is divided by a power splitter or a directional coupler into reference and measurement paths. The directional bridge or coupler measures the reflected wave of the DUT. The analyzer forms the ratio A/R which is proportional to the input reflection coefficient of the DUT. Using a third detector, the attenuation is measured by calculating the ratio B/R . Most scalar network analyzers are microprocessor- or computer-controlled and offer simple correction methods. The calibration for reflection measurements is frequently done by using open and short circuits, and a connect through normalization is used for the transmission path. Because these analyzers are broadband systems, they operate very fast and are easily expandable to highest frequencies. Commonly used scalar network analyzers operate from 10 MHz to 18 GHz or 26 GHz, and often their frequency range can be extended to 50 GHz in coaxial lines and to 110 GHz in wave guides. The dynamic range is limited to about 75 dB by the square law characteristic of the detectors and noise. The measurement accuracy achieved is quite reasonable for example, 0.6 dB measuring a 30 dB attenuator. The insertion points have to be well matched.

The following errors influence the measurement uncertainty:

- the harmonics of the sweep generator;
- the matching of the insertion points;
- the square law characteristic of the detectors; and
- the sweep generator level stability.

Vector Measurement. Vector measurements enable characterising a two-port circuit completely. In addition to the magni-

tude, the phase of the scattering parameters is also determined. There are two major concepts for measuring the complex parameters of a two port device: the vector network analyzer and the six-port technique.

Modern vector network analyzers (8,39,40) measure all four scattering parameters: s_{11} , s_{21} , s_{12} , and s_{22} without the necessity of turning the DUT around. Therefore they are symmetrical (Fig. 27) and measure in both directions.

The basic concept looks similar to that of the scalar network analyzer. The signal of the generator is divided into reference and measurement paths. In forward measurements, the directional bridge A determines the reflected signal, bridge B determines the transmitted signal, and vice versa for the reverse case. Instead of using diode detectors, the signals are down-converted to an intermediate frequency and analyzed in magnitude and phase. Synthesized sweep generators and synchronous detection are being used to obtain high accuracy for magnitude and phase measurements. Because the complex signals are measured, the main errors due to component imperfections may be corrected. Frequently a 12-term error model is applied to correct the source and load match, the bridge characteristics, the transmission leakage cross talk and down-converter characteristics. In the first phase well-known standards (e.g., open, short, line) are measured, and the 12 error parameters are determined. In the second phase the DUT is measured and the data corrected according to the calculated error terms. Several different techniques for measuring the error parameters are used, such as open-short-load, transmission-reflect-line, line-reflect-line, etc. Each technique uses different kinds of reference standards, such as short and open circuits, well defined lines, known loads.

Excellent performance is achieved by using the 12-term error correction technique. For example, at 20 GHz, load and source match better than -38 dB return loss, transmission tracking is better than 0.06 dB and cross talk is less than -104 dB. As a result a 30 dB attenuator can be measured at 20 GHz with an uncertainty of 0.07 dB.

Vector network analyzers are commercially available in coaxial configurations in frequency bands from about 100 kHz to 50 GHz and in waveguides up to 110 GHz. Some specially dedicated systems operate in waveguides up to 1000 GHz.

The measuring uncertainty is defined mainly by the following parameters:

- accuracy of the reference standards
- stability of the generator and of the detection system

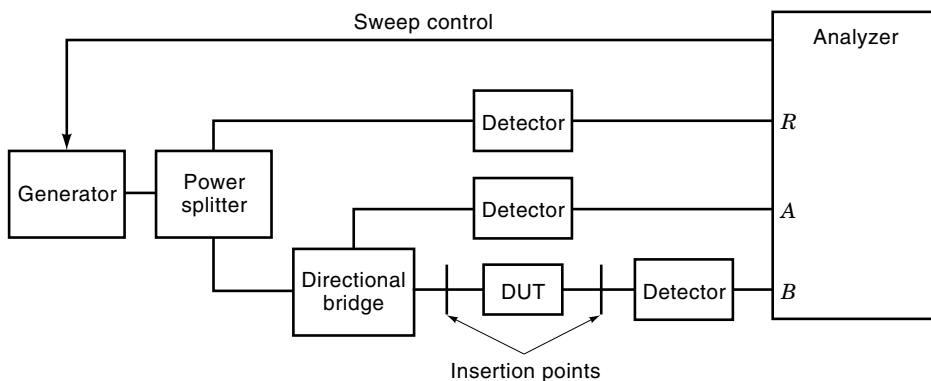


Figure 26. Principle of a scalar network analyzer.

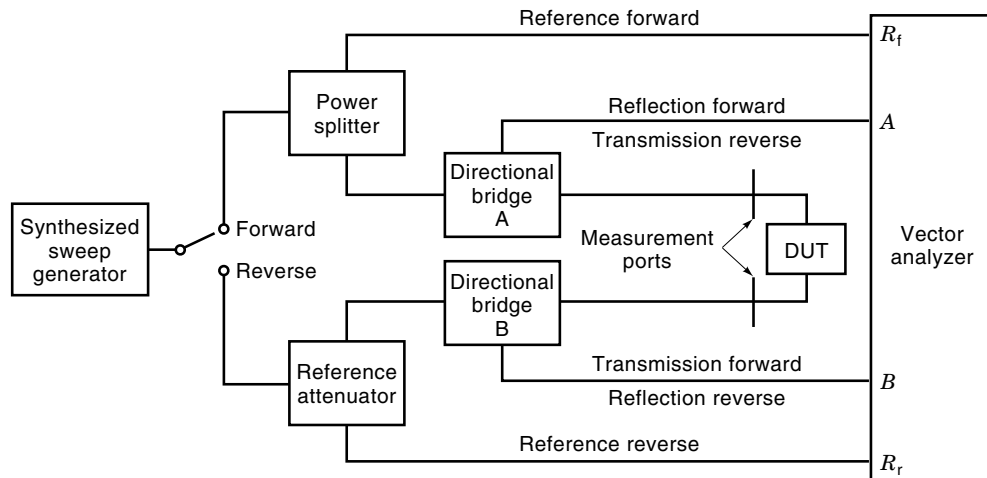


Figure 27. Principle of a vector network analyzer.

stability of the connection cables
 repeatability of the connectors
 accuracy of the built-in software to calculate the error parameters and DUT scattering parameters

The six-port technique is another method for measuring the complex scattering parameters of a device. The magnitude and the phase of the signal are calculated from four scalar power measurements made with detectors arranged as shown in Fig. 28 (8,40).

The four power sensors gather enough information to calculate the magnitude and phase of the reflection of the DUT and the power launched into it. The calibration of the six-port device is rather complicated because a set of quadratic equations has to be solved. The quadratic equations can be linearized and solved for both calibration and measurement (8,40).

The simplicity of the detection system is an advantage of the six-port device especially for wideband applications and very high operating frequencies. Compared to the vector network analyzer, the six-port device requires more calibration and more complicated mathematics.

Two six-port devices connected in the configuration shown in Fig. 29 are required to provide attenuation measurements.

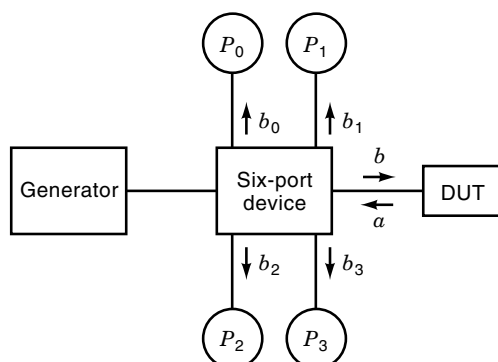


Figure 28. Principle of six-port technique for reflection measurement.

The dividing circuit includes phase adjustments to obtain different ratios b_1/b_2 at the terminals of the DUT. Using state-of-the-art power sensors the dynamic range of a dual six-port device is as large as 60 dB.

To achieve maximum accuracy, through connection, reflection line TRL-calibration is frequently used. Because the six-port device determines the complex parameters during the calibration process, the test ports appear well matched. The measurement uncertainties are primarily limited by the calibration standards, mainly the reflection standard (short or open). Real-time operation is limited by the computing time and the response time of the power sensors.

Fiber Optics

Three main methods (41,42) are used for attenuation measurements: the insertion loss technique, the cut-back technique, and the backscattering method. The first two methods perform two-point (end to end) measurements and the last performs a one-ended characterization. Some of the methods are standardized (43,44).

Insertion Loss Method. The insertion loss technique consists of a stable laser source and a stable, accurate, optical power meter. The power P_2 of the laser source is sent into the DUT (e.g., an optical fiber), and the power P_3 is measured at the far end. The attenuation is given by the ratio of the two power

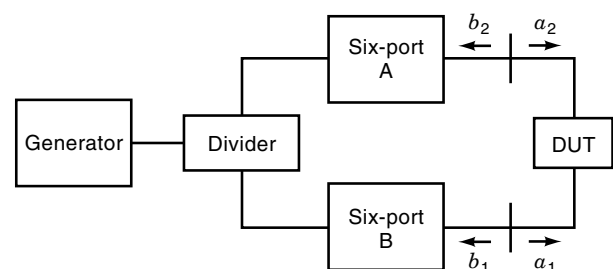


Figure 29. Principle of a dual six-port for s-parameter measurement.

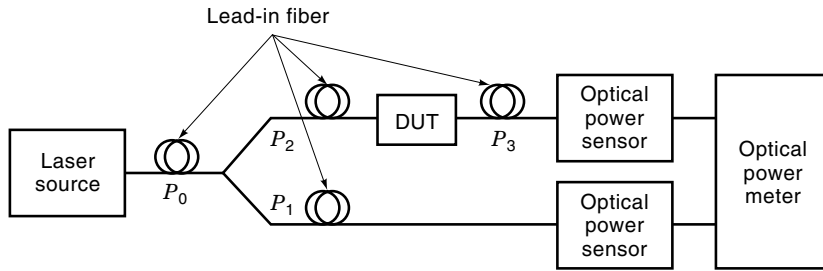


Figure 30. Principle of the insertion-loss method.

levels as

$$A(\text{dB}) = 10 \log \left(\frac{P_3}{P_2} \right) \quad (28)$$

To achieve more accurate measurements in the first phase, the power of the source is directly measured and is remeasured in the second phase with the DUT inserted. More sophisticated measuring systems use a configuration shown in Fig. 30.

A second power sensor measures the power level of the source instantaneously by a power divider. In this configuration the power stability of the source is less important because P_1 is always used as a reference. By using cooled detectors, a dynamic range up to 90 dB is achieved.

The accuracy of the measurements is determined by the following factors:

- the power level and wavelength stability of the source
- the calibration and stability of the power sensors
- the reproducibility of the connectors
- the linearity of the detectors

The measurement uncertainties for the insertion-loss technique are on the order of 0.9 dB including the connector reproducibility. Sophisticated systems reach over a limited dynamic range of 50 dB and uncertainty of 0.2 dB.

Cut-Back Method. The cut-back method (41,45) (Fig. 31) is the most accurate technique, but it is destructive. This method was developed to measure the attenuation of fibers as a function of the wavelength. Using a light source combined with a monochromator, a fiber can be tested at any wavelength from 800 nm to 1600 nm with a spectral width of 3 nm. The light from the source is projected into the fiber by a lens.

The power $P_2(\lambda)$ is measured at the far end of the fiber (test length l_t) by using a cooled detector. Then the fiber is cut back

to a short length of 2 m to 3 m without changing the projecting conditions, and the power $P_1(\lambda)$ is recorded. If the power loss in the short length of fiber is assumed to be negligible, the attenuation is given by the following equation:

$$A(\lambda) = 10 \log[P_2(\lambda)/P_1(\lambda)] \quad (29)$$

Assuming a uniform fiber, the attenuation coefficient per unit length of the fiber is given by

$$\alpha(\lambda) = \frac{A(\lambda)}{(l_t - l_r)} \quad (30)$$

where l_t and l_r are given in kilometers. The achieved uncertainty for cable length of several kilometers is about 0.02 dB/km for multimode fibers and 0.004 dB/km for single mode.

Backscattering Method. The backscattering method is a one-ended measurement based on Rayleigh scattering in an optical fiber (41,46). Figure 32 shows the principle of an optical time-domain reflectometer (OTDR).

A laser pulse is projected into the fiber by a coupler, and the backscattered power is measured. The backscattered power is related to the attenuation loss of the fiber and the measured time delay is related to the distance in the fiber. The attenuation is calculated by using two values of the backscattered power at different time delays (different locations along the fiber). The OTDR has the advantage of providing attenuation and reflection information along the fiber. A typical recording is shown in Fig. 33.

The length of the pulse is responsible for the dead zone where no measurement is possible. A lead-in fiber allows masking the dead zone. From the measured data details in the fiber path, such as connector loss or splice loss, irregularities of attenuation and defects are analyzed.

Commercially available OTDRs have a dynamic range of about 35 dB and cover distances up to 65 km depending on the

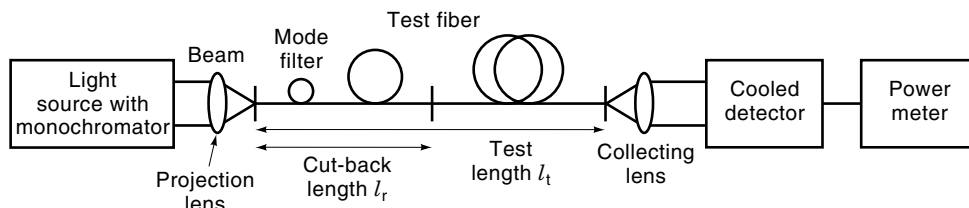


Figure 31. Principle of the cut-back method.

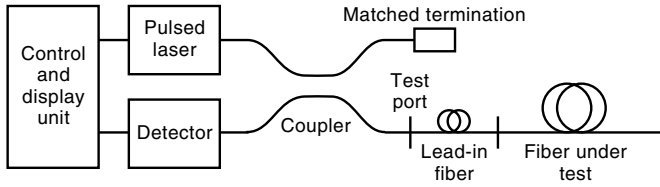


Figure 32. Principle of an optical time-domain reflectometer.

fiber loss. A well-calibrated OTDR can produce a measurement uncertainty of about 0.02 dB/km.

ERRORS AND UNCERTAINTIES IN ATTENUATION MEASUREMENTS

Whenever measurements are made, the results differ from the true or theoretically correct values. The differences are the result of errors in the measurement system, and it should be the aim to minimize these errors. In practice there are limits because no measurement instruments operate perfectly. A statement of measurement uncertainty reflects the quality of the measured results, and it has to be accompanied by a statement of confidence.

The International Committee for Weights and Measures (CIPM) (47) has published a guide for expressing uncertainty in measurements which has been adopted by the European Cooperation for Accreditation of Laboratories (EA) (48). According to the guide, uncertainty is grouped in two categories: Type A and Type B.

Type A evaluation is the result of statistical analysis of a series of repeated observations and therefore includes random effects.

Type B evaluation is by definition other than Type A, for example, judgment based on data of calibration certificates, experiences with instruments, and manufacturers' specifications.

Type A Evaluation of Uncertainty Components

Random effects result in errors that vary unpredictably. For an estimate of the standard deviation $s(q_k)$ of a series of n readings, q_k is obtained from

$$s(q_k) = \sqrt{\frac{1}{(n-1)} \sum_{k=1}^n (q_k - \bar{q})^2} \quad (31)$$

where $\bar{q}$ is the mean of n measurements.

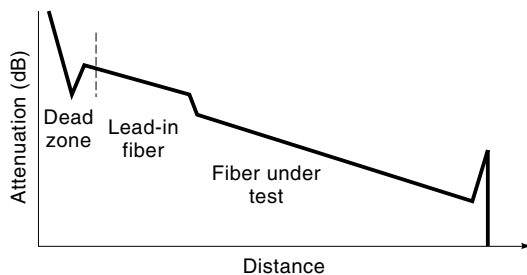


Figure 33. Typical backscattering signature of a fiber.

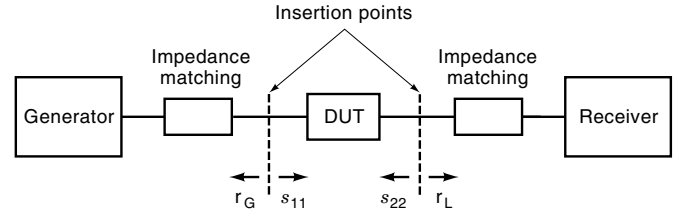


Figure 34. Example of an attenuation measurement system.

The random component of uncertainty is reduced by repeating the measurements. This yields the standard deviation of the mean $s(\bar{q})$

$$s(\bar{q}) = \frac{s(q_k)}{\sqrt{n}} \quad (32)$$

The standard uncertainty of the input estimate $\bar{q}$ is the experimental standard deviation of the mean (for $n \geq 10$)

$$u(\bar{q}) = s(\bar{q}) \quad (33)$$

Type B Evaluation of Uncertainty Components

Systematic effects that remain constant during measurements but change if the measurement conditions are altered cannot be corrected and therefore contribute to uncertainty. Other contributions arise from errors that are not possible or impractical to correct for, such as from calibration certificates or manufacturers' specifications. Most of these contributions are adequately represented by a symmetrical distribution. In RF metrology three main distributions are of interest: normal, rectangular, and U-shaped.

Normal Distribution. Uncertainties derived from multiple contributions are assumed to be normally distributed. Accredited calibration laboratories issue calibration certificates calculated for a normal distribution and a minimum level of confidence of 95% (approximate coverage factor $k = 2$). The standard uncertainty associated with the estimate x_i is given by

$$u(x_i) = \frac{\text{uncertainty}}{k} \quad (34)$$

Rectangular Distribution. This means that there is equal probability that the true value lies between limits. This is the case for most manufacturers' specifications that give a semi-range limit a_i :

$$u(x_i) = \frac{a_i}{\sqrt{3}} \quad (35)$$

U-Shaped Distribution. This distribution is applicable to mismatch uncertainty (49). Because the phases of the reflection coefficients (of source, DUT, load) in scalar measurement are not known, the mismatch loss has to be taken into account as an uncertainty. The mismatch uncertainty is asymmetrical to the measured result, and normally the larger of the two limits $M = 20 \log(1 - |\Gamma_G| |\Gamma_L|)$ is used. The standard uncertainty is

calculated as

$$u(x_i) = \frac{M}{\sqrt{2}} \quad (36)$$

Combined Standard Uncertainty

The combined standard uncertainty for uncorrelated input quantities is calculated as the square root of the sum of the squares of the individual standard uncertainties:

$$u_c(y) = \sqrt{\sum_{i=1}^m u_i^2(y)} \quad (37)$$

Expanded Uncertainty

The expanded uncertainty U defines an interval in which there is the true value with a specified confidence level. Normally accredited calibration laboratories are asked to use the coverage factor $k = 2$ (approximately 95% confidence level) giving:

$$U = k \cdot u_c(y)$$

Uncertainty in Attenuation Measurement

Let us assume a simple attenuation measuring setup, shown in Fig. 34, consisting of a generator, two matching circuits and a receiver. In the first phase when the two insertion points are clamped together, the receiver measures $P_1(f)$ (often called a normalization). In the second phase the DUT is inserted, and the receiver reads the values $P_2(f)$.

Attenuation as a function of the frequency is calculated from the ratio of the two sets of readings:

$$A(f) = 10 \log \left[\frac{P_2(f)}{P_1(f)} \right] \quad (38)$$

The following errors contribute to the uncertainty of the measurement:

The statistical errors of n repeated measurements (Type A) are given by the arithmetic experimental standard deviation:

$$s(A) = \sqrt{\frac{1}{(n-1)} \sum_{k=1}^n (A_k - \bar{A})^2} \quad (39)$$

($\bar{A}$ is the arithmetic mean of the measurements)

The standard uncertainty is calculated from

$$u_s(A) = s(\bar{A}) = \frac{s(A)}{\sqrt{n}} \quad (40)$$

The generator level stability a_G is taken from the manufacturer's specification and is assumed to be rectangularly distributed. The uncertainty is calculated as follows:

$$u_G = \frac{a_G}{\sqrt{3}} \quad (41)$$

The receiver level linearity and stability a_R is taken from the manufacturer's specification. The uncertainty is calculated as

$$u_R = \frac{a_R}{\sqrt{3}} \quad (42)$$

The noise level of the receiver influences the measurement of high attenuation values. It is given in the manufacturer's specification and contributes to the uncertainty as

$$u_N = \frac{a_N}{\sqrt{3}} \quad (43)$$

The cross talk of the measurement system a_I is determined by measurements and regarded as limits, and therefore contributes to the uncertainty as

$$u_I = \frac{a_I}{\sqrt{3}} \quad (44)$$

Two mismatch losses have to be taken into account, one during the normalization (often also called calibration) phase and the second while measuring the DUT.

Normalization Phase. The maximum mismatch loss (49) is calculated from the reflection coefficients of the source and the receiver as

$$M_C = 20 \log(1 - |r_G||r_L|) \quad (45)$$

As in scalar measurements, the phases of the reflection coefficients are unknown. The mismatch loss contributes to the measurement uncertainty and is normally assumed to be U-shaped distributed:

$$u_C = \frac{M_C}{\sqrt{2}} \quad (46)$$

Measurement Phase. There are two mismatch losses (49) that have to be considered: one between the generator and the input of the DUT and the other between the output of the DUT and the receiver. In addition, for small attenuation values the interaction between the input and the output connection has to be considered. The maximum limits of the mismatch loss which have to be used for the uncertainty are given by

$$M_m = 20 \log \frac{|1 - |r_G s_{11}| - |r_L s_{22}| - |r_G r_L s_{11} s_{22}| - |r_G r_L s_{21} s_{12}|}{1 - |r_G r_L|} \quad (47)$$

The uncertainty is given by

$$u_m = \frac{M_m}{\sqrt{2}} \quad (48)$$

The total uncertainty is calculated either from linear values or from decibel values as long as they are small:

$$u_c(A) = \sqrt{u_s^2 + u_G^2 + u_R^2 + u_N^2 + u_I^2 + u_C^2 + u_m^2} \quad (49)$$

The expanded uncertainty is calculated using a coverage factor $k = 2$ (approximately 95% confidence level) as

$$U(A) = k \cdot u_c(A) = 2 \cdot u_c(A) \quad (50)$$

The uncertainty has to be calculated for all of the measurement frequencies to find the maximum value of the uncertainty.

BIBLIOGRAPHY

1. D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*, New York: Pergamon, 1967.
2. R. W. Beatty, *Applications of Waveguide and Circuit Theory to the Development of Accurate Microwave Measurement Methods*, NBS Monograph 137, Washington, DC: US Government Printing Office, 1973.
3. R. W. Beatty, Insertion loss concepts, *Proc. IEEE*, **52**: 663, 1964.
4. F. L. Warner, *Microwave Attenuation Measurement*, IEE Monograph Series 19, London: Peregrinus, 1977.
5. S. J. Mason, Feedback theory-some properties of signal flow graphs, *Proc. IRE*, **41**: 1144–1156, 1953.
6. N. Kuhn, Simplified signal flow graph analysis, *Microw. J.*, (11): 59–66, 1963.
7. S. F. Adam, *Microwave Theory and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1969.
8. G. H. Bryant, *Principles of Microwave Measurements*, IEE Electrical Measurement Series 5, London: Peregrinus, 1993.
9. T. S. Laverghetta, *Modern Microwave Measurements and Techniques*, Norwood, MA: Artech House, 1998.
10. *Reference Data for Engineers*, 7th ed., Indianapolis, IN: H. W. Sams, 1985.
11. E. Weber, Precision Metalized Glass Attenuators, *Technique of Microwave Measurements*, MIT Radiation Lab Series, vol. 11, New York: McGraw-Hill, pp. 751–774.
12. S. F. Adam, Precision thin-film coaxial attenuators, *Hewlett-Packard J.*, 12–19, June 1967.
13. C. G. Montgomery, *Technique of Microwave Measurements*, MIT Radiation Lab Series Vol. 11, New York: McGraw-Hill.
14. W. Bolinger, private communication.
15. H. L. Kundsén, Champs dans un guide rectangulaire à membrane conductrice, *L'Onde Electrique*, April 1953.
16. B. P. Hand, A precision waveguide attenuator which obeys a mathematical law, *Hewlett Packard J.*, **6** (5): 1955.
17. G. C. Southworth, *Principles and Applications of Waveguide Transmission*, Van Nostrand: Princeton, NJ: 1950, pp. 374–376.
18. B. P. Hand, Broadband rotary vane attenuator, *Electronics*, **27**: 184–185, 1954.
19. P. F. Mariner, An absolute microwave attenuator, *Proc. IEE*, **109B**: 415–419, 1962.
20. T. Imanura, MN 9002A Standard optical attenuator, *Anritsu Techn. Rev.*, **14**: 32–41, 1991.
21. C. F. Varley, On a new method of testing electronic resistance, *Math. Phys. Sect. Br. Assoc. Adv. Sci.*, 14–15, 1866.
22. M. L. Morgan and J. C. Riley, Calibration of a Kelvin–Varley standard divider, *IRE Trans.*, **I-9** (1): 273–243, 1960.
23. A. F. Dunn, Calibration of a Kelvin–Varley voltage divider, *IEEE Trans. Instrum. Meas.*, **IM-3**: 129–139, 1964.
24. J. J. Hill and A. P. Miller, A seven-decade adjustable-ratio-inductively-coupled voltage divider with 0.1 part per million accuracy, *Proc. IEE*, **109**: 157–162, 1962.
25. S. Avramov et al., Automatic inductance voltage divider bridge for operation from 10 Hz to 100 kHz, *IEEE Trans. Instrum. Meas.*, **42**: 131–135, 1993.
26. R. Yell, NPL MK 3 WBCO attenuator, *IEEE Trans. Instrum. Meas.*, **IM-27**: 388–391, 1978.
27. R. Yell, Developments in waveguide below cutoff attenuators at NPL, *IEE Colloquium Dig.*, **49**: 1/1–1/5, 1981.
28. H. Bayer, Consideration of a rectangular waveguide below cutoff piston attenuator as a calculable broad-band attenuation standard between 1 MHz and 2.6 GHz, *IEEE Trans. Instrum. Meas.*, **IM-29**: 467–471, 1980.
29. F. L. Warner, D. O. Watton, and P. Herman, A very accurate X-band rotary vane attenuator with an absolute digital angular measurement system, *IEEE Trans. Instrum. Meas.*, **IM-21**: 446–450, 1972.
30. W. E. Little, W. Larson, and B. J. Kinder, Rotary vane attenuator with an optical readout, *J. Res. NBS*, **75C**: 1–5, 1971.
31. W. Larson, *The Rotary Vane Attenuator as an Interlaboratory Standard*, NBS Monograph 144, Washington, DC: US Government Printing Office, November 1975.
32. H. Bayer, F. Warner, and R. Yell, Attenuation and ratio-national standards, *Proc. IEEE*, **74**: 46–59, 1986.
33. T. Imamura, MN9002 Standard optical attenuator, *Anritsu Techn. Rev.*, **14**: 32–41, 1991.
34. G. F. Engen and R. W. Beatty, Microwave attenuation measurements with accuracies from 0.0001 to 0.06 dB over a range of 0.01 to 50 dB, *J. Res. NBS*, **64C**, 139–145, 1960.
35. H. Bayer, An error analysis for the RF-attenuation measuring equipment of the PTB applying the power method, *Metrologia*, **11**: 43–51, 1975.
36. D. L. Hollway and F. P. Kelly, A standard attenuator and the precise measurement of attenuation, *IEEE Trans. Instrum. Meas.*, **IM-13**: 33–44, 1964.
37. F. L. Warner, P. Herman, and P. Cumming, Recent improvements to the UK national microwave attenuation standards, *IEEE Trans. Instrum. Meas.*, **IM-32** (1): 33–37, 1983.
38. F. K. Weinert, High performance microwave ratio meter employs parallel if complex vector substitution, *Microw. J.*, **24**: 51–85, 1981.
39. Hewlett-Packard, *Understanding the Fundamental Principles of Vector Network Analysis*, HP Application Note 1287-1, 1997; *Exploring the Architectures of Network Analyzers*, HP Application Note 1287-2, 1997; *Applying Error Correction to Network Analyzer Measurement*, HP Application Note 1287-3, 1997.
40. G. F. Engen, Microwave circuit theory and foundations of microwave metrology, *IEE Electrical Measurement Series*, **9**, London: Perigrinus, 1992.
41. O. D. D. Soares (ed.), Trends in optical fibre metrology, *Part VI: Optical Fibre Characterisation, Calibration Standards*, Fibre Characterization and Measurement, p. 353–397; NATO ASI Series E, Applied Sciences, Dordrecht: Kluwer, 1995, Vol. 285.
42. C. Hentschel, *Fibre Optics Handbook*, Hewlett Packard, HP 13100-5952-9654, Boblingen, Germany: Hewlett Packard, 1989.
43. International Electrotechnical Commission, Optical fibres-Part 1: Generic specification IEC 793-1 (11/92).
44. International Telecommunications Union, Definition and test methods for the relevant parameters of single mode fibres, ITU-T G 650 (03/93).
45. T. Jones, Attenuation and cut-off wavelength measurement, National Physical Laboratory, Optical Fibre Measurement Course, April 27–28, 1993.

46. D. J. Ives and B. Walker, Optical time domain reflectometry, National Physical Laboratory, Optical Fibre Measurement Course, April 27–28, 1993.
47. International Organization for Standardization, Guide to the expression of uncertainty in measurement, 1st ed., Geneva, Switzerland: International Organization for Standardization, ISBN 92-67-10188-9. 1993, corrected and reprinted 1995.
48. European Cooperation for Accreditation of Laboratories, *Expression of the Uncertainty of Measurement in Calibration*, EAL-R2, Edition 1, April 1997; *Examples*, EAL-R2-S1, Edition 1, November 1997.
49. I. A. Harris and F. L. Warner, Re-examination of mismatch uncertainty when measuring microwave power and attenuation, *IEEE Proc., Pt. H*, **128** (1): 35–41, 1981.

KURT HILTY

Swiss Federal Office of Metrology

LOSS-ANGLE MEASUREMENT

Dielectrics, in the most general sense, may be considered as the broad class of *nonmetals* from the standpoint of their interaction with electric and magnetic fields. Hence gases, liquids, and solids can all be included, both with respect to their ability to store electric and magnetic field energy, as well as the accompanying dissipative processes occurring in the material when placed in an electric or magnetic field. How these phenomena can be described macroscopically and interpreted from the standpoint of molecular theory, how they can be accurately measured, and what the properties of various materials are provides the groundwork for their use in the rapidly growing microelectronic applications that shape much of our world today.

In a more narrow sense, dielectric materials may be classified from basic principles of the *energy-band model* as electrical insulators. The energy-band model forms the basis for development of all components in solid-state electronics. This model shows that the possible energies of electrons in a crystal are grouped in a certain number of allowed energy bands separated from one another by forbidden energy bands. The position of the Fermi energy, or energy of the topmost filled electron level in the sequence of allowed and forbidden bands, permits a unified treatment of metals, semiconductors, and insulators. According to the energy-band model, matter becomes dielectric (a poor conductor of electricity) when the conduction band and the valence band are separated by an energy gap higher than 5 eV. At normal temperatures only a small number of electrons have sufficient thermal energy necessary to make a transition to the conduction band. When temperature increases, the transition probability increases and the electrical conductivity σ can be written as

$$\sigma = \sigma_0 e^{-W/k_B T} \quad (1)$$

where W is the activation energy of the conduction process, k_B is Boltzmann's constant (1.38062×10^{-23} J/K), and T is the absolute temperature. In general, Eq. (1) is valid for crystalline and amorphous dielectrics, as well as partially crystalline dielectrics such as polymers.

Materials, whether in the solid, liquid, or gaseous states, may be electrically nonlinear, anisotropic, inhomogeneous, and dispersive both with respect to frequency and temperature.

Dissipation results from loss mechanisms that can differ in different types of materials. Because of all these complicating factors, both measurement technique and accuracy for evaluation of dielectric properties are requisite for physical understanding. Dielectric-loss-angle measurements reflect the different loss mechanisms occurring in a material placed in an electric field. This article addresses various measurement techniques for dielectric loss angle and permittivity evaluation of materials and is organized as follows:

- Electromagnetic characteristics of materials (constitutive equations, anisotropy, polar versus nonpolar materials, free charge versus bound charge, complex refractive index, polarization mechanisms, dispersion and relaxation processes)
- Permittivity and dielectric-loss-angle measurements (low-frequency complex impedance, free-space measurements for solids, liquids and gases, waveguide transmission and reflection techniques, resonance methods, and anisotropic material measurements)

ELECTROMAGNETIC CHARACTERISTICS OF MATERIALS

Physical Concepts Governing Electromagnetic Behavior

Any material is electromagnetically characterized by its permittivity ϵ (F/m), magnetic permeability μ (H/m), and electrical conductivity σ (S/m). Maxwell's equations, together with the constitutive equations relating field quantities in terms of material properties, completely govern electromagnetic wave propagation and behavior in that medium.

The constitutive equations for a linear, homogeneous, and isotropic medium may be expressed in the frequency domain as

$$\begin{aligned} \mathbf{B} &= \mu \mathbf{H} \\ \mathbf{J} &= \sigma \mathbf{E} \\ \mathbf{D} &= \epsilon \mathbf{E} \end{aligned} \quad (2)$$

where the magnetic induction $\mathbf{B}$ (Wb/m²) is related to the magnetic field $\mathbf{H}$ (A/m) by the magnetic permeability, the current density $\mathbf{J}$ (A/m²) is related to the electric field $\mathbf{E}$ (V/m) by the conductivity, and the dielectric displacement field $\mathbf{D}$ (C/m²) is related to the electric field by the permittivity. Any deviation from linearity is usually included by making ϵ , μ , or σ field dependent. For anisotropic media, ϵ , μ , or σ is a second-rank tensor as opposed to just a scalar function of frequency or simply a constant. For inhomogeneous media, ϵ , μ , or σ is a function of spatial coordinates. Material dielectrics that are linear, isotropic, and homogeneous when placed in an electric field at one frequency may not be isotropic or homogeneous when placed in an electric field at another frequency or under different temperature or pressure conditions. Similarly, dielectrics that behave linearly when placed in weak electric fields may not be linear in strong fields (or at high temperatures).

Anisotropy

When a dielectric is placed in an electric field $\mathbf{E}$, the material becomes polarized, and the dielectric displacement field is often

written

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (3)$$

where $\mathbf{P}$ is defined as the electric polarization of the material (dipole moment per unit volume), and is related to the electric field as

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad (4)$$

The proportionality constant χ is called the electric susceptibility, and the factor ϵ_0 (free-space permittivity equal to 8.854×10^{-12} F/m) is included in Eq. (3) to make χ dimensionless. Then Eq. (2) becomes

$$\mathbf{D} = \epsilon_0(1 + \chi) \mathbf{E} \quad (5)$$

or

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E} \quad (6)$$

where $\epsilon_r = 1 + \chi$ is called the complex permittivity of the medium relative to a vacuum. The presence of a dielectric always affects the ratio of $\mathbf{D}$ to $\mathbf{E}$ by a factor of ϵ_r . For *linear* materials the dipole moment induced in a dielectric by an external field $\mathbf{E}$ is directly proportional to $\mathbf{E}$. As long as the electric properties of the dielectric are independent of direction of the applied electric field, it is *isotropic*; that is, $\mathbf{P}$ and $\mathbf{E}$ are collinear. For an *anisotropic* material, however, the polarization (or charge separation) obtained when an electric field is applied along one coordinate axis will be different from that produced by the same field applied along a different coordinate axis. Quantitatively, this can be expressed by writing

$$\mathbf{P} = \epsilon_0 \bar{\bar{\chi}} \cdot \mathbf{E} \quad (7)$$

where $\bar{\bar{\chi}} = \chi_x \mathbf{i}\mathbf{i} + \chi_y \mathbf{j}\mathbf{j} + \chi_z \mathbf{k}\mathbf{k}$ and χ_x, χ_y, χ_z are the principal components of the electric susceptibility tensor expressed in dyadic form. For isotropic materials $\chi_x = \chi_y = \chi_z$, and Eq. (7) reduces to Eq. (4). Equation (7) shows that $\mathbf{P}$ and $\mathbf{E}$ are not collinear when $\chi_x \neq \chi_y = \chi_z$ or when $\chi_x = \chi_y \neq \chi_z$ or when $\chi_x \neq \chi_y \neq \chi_z$ (for two- or three-dimensional anisotropy), so that the electric susceptibility tensor may, in general, be viewed as an operation that takes a vector $\mathbf{E}$ and converts it into a new vector $\mathbf{P}$ that is not collinear with $\mathbf{E}$.

Polar versus Nonpolar Materials

Dielectric materials may also be divided into one of two categories: *polar* and *nonpolar*. A nonpolar material (such as inert or rare gases) is simply one that contains no (equivalent) dipoles (or separation of charge) when the material is not in an electric field. A polar material, on the other hand, possesses permanent polarization, even in the absence of an electric field, due to its molecular structure. Polar materials have permanent dipole moments at the microscopic or molecular level. A common polar molecule is the water molecule whose equivalent dipole moment results from a 104° angle between the two OH^- anions; this can be contrasted with the CO_2 molecule, in which the individual moments of each CO pair cancel, yielding a zero permanent moment.

In general, any distribution of charge may be described in terms of its multipole moments (1). The relevance of this discussion to dielectric material properties is that the existence of permanent dipole moments on the molecular level gives rise to a type of polarization mechanism when an electric field is applied that is frequency-dependent. Without an applied electric field in the case of the water molecule, the individual molecular dipole moments point in random directions, so that macroscopically their vector sum vanishes. In the presence of the applied electric field $\mathbf{E}$, though, there is a pronounced tendency of the dipoles to line up in the direction of $\mathbf{E}$, creating an *orientational* polarization whose magnitude can be computed and measured (2).

Ferroelectric materials are those in which there is spontaneous alignment of electric dipole moments at the molecular level. This occurs in ferroelectric materials at the Curie temperature. The permittivity of a ferroelectric material is field-strength dependent, which allows ferroelectric materials to be used in a variety of nonlinear devices, such as piezoelectric transducers, voltage-controlled oscillators, varactors, tunable filters, and phase shifters.

Complex Material Constituent Properties

The solution of Maxwell's equations yields all of the quantities that describe the propagation of electromagnetic waves in terms of the propagation constant $\sigma = jk$, where k is the complex wavenumber defined by

$$k^2 = \omega\mu(\omega\epsilon - j\sigma) \quad (8)$$

for $\exp(+j\omega t)$ time dependence for angular frequency ω and time t .

In general, the constituent electrical properties may be written as complex quantities; that is, for $\exp(+j\omega t)$ time dependence

$$\begin{aligned} \epsilon &= \epsilon' - j\epsilon'' = (\epsilon'_r - j\epsilon''_r)\epsilon_0 = \epsilon_r\epsilon_0 \\ \sigma &= \sigma' + j\sigma'' \\ \mu &= \mu' - j\mu'' = (\mu'_r - j\mu''_r)\mu_0 = \mu_r\mu_0 \end{aligned} \quad (9)$$

where μ_0 is the free-space permeability equal to $4\pi \times 10^{-7}$ H/m. Each component of ϵ , σ , or μ (which for anisotropic materials are tensor matrices) is, in general, a complex quantity. The imaginary part of the propagation constant contains all necessary information about energy loss in a material medium during wave propagation. If magnetic properties are ignored, we may consider only the complex form of ϵ and σ in Eq. (8):

$$\omega\epsilon - j\sigma = \omega(\epsilon' - j\epsilon'') - j(\sigma' + j\sigma'') = (\sigma'' + \omega\epsilon') - j(\sigma' + \omega\epsilon'') \quad (10)$$

Here $\omega\epsilon' + \sigma''$ may be considered an effective permittivity and $\sigma' + \omega\epsilon''$ as an effective conductivity. The term $\sigma' + j\sigma''$ physically represents carrier transport due to Ohmic and Faraday diffusion mechanisms, whereas $\epsilon' - j\epsilon''$ represents dielectric relaxation mechanisms. From Eq. (10), the loss tangent is simply defined as

$$\tan \delta = \tan\left(\psi + \frac{\pi}{2}\right) \equiv \frac{\sigma' + \omega\epsilon''}{\sigma'' + \omega\epsilon'} \quad (11)$$

where ψ is the phase between $\mathbf{E}$ and $\mathbf{J}$. If there are no dielectric losses, $\epsilon'' \rightarrow 0$. Similarly, if there are no Faraday losses, $\sigma'' \rightarrow 0$; hence

$$\tan \delta = \frac{\sigma'}{\omega\epsilon'} \quad (12)$$

which describes losses physically due to Ohmic conductivity.

Distinction between Ohmic Conductivity (σ') and Dielectric Loss Factor (ϵ'') or between Faraday Diffusion Transport (σ'') and In-Phase Polarization Phenomena (ϵ')

It is often stated that it is artificial to make distinctions between Ohmic carrier transport phenomena and dielectric loss characteristics of a material when that material is placed in a time-varying electric field. Actual dielectric measurements are indifferent to the underlying physical processes. To the extent, however, that physical and chemical processes are understood, distinctions can be made and materials designed to have certain electromagnetic characteristics.

The lack of distinction between two loss mechanisms from only measurement data can be seen by inspection of Eq. (11). Another way to see the relation between conductivity and the imaginary part of the permittivity is to write Ampere's law,

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (13)$$

where $\mathbf{J}$ is the conduction current in a material medium due to an applied electric field. For sinusoidal fields, Eq. (13) can be rewritten as

$$\begin{aligned} \nabla \times \mathbf{H} &= \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \\ &= (\sigma + j\omega\epsilon_0\epsilon_r) \mathbf{E} \\ &= j\omega \left[\epsilon_r' - j \left(\epsilon_r'' + \frac{\sigma}{\epsilon_0\omega} \right) \right] \epsilon_0 \mathbf{E} \end{aligned} \quad (14)$$

Hence, without other physical information, a conductivity σ is equivalent to an imaginary part of ϵ given by

$$\epsilon_r'' = \frac{\sigma}{\omega\epsilon_0} \quad (15)$$

so an equivalent form of Eq. (12) is

$$\tan \delta = \frac{\epsilon''}{\epsilon'} = \frac{(\text{loss current})}{(\text{charging current})} \quad (16)$$

Time-Domain Dielectric Constitutive Equation

The permittivity ϵ in Eq. (2) is defined for a time-harmonic electromagnetic phasor field at a certain frequency ω . If the permittivity is independent of frequency, the relationship in the time domain is simply expressed by

$$\mathbf{D}(\mathbf{r}, t) = \epsilon \mathbf{E}(\mathbf{r}, t) \quad (17)$$

However, if the permittivity is a function of frequency, the

temporal relationship is a Fourier transform of the product of $\epsilon(\omega)$ and $\mathbf{E}(\omega)$ and is therefore given by the convolution integral

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t h(t - \tau) \mathbf{E}(\mathbf{r}, \tau) d\tau \quad (18)$$

where

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \epsilon(\omega) e^{j\omega t} d\omega \quad (19)$$

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, \omega) e^{j\omega t} d\omega \quad (20)$$

The medium whose permittivity is a function of frequency $\epsilon(\omega)$ is called dielectrically dispersive. Although, strictly speaking, all media are dispersive, a medium can often be treated as nondispersive within a frequency range used for a particular problem.

If a medium is linear but varies with time, the relationship between $\mathbf{D}$ and $\mathbf{E}$ cannot be expressed as the convolution integral, Eq. (18). The general relationship should then be given by

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t h(t, t - \tau) \mathbf{E}(\mathbf{r}, \tau) d\tau \quad (21)$$

Substitution of Eq. (20) into Eq. (21) yields

$$\mathbf{D}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \epsilon(t, \omega) \mathbf{E}(\mathbf{r}, \omega) e^{j\omega t} d\omega \quad (22)$$

where $\epsilon(t, \omega)$ is the time-varying permittivity given by

$$\epsilon(t, \omega) = \int_0^{\infty} h(t, \tau) e^{j\omega\tau} d\tau \quad (23)$$

Time-varying media will not be discussed in this article.

Complex Refractive Index

For a nonabsorbing medium the refractive index is defined as c/u , where $c = 1/\sqrt{\mu_0\epsilon_0}$ is the velocity of propagation in free space and u is the phase velocity in the material ($c/\sqrt{\mu_r\epsilon_r}$). For a dielectrically lossy medium a *complex refractive index* $n - j\kappa$ is defined such that

$$n - j\kappa = \frac{c}{u} = \frac{ck}{\omega} \quad (24)$$

so that

$$n^2 - \kappa^2 = \epsilon_r' \mu_r' \quad (25)$$

and

$$2n\kappa = \frac{\sigma \mu_r'}{\epsilon_0 \omega} \quad (26)$$

where κ is the *absorption loss index*. The *absorption coefficient* α is related to the absorption loss index by $\alpha = 2\omega\kappa/c$. For a

nonmagnetic material, Eqs. (25) and (26) may be solved for the refractive index and absorption loss index,

$$n = \frac{1}{\sqrt{2}} \left[\sqrt{(\epsilon'_r)^2 + (\epsilon''_r)^2} + \epsilon'_r \right]^{1/2} \quad (27)$$

and

$$\kappa = \frac{1}{\sqrt{2}} \left[\sqrt{(\epsilon'_r)^2 + (\epsilon''_r)^2} - \epsilon'_r \right]^{1/2} \quad (28)$$

For small dissipation, where $\tan \delta \sim \delta$, $n \simeq \sqrt{\epsilon'_r}$, and $\kappa \simeq \delta \sqrt{\epsilon'_r}/2 = \epsilon''_r/(2\sqrt{\epsilon'_r})$.

Quality Factor of Dielectric

The *quality factor* Q of a dielectric at frequency f is defined as the inverse of the loss tangent,

$$\begin{aligned} Q &= \frac{1}{\tan \delta} = \frac{\epsilon'}{\epsilon''} = \frac{\omega \epsilon' E_0^2}{\omega \epsilon'' E_0^2} \\ &= 2\pi f \frac{\frac{1}{2} \epsilon' E_0^2}{\frac{1}{2} \sigma E_0^2} \\ &= 2\pi \frac{(\text{average energy stored per half cycle})}{(\text{energy dissipated per half cycle})} \end{aligned} \quad (29)$$

The quality factor Q is used as a descriptive characteristic of any dielectric material.

Polarization Mechanisms Intrinsic to Materials. A pulse or "signal" of any specified *initial* form can be constructed by superposition of harmonic wave trains of infinite length and duration. The velocities with which the constant-phase surfaces of these component waves are propagated depend on the propagation constant or on the parameters ϵ , μ , and σ . If the medium is nonconducting and the quantities ϵ and μ are independent of the frequency of the applied field, the phase velocity is constant and the signal is propagated without distortion. The presence of a loss mechanism, however, yields a functional relation between the frequency and phase velocity, as well as between frequency and attenuation. Hence in a lossy or absorptive medium the harmonic components suffer relative displacements in phase in the direction of propagation, and the signal arrives at a distant point in a modified form. The signal is *dispersed* and attenuated, and a medium in which the phase velocity is a function of frequency f (or in which the complex dielectric constant ϵ is a function of frequency) is said to be electrically dispersive.

The quantity $\epsilon'(f; x, y, z)$ is a measure of the polarization of the material. There can be a number of different polarizing mechanisms, each having a characteristic relaxation frequency and dielectric dispersion centered around this relaxation frequency. At the relaxation frequency there is maximal absorption. Figure 1 illustrates the dispersion of ϵ (and σ) that may be observed in materials in the frequency range 10^3 Hz to 10^{15} Hz. At the highest frequencies, the polarizing species in a material are the electrons. Electronic polarization occurs when an applied electric field causes a net displacement of the electron cloud of an atom with respect to its nucleus. At frequencies below about 10^{13} Hz, there is also a contribution from

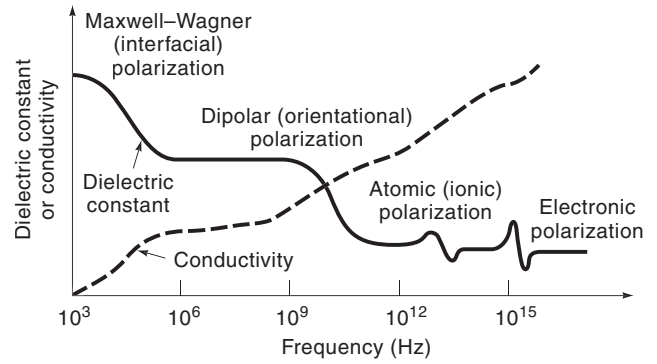


Figure 1. Dielectric dispersion for various types of polarization.

atomic polarization. Atomic polarization occurs in structures (molecules, solutions) in which atoms do not share electrons equally and electric fields displace the electron clouds preferentially towards the stronger binding atoms. It also occurs when charged atoms are displaced with respect to each other. Dipolar polarization, that is, the orientation of polar molecules (molecules with asymmetric charge distributions), occurs at frequencies below about 10^{10} Hz.

At frequencies below about 10^5 Hz, there are various types of charge polarization that may be collectively referred to as Maxwell-Wagner mechanisms (3,4). One of these, interfacial (space-charge) polarization, occurs when migrating charge carriers are trapped or impeded in their motion by local chemical or electric potentials, causing local accumulations of charge and a macroscopic field distortion. Another low-frequency mechanism that can occur is due to mixtures of materials having differing electrical properties (such as conducting spheres embedded in a dielectric). Several different equations are available to describe the resultant properties for various geometries of the embedded conductor (5-7). The common cause of these effects is the distributions of charge that occur at conductor-dielectric boundaries and the resultant action under applied electric fields that can yield very large low-frequency dielectric constants.

Still another dispersion mechanism for dielectric behavior at low frequencies, which is often distinguished from Maxwell-Wagner effects, is that which occurs in colloidal suspensions. Maxwell-Wagner effects occur when the charge around conducting particles in a dielectric medium is a thin coating that is much smaller than the particle dimensions; the charge responds to an applied electric field independent of the charge on nearby particles. In colloidal suspensions, on the other hand, the charge layer is on the same order of thickness or larger than the particle dimensions; hence it is affected by the charge distributions of adjacent particles. Colloidal polarization responses result in far higher low-frequency dielectric constants than those resulting from typical Maxwell-Wagner mechanisms, with dielectric constants on the order of 10^5 not uncommon.

Dispersion and Relaxation Processes in Materials. Polarization occurring in material media as a result of electromagnetic wave propagation is physically damped by either resonance or relaxation. Resonance is the state of a harmonic oscillator that is driven at its preferred frequency. Relaxation, on the

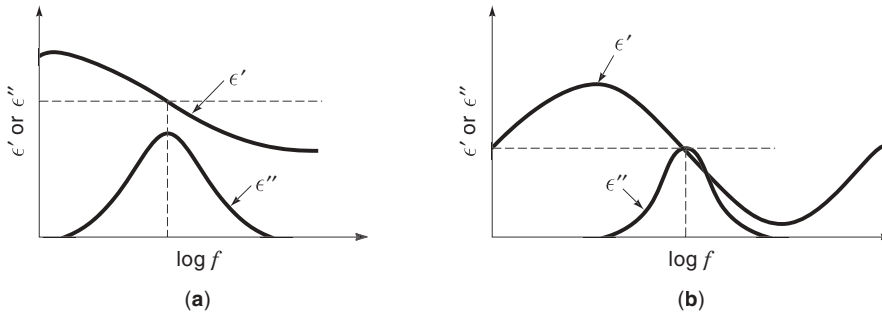


Figure 2. Two differing types of dispersion and absorption processes occurring in dielectrics as a function of frequency: (a) relaxation and (b) resonance.

other hand, is the state of a critically damped or overdamped oscillator. The characteristics of ϵ' and ϵ'' for these two differing types of dispersion and absorption processes are shown in Fig. 2.

At microwave frequencies, dipolar or orientation polarization phenomena principally occur. In this case, the frequency is sufficiently low so that the rotation of polar molecules has time to take place. At a frequency of $\omega = 1/\tau$, ϵ' decreases because the individual dipoles can no longer keep in step with the applied field. The relaxation time τ represents the time required for the dipoles to revert to a random distribution. This is a diffusion process that is represented by Fig. 2(a). Atomic and electronic polarization processes take place in the infrared and optical portion of the spectrum (1 THz and above) and lead to the resonance-type dispersion and absorption phenomenon represented by Fig. 2(b). A given medium may display any or all of these characteristic polarization phenomena, depending on its composition and molecular or atomic structure.

Relaxation processes are those observed in dielectric materials at microwave frequencies and below. Relaxation models are based on the general equation of charge motion,

$$\ddot{q} + (\mu\sigma)^{-1}\dot{q} + (\mu\epsilon)^{-1}q = 0 \quad (30)$$

where q is the charge and the overdot represents differentiation with respect to time. Another relaxation model is based on the diffusion of charged ions whose concentration is spatially variable. In this case

$$\begin{aligned} & \frac{\partial^2}{\partial x^2} \overline{Q}(t; x, y, z) + \frac{\partial^2}{\partial y^2} \overline{Q}(t; x, y, z) + \frac{\partial^2}{\partial z^2} \overline{Q}(t; x, y, z) \\ &= \frac{\partial^2 \overline{Q}(t; x, y, z)}{\partial t^2} + \frac{1}{K} \frac{\partial}{\partial t} \overline{Q}(t; x, y, z) + \frac{\eta}{K} \overline{Q}(t; x, y, z) \end{aligned} \quad (31)$$

where $\overline{CTQ}$, the concentration of charged ions, is a function of spatial coordinates and time. $K(t)$ is the diffusion coefficient and η is a constant. In the latter model, spatial derivatives must be taken in determining diffusion relaxation, which, in terms of electrical circuit analogs, lead to generalized distributed impedances (as opposed to lumped impedances) and nonlinear behavior.

Debye Relaxation

Materials having single relaxation time constants are called Debye materials. The complex permittivity in a Debye material

is given by (8–10)

$$\epsilon' - j\epsilon'' = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + \omega^2\tau^2} - j \frac{(\epsilon_s - \epsilon_\infty)\omega\tau}{1 + \omega^2\tau^2} \quad (32)$$

where τ is the relaxation time, ϵ_s is the relative dielectric constant at zero frequency ($\epsilon_{dc} = \epsilon_s\epsilon_0$) and ϵ_∞ is the relative dielectric permittivity at infinite frequency. In general, apart from liquid dielectrics, single relaxations are seldom observed. Multiple relaxations or distributions of relaxations are instead found.

Generalized Relaxation Distributions

A generalized expression for material media in which multiple relaxations are found may be written as (11)

$$\epsilon' - j\epsilon'' = \epsilon_\infty + (\epsilon_s - \epsilon_\infty) \int_0^\infty \frac{D(\tau)(1 - j\omega\tau)}{1 + \omega^2\tau^2} d\tau \quad (33)$$

where $D(\tau)$ is the time-constant distribution function, normalized such that

$$\int_0^\infty D(\tau) d\tau = 1 \quad (34)$$

One of the most commonly observed simple relaxation distributions in lossy media is the Cole–Cole distribution. In the Cole–Cole distribution Eq. (33) reduces to

$$\epsilon' - j\epsilon'' = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (j\omega\tau)^{1-m}} \quad (35)$$

where $0 \leq m \leq 1$. The loss tangent for the Cole–Cole distribution is

$$\tan \delta = \frac{\epsilon''}{\epsilon'} = \frac{\theta(\omega\tau)^{1-m} \sin \left[(1-m) \frac{\pi}{2} \right]}{1 + \theta + (2 + \theta)(\omega\tau)^{1-m} \cos \left[(1-m) \frac{\pi}{2} \right] + (\omega\tau)^{2(1-m)}} \quad (36)$$

where $\theta = (\epsilon_s - \epsilon_\infty)/\epsilon_\infty$. The $m = 0$ case corresponds to a Debye material (single relaxation). The $m = 1$ case corresponds to an infinitely broad continuous distribution (one having no relaxation). In the latter case the imaginary part of the complex permittivity disappears, and the real part becomes frequency independent.

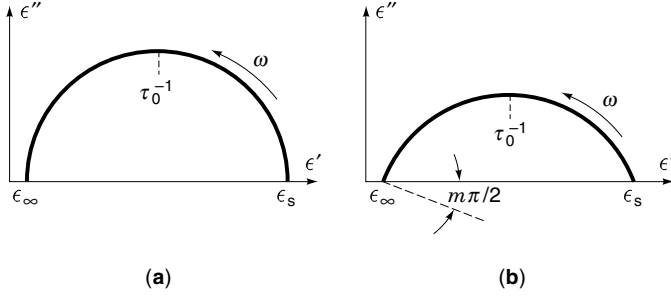


Figure 3. ϵ'' versus ϵ' plots for (a) Debye and (b) Cole–Cole materials.

The Cole–Cole distribution corresponds to a symmetric distribution of relaxation times of width m . Whereas a Debye material yields a plot of ϵ'' (ϵ') that is a semicircle whose center lies on the $\epsilon'' = 0$ axis, a Cole–Cole ϵ'' (ϵ') plot is a semicircle whose center lies below the horizontal $\epsilon'' = 0$ axis, on a line drawn from ($\epsilon' = \epsilon_\infty$, $\epsilon'' = 0$) that makes an angle of $m\pi/2$ with the horizontal axis. This is shown in Fig. 3. In addition to the Cole–Cole expression, there are other empirical relations commonly used to describe a non-Debye response. These are the Cole–Davidson (12), the combined Cole–Cole, and the Williams–Watkins (13) expressions. A characteristic feature of all these empirical relations, besides being based on Eq. (30), is that at frequencies away from the (dominant) relaxation frequency, they reduce to expressions showing a power-law dependence (14) on frequency for both ϵ' and ϵ'' .

Generalized Relation between Permittivity and Dielectric Loss Index

A generalized relation between ϵ' and ϵ'' for linear dielectric materials possessing an arbitrary number of relaxation times may be derived by regarding the permittivity as a system function characterizing the electrical properties of a material with the applied electric field as input and the displacement field as output. In the time domain the material permittivity is simply the transient (causal) system response, which can always be decomposed into the sum of an even and odd function whose Fourier transforms yield the (real) permittivity and (imaginary) dielectric loss index. The real permittivity and dielectric loss index are then related by the following Hilbert transforms, also known as the Kramers–Krönig relations,

$$\epsilon''(\omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\epsilon'(v)}{\omega - v} dv \quad (37)$$

and

$$\epsilon'(\omega) = \epsilon_\infty - \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\epsilon''(v)}{\omega - v} dv \quad (38)$$

where P denotes the Cauchy principal value. The application and limitations of Eqs. (37) and (38) for band-limited measurement data have been described in Ref. 11, as well as the use of an inverse power law of the distribution function for predicting expected changes in the dielectric loss tangent from measured changes in permittivity at two selected frequencies.

Effect of Temperature Changes

A classical statistical thermodynamic model using a double potential well was used (11) to describe the dispersive dielectric behavior for a bistable dielectric as a function of temperature and frequency in terms of the dipolar polarizability $\alpha_D = Np_E^2/k_B T$, the activation energy U , and the high-frequency (optical) permittivity at temperature T , where N is the total number of bistable dipoles in the material having dipole moment p_E . The results are

$$\epsilon'(\omega, T) = \epsilon_\infty(T) + \frac{\alpha_D}{1 + \omega^2 \tau^2} \quad (39)$$

and

$$\tan \delta(\omega, T) = \frac{\alpha_D \omega \tau}{\alpha_D + \epsilon_\infty(T)(1 + \omega^2 \tau^2)} \quad (40)$$

where $\tau = e^{-U/k_B T}/2A$ and A is a constant that may or may not depend on temperature describing the number of dipoles within the dielectric jumping per unit time from one potential energy state to a higher state. Equations (39) and (40) are limited to dielectric materials for which interaction between individual dipoles can be neglected and for conditions in which $p_E E \ll k_B T$ (nonsuperconducting states).

Langevin considered the electrostatic case of interacting molecules from a Maxwell–Boltzmann statistical ensemble average of the angular alignment with an applied electric field E of point dipoles having equal dipole moments in thermal equilibrium at temperature T . He derived the well-known Langevin function shown in Fig. 4,

$$\langle \cos \theta \rangle = \coth y - 1/y \quad (41)$$

where θ is the angle between field and dipole and y is $p_E E/(k_B T)$. The ensemble average $\langle \cos \theta \rangle$ increases with increasing y ; for high values of E/T , the orienting action of the electric field dominates over the disorienting action of the temperature. Implicit in the derivation of the Langevin function are the assumptions that the molecules are *point* dipoles that have *isotropic* polarizability, that ergodicity holds, and that the system obeys the classical Maxwell–Boltzmann statistics.

Additional discussion on relaxation models is to be found in the classical texts of Von Hippel (15) and Böttcher (16). The use of these physical relaxation models provides insight into what dispersive permittivity and dielectric loss tangents might be expected both as a function of temperature and frequency.

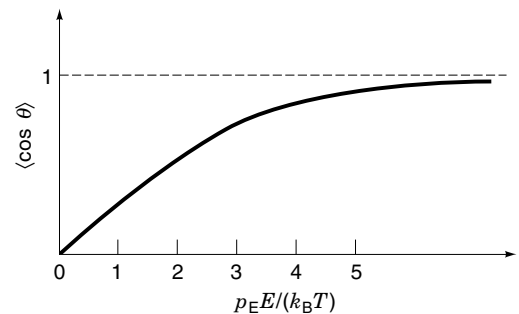


Figure 4. Behavior of Langevin function versus $p_E E/k_B T$.

However, their applicability and validity must be examined by accurate measurements.

PERMITTIVITY AND DIELECTRIC-LOSS-ANGLE MEASUREMENTS

Low-Frequency Complex Impedance Measurements

The use of a plane-parallel capacitor having a vacuum capacitance $C_0 = \epsilon_0 S/h$, where S and h are, respectively, the surface area of the electrode plates and separation of the plates, is commonly used for low-frequency dielectric measurements. If a low-frequency voltage $V = V_0 e^{j\omega t}$ is applied to this capacitor, a charge $Q = C_0 V$ appears on the electrodes that is in phase with the applied voltage. The nondissipative displacement current in the external circuit is then given by

$$I = \dot{Q} = j\omega C_0 V \quad (42)$$

which is 90° out of phase with the applied voltage. If the volume between the electrodes is filled with a lossless, nonpolar insulating material, the capacitor has a capacitance $C = \epsilon'_r C_0$. In this case the new displacement current is

$$I_{\text{diel}} = \dot{Q}_{\text{diel}} = j\omega C V = \epsilon'_r I \quad (43)$$

The capacitance is larger than the vacuum capacitance, but remains 90° out of phase with respect to the applied voltage. For lossy dielectric materials, the current is not 90° out of phase with the voltage since there is a small conduction GV due to charge motion in phase with the applied voltage. If the charges are free, the conductance G is independent of frequency. However, if the charges are bound, G is frequency dependent, and the dipole relaxation phenomena previously described become relevant. In general,

$$I = (j\omega C + G)V \quad (44)$$

where $G = \sigma S/h$ (if G is the conductance due to free charges) and $C = \epsilon'_r S/h$. Whenever dissipation is not exclusively due to free charges, but is also due to bound charges, the conductivity is itself a complex frequency-dependent quantity and a distinction cannot be made between Ohmic conductivity and dielectric loss factor or between Faradaic diffusion transport and in-phase polarization.

Free-Space Measurement

Free-space measurements of the complex permittivity and complex permeability usually involve placing a plate specimen orthogonal to the axis between the transmitting and receiving antennas. A plane electromagnetic wave is passed through the specimen. The complex permittivity or permeability can then be evaluated from measurements of the propagation constant $\gamma_s = jk_s = j\omega \sqrt{\mu_0 \epsilon_0 \epsilon_{r,s} \mu_{r,s}}$ of the plane electromagnetic wave propagating in the specimen or from the measured impedance $Z_s = \sqrt{\mu_0 \mu_{r,s} / \epsilon_0 \epsilon_{r,s}}$ of the specimen. The accuracy of free-space measurements depends on the appropriate choice of a theoretical model best representing the experimental measurement system and the accuracy of the measurement system.

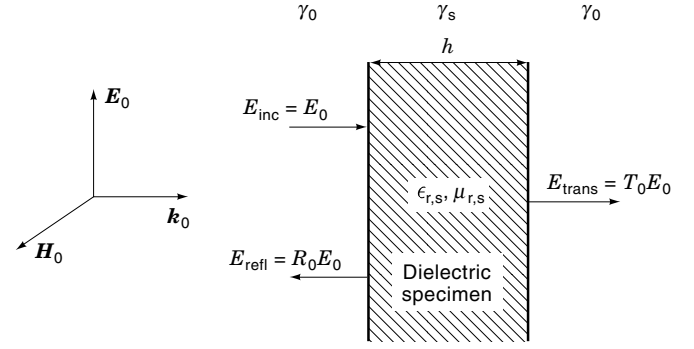


Figure 5. Reflection and transmission coefficients for an electromagnetic plane wave normally incident on an infinite dielectric plate specimen.

Solid Dielectric Specimens. For a normally incident transverse electromagnetic (TEM) wave on the specimen surrounded by air (see Fig. 5) the transmission and reflection coefficients, T_0 and R_0 , are given by

$$T_0 = \frac{4\gamma_s \gamma_0}{(\gamma_0 + \gamma_s)^2 e^{-\gamma_s h} - (\gamma_0 - \gamma_s)^2 e^{\gamma_s h}} \quad (45)$$

$$R_0 = \frac{(\gamma_0^2 - \gamma_s^2) e^{-\gamma_s h} - (\gamma_0^2 - \gamma_s^2) e^{\gamma_s h}}{(\gamma_0 + \gamma_s)^2 e^{-\gamma_s h} - (\gamma_0 - \gamma_s)^2 e^{\gamma_s h}} \quad (46)$$

where $\gamma_0 = j2\pi/\lambda_0$ and $\gamma_s = j2\pi \sqrt{\epsilon_{r,s} \mu_{r,s}}/\lambda_0$. Equations (45) and (46) may be solved for the complex permittivity and permeability of a magnetic plane-parallel plate. If the specimen is nonmagnetic, the transmission coefficient may be solved for ϵ_r . A common reflection technique for complex permittivity evaluation is to place a conducting plate (short) behind the specimen and measure the reflection coefficient. In this case,

$$R_{0|\text{short}} = \frac{(\gamma_0 - \gamma_s) e^{-\gamma_s h} - (\gamma_0 + \gamma_s) e^{\gamma_s h}}{(\gamma_0 + \gamma_s) e^{-\gamma_s h} - (\gamma_0 - \gamma_s) e^{\gamma_s h}} \quad (47)$$

Generally, complex permittivity evaluations are more accurate in reflection (one-port scattering parameter) measurements when the specimen is surrounded by air, whereas permeability evaluations are most accurate from reflection measurements when the specimen is backed by a shorting plane.

Liquid, Gas, or Plasma Specimens. The preceding free-space experimental setup is used mainly for measuring solid dielectric specimens that have plane-parallel plate geometries. For liquids, gases, or plasmas that must be held in containers whose walls are transparent to probing electromagnetic waves, the (two-dimensional) analysis becomes somewhat more complicated (see Fig. 6). In this case the amplitude reflection and transmission coefficients are given by (17),

$$R_0 = \frac{m \left(\frac{Z_s + Z_w}{Z_s - Z_w} + e^{2\gamma_w d} \frac{Z_w + Z_0}{Z_w - Z_0} \right) + n \left(1 + e^{2\gamma_w d} \frac{Z_s + Z_w}{Z_s - Z_w} \frac{Z_w + Z_0}{Z_w - Z_0} \right) e^{2\gamma_s h}}{m \left(\frac{Z_s + Z_w}{Z_s - Z_w} \frac{Z_w + Z_0}{Z_w - Z_0} + e^{2\gamma_w d} \right) + n \left(\frac{Z_w + Z_0}{Z_w - Z_0} + e^{2\gamma_w d} \frac{Z_s + Z_w}{Z_s - Z_w} \right) e^{2\gamma_s h}} \quad (48)$$

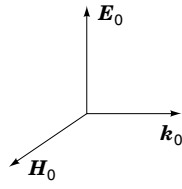


Figure 6. Free-space model for measuring liquids, gases, or plasmas in transparent container.

and

$$T_0 = \left\{ \cosh(\gamma_s h) \left[\cosh^2(\gamma_w d) + \sinh^2(\gamma_w d) - \frac{1}{2} \left(\frac{Z_0}{Z_w} + \frac{Z_w}{Z_0} \right) \sinh(2\gamma_w d) \right] + \frac{1}{2} \sinh(\gamma_s h) \left[\left(\frac{Z_s}{Z_w} + \frac{Z_w}{Z_s} \right) \sinh(2\gamma_w d) - \left(\frac{Z_s}{Z_0} + \frac{Z_0}{Z_s} \right) \cosh^2(\gamma_w d) - \left(\frac{Z_w^2}{Z_0 Z_s} + \frac{Z_0 Z_s}{Z_w^2} \right) \sinh^2(\gamma_w d) \right] \right\}^{-1} \quad (49)$$

where

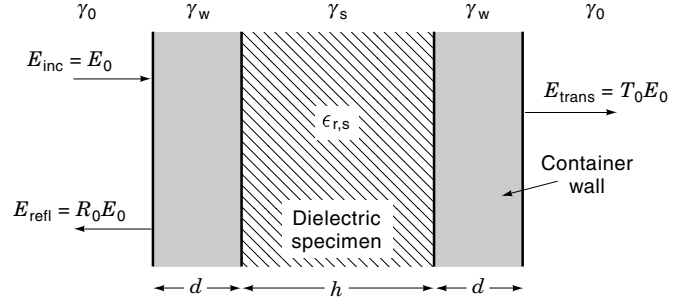
$$m = \left(\frac{Z_0}{Z_s} + 1 \right) \cosh(\gamma_w d) - \left(\frac{Z_0}{Z_w} + \frac{Z_w}{Z_s} \right) \sinh(\gamma_w d) \quad (50)$$

$$n = \left(\frac{Z_0}{Z_s} - 1 \right) \cosh(\gamma_w d) - \left(\frac{Z_0}{Z_w} - \frac{Z_w}{Z_s} \right) \sinh(\gamma_w d) \quad (51)$$

and Z_0 is the free-space impedance, $Z_w = Z_0/\sqrt{\epsilon_{r,w}} = \gamma_0 Z_0/\gamma_w$ is the container wall impedance, and $Z_s = Z_0/\sqrt{\mu_{r,s}\epsilon_{r,s}} = \gamma_0 Z_0/\gamma_s$ is the impedance of the medium under test.

Waveguide Transmission-Line Methods

The use of waveguide transmission and reflection techniques for evaluating complex permittivity and complex permeability has a long history, and the literature describing various techniques is extensive (18–37). Transmission-line techniques, usually made in rectangular or coaxial waveguides, are the simplest of the relatively accurate ways of measuring permeability and permittivity. Coaxial lines are broadband in the TEM dominant mode and therefore are attractive for spectral characterization of lossy magnetic materials, despite the problems of measurement uncertainty in complex permittivity determination introduced by potential air gaps between the sample and the coaxial line center conductor. Details of two-port, reference-plane invariant scattering parameter relations that can be used for determining permittivity and permeability are given elsewhere (37). One set of equations for dielectric and magnetic measurements of a single sample, in terms of two-port scattering parameters that can be taken with an automatic network



analyzer, is

$$S_{11}S_{22} - S_{21}S_{12} = \exp(-2\gamma_0(L_{\text{air}} - L)) \frac{R^2 - T^2}{1 - R^2T^2} \quad (52)$$

and

$$(S_{12} + S_{21})/2 = \exp[-\gamma_0(L_{\text{air}} - L)] \frac{T(1 - R^2)}{1 - R^2T^2} \quad (53)$$

where

$$R = \frac{\mu\gamma_0 - \mu_0\gamma}{\mu\gamma_0 + \mu_0\gamma} \quad (54)$$

$$T = \exp(-\gamma L) \quad (55)$$

$$\gamma_0 = \sqrt{\left(\frac{2\pi}{\lambda_c}\right)^2 - \left(\frac{\omega}{c_{\text{lab}}}\right)^2} \quad (56)$$

$$\gamma = \sqrt{\left(\frac{2\pi}{\lambda_c}\right)^2 - \frac{\omega^2\mu_r\epsilon_r}{c_{\text{vac}}^2}} \quad (57)$$

c_{vac} and c_{lab} are the speed of light in vacuum and laboratory, ω is the angular frequency, λ_c is the cutoff transmission-line wavelength, ϵ_r and μ_r are the specimen relative complex permittivity and permeability relative to vacuum, and L_{air} and L are air-line and specimen lengths. Equations (52) and (53) may be solved either explicitly or implicitly as a system of nonlinear scattering equations at each frequency or by using a nonlinear regression model over the entire frequency range. The total attenuation loss α_{TEM} of a sample under test for TEM mode structure in a coaxial transmission line is given by

$$\begin{aligned} \alpha_{\text{TEM}} &= \text{Re}(\gamma) \\ &= \text{Re} \left(j \frac{\omega}{c_{\text{vac}}} \sqrt{(\epsilon'_r - j\epsilon''_r)(\mu'_r - j\mu''_r)} \right) \\ &= \frac{\sqrt{2}\omega}{2c_{\text{vac}}} \sqrt{\epsilon'_r\mu'_r[(1 + \tan^2\delta_e)^{1/2}(1 + \tan^2\delta_m)^{1/2} + \tan\delta_e \tan\delta_m - 1]} \end{aligned} \quad (58)$$

where $\tan\delta_e = \epsilon''_r/\epsilon'_r$ and $\tan\delta_m = \mu''_r/\mu'_r$. One disadvantage of microwave measurements of the complex permittivity in waveguide is that specimens have very small tolerances in properly machined dimensions. If the specimen does not fill the entire cross section of the waveguide perfectly, corrections must be made for air gaps. For high-permittivity samples in either

rectangular or coaxial transmission lines, air gaps can lead to dielectric depolarization, which yields severe underestimates of actual specimen permittivity.

Thin Film Specimens

Techniques useful for the evaluation of the dielectric properties of thin film structures are important for various applications, such as dynamic random access memory (DRAM) cells and high-frequency bypass on-chip capacitors for monolithic microwave integrated circuits (MMICs). The dielectric properties of thin film structures, commonly 8 nm to 100 nm thick, on semiconductor substrates often differ appreciably from their bulk counterparts. For example, the real permittivities of thin films are usually smaller and dielectric losses significantly larger than those of corresponding bulk materials. These differences may be due to nonuniform strain arising from film-substrate lattice mismatches, film compositional inhomogeneities, film conductivity, or other structural imperfections.

Above 1 GHz, little is known about thin film capacitive characteristics. At these frequencies measurements become more difficult as a result of stray admittance effects around the thin film deposited structures. One technique outlined in (38–40) is based on an equivalent circuit model of the capacitive on-chip test structure. The equivalent circuit is expressed by lumped elements of intrinsic capacitor admittance Y_c , and parallel and series parasitic admittance Y_p and Y_s . The measured thin film test structure has an admittance Y_t which is corrected for Y_p and Y_s to obtain the intrinsic admittance of the thin film Y_c . The parasitic admittances Y_p and Y_s are evaluated by measuring the open- and short-test circuit device admittances, Y_{open} and Y_{short} . These admittances may be determined with microwave probe station scattering parameter reflection coefficient (S_{11}) measurements. The intrinsic admittance is calculated from

$$Y_c = (Y_t - Y_{open})(Y_{short} - Y_{open}) / (Y_{short} - Y_t) = G + j\omega C \quad (59)$$

where G and C are the conductance and capacitance of the intrinsic thin-film capacitor determined by the real and imaginary part of Y_c .

The dielectric loss tangent is given by

$$\tan \delta = G / (\omega C) \quad (60)$$

Resonance Methods

Resonant methods employing either closed and open cavities or dielectric resonators provide the highest measurement accuracy for evaluating complex permittivity and dielectric-loss tangent of low-loss materials at microwave frequencies (41–43). Generally, the (real) permittivity is calculated from the measured resonant frequency of one of the dominant modes of the resonant measurement system and the dimensions of the resonant structure. As long as specimen losses are low, they do not affect resonant frequencies. Exact relations between permittivity, sample dimensions, and resonant frequency exist only for simple rectangular, cylindrical, or spherical resonant structure geometries and when any permittivity inhomogeneity in the measurement fixture varies in only one of the principal coordinate directions. Resonant fixtures commonly used in practice for cylindrically shaped disk or rod specimens are shown

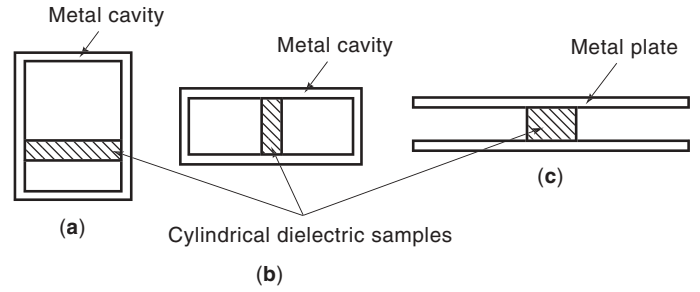


Figure 7. Typical cylindrical cavities and dielectric rod resonators used for complex permittivity measurements for which closed-form solutions are available: (a) TE_{01p} mode cavity, (b) TM_{010} mode cavity, (c) TE_{011} mode dielectric resonator.

in Fig. 7. An eigenvalue relationship, derived from Maxwell's equations and application of boundary conditions for the particular fixture of interest, always exists for a specific mode family that gives a transcendental equation relating permittivity, resonant frequency, and sample or fixture dimensions,

$$F(f_r, \epsilon_r', \text{dimensions}) = 0 \quad (61)$$

The permittivity is a numerical root of Eq. (61) for a given f_r and sample and fixture dimensions. Although there is more than one (mathematical) root to the eigenvalue equation, it is usually possible to pick the correct root, since many roots are non-physical or the permittivity is approximately known. It is also possible, in principle, to obtain a unique solution by using two independent measurements with different mode field configurations or by using two samples having different dimensions. The resonant fixtures shown in Fig. 7 may be practically used, when properly dimensioned, for complex permittivity evaluations of low- and medium-loss materials over the frequency range 1 GHz to 50 GHz.

Several resonant fixtures of practical use possess geometries for which analytical solutions are not available. These fixtures may be analyzed with numerical mode-matching, Rayleigh–Ritz, or finite-element methods. All these techniques allow accuracy improvements by incorporating more terms in field expansions or by mesh refinement. Mode-matching (44,45) is one of the most accurate methods for computation of resonant frequencies of axially symmetric resonant fixtures. In radial mode-matching, a resonant structure is first subdivided into cylindrical regions having dielectric inhomogeneity only along the axial direction. The electromagnetic field components are then separately expanded into series modal expansions in each region. Boundary conditions that are applied at the interfaces between adjacent regions yield a system of matrix equations (with respect to the field expansion coefficients) that have nonzero solutions only when the determinant vanishes. The resonant frequencies are the values that make the determinant of the square matrix vanish.

Generally, measurement uncertainties for (real) permittivity depend on

- Presence of air gaps between the specimen and conducting parts of the resonant structure that cause depolarization [some structures are not sensitive to air gaps, such as those in Figs. 7(a) and 7(c)]

- Computational inaccuracies
- Uncertainties from physical dimensions of specimen and resonant structure

Air gaps limit the measurement accuracy of high-permittivity solid materials when the electromagnetic field structure in the measurement system has an electric field component normal to the sample surface. In these cases, a discontinuity in the normal electric field leads to depolarization. When the applied electric field is continuous across a specimen boundary, such as with cylindrical samples in TE_{0np} or quasi- TE_{0np} ($TE_{0\delta v}$) mode resonant fixtures (45,46), high measurement accuracies are generally achieved. In the latter case, air gaps do not play a significant role. The depolarizing effects of air gaps can be mitigated by metallization of the sample surface contacting the fixture conductors. This added procedure improves real permittivity measurement accuracy, but can substantially degrade measurement accuracy of the dielectric loss factor. When the depolarizing effects of air gaps are either not important or have been mitigated, and the numerical method used for field analysis is sufficiently accurate, real permittivity measurement uncertainty depends only on the uncertainties associated with the physical dimensions of the sample under test and the measurement system.

Evaluation of the dielectric loss index or dielectric loss tangent is always based on the expression

$$Q_u^{-1} = p_{es} \tan \delta_s + p_{ed} \tan \delta_d + R_s/G + Q_r^{-1} \quad (62)$$

Equation (62) is valid for any resonant system containing isotropic dielectric materials. It defines the unloaded Q factor of the resonant system, Q_u , in terms of the sample partial electric energy filling factor p_{es} and sample dielectric loss tangent $\tan \delta_s$, the electric energy filling factors of the dielectric supports p_{ed} having relative real permittivity ϵ'_d and loss tangent $\tan \delta_d$ inside the resonant measurement system, the surface resistance R_s of any conducting shields, the geometrical factor G of the resonant system, and any radiation Q factor Q_r of the measurement fixture. The sample electric energy filling factor is defined as

$$p_{es} = \frac{W_{es}}{W_{et}} = \frac{\int_{V_s} \epsilon'_{r,s} \mathbf{E} \cdot \mathbf{E}^* dv}{\int_V \epsilon'_r(v) \mathbf{E} \cdot \mathbf{E}^* dv} \quad (63)$$

where W_{es} is the electric energy stored in the sample, W_{et} is the total electric energy stored in the resonant measurement fixture, $\epsilon'_{r,s}$ is the relative real permittivity of the sample, $\epsilon'_r(v)$ is the relative spatially dependent permittivity in the resonant structure, and the asterisk denotes complex conjugate. The electric energy filling factor for the dielectric support is given by

$$p_{ed} = \frac{W_{ed}}{W_{et}} = \frac{\int_{V_s} \epsilon'_{r,d} \mathbf{E} \cdot \mathbf{E}^* dv}{\int_V \epsilon'_r(v) \mathbf{E} \cdot \mathbf{E}^* dv} \quad (64)$$

where W_{ed} is the electric energy stored in the dielectric support. The geometrical factor is defined by

$$G = \frac{\omega \int_V \mu_0 \mathbf{H} \cdot \mathbf{H}^* dv}{\int_S \mathbf{H}_t \cdot \mathbf{H}_t^* dS} \quad (65)$$

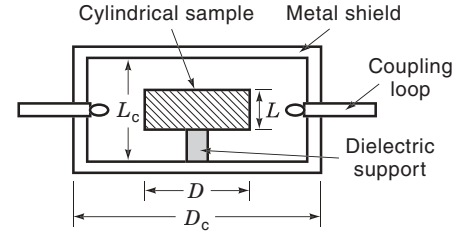


Figure 8. $TE_{01\delta}$ mode dielectric resonator measurement fixture.

where H_t is the magnetic field tangential to any conducting shield having surface area S .

For highest accuracy in dielectric-loss tangent measurements, the first term on the right-hand side of Eq. (62) must dominate, or the last three terms on the right-hand side of Eq. (62) should be minimized and well-characterized. Evaluation of conductor losses and those due to dielectric materials other than the sample under test within the measurement system (such as dielectric supports) often requires the rigorous use of numerical computation methods. The surface resistance of metal shields must also be well-characterized at the measurement frequency and temperature of interest. If an open resonator is used, radiation losses must be considered. For most resonant measurement fixtures used in practice, uncertainties in dielectric-loss-tangent evaluations are limited by conductor losses. Conductor losses decrease as the surface resistance becomes small and as the geometric factor increases. One common procedure to minimize conductor losses is to situate the dielectric specimen in a position away from the conductor walls as shown in Fig. 8 (47–52). Usually quasi- TE_{011} modes of this structure (often called $TE_{01\delta}$ modes) are used for dielectric-loss-tangent measurements. For this mode, geometric factors [approach] a maximum when dimensions of the metal shield increase. The optimal value of the geometric factor (optimal positioning of the specimen relative to metal shielding) depends on sample permittivity. If the distance of the metal shield from the specimen becomes greater than the optimum value, the electric energy filling factor of the sample decreases rapidly, and the field distribution becomes essentially the same as in an empty TE_{011} cavity. With optimal shield dimensions and a metal surface resistance equal to 30 mΩ (that of copper at 10 GHz and 297 K), the Q factor due to conductor losses can be large (2×10^5 for a sample having a relative real permittivity equal to 30). Sample dielectric losses should be at least 10% of the overall losses in Eq. (62) for accurate dielectric loss measurements. Hence the lower bound on the dielectric-loss-tangent measurement is about 5×10^{-7} (for sample permittivities greater than or equal to 30) when using a $TE_{01\delta}$ mode dielectric resonator with optimal shielding conditions. For this resolution, the geometric factor must be accurately known. Another way to minimize conductor losses is to employ high-temperature superconductors as the shielding material (53–55); however, this can only be done at cryogenic temperatures.

The most effective way to decrease conductor losses (or increase the geometric factor) for accurate dielectric-loss-tangent measurements is to use higher-order hybrid modes having high azimuthal mode numbers. These hybrid modes, called *whispering-gallery modes*, are excited in cylindrical specimens (56) as shown in Fig. 9. Conductor losses decrease very rapidly with increasing azimuthal mode numbers for

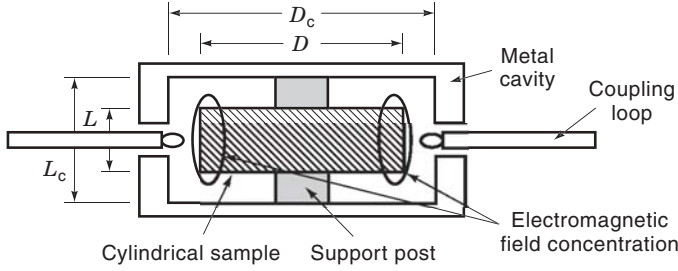


Figure 9. Whispering-gallery mode resonant structure.

whispering-gallery modes. Therefore they can be used for high-resolution dielectric-loss-tangent measurements of ultralow-loss materials. In addition, this method can be used for measurements of low-dielectric-loss materials up to frequencies of 100 GHz.

The most commonly used dielectric materials in the electronics industry are those used for printed wiring board substrates. Typically these dielectric materials exhibit losses in the range, $10^{-4} < \tan \delta < 10^{-2}$. Because printed wiring board substrates are relatively thin dielectric sheets, the resonant measurement techniques described previously are not applicable. For these samples the most convenient measurement technique is the split (tuned) dielectric post resonator illustrated in Fig. 10. This nondestructive method can be used practically at all frequencies from 1 GHz to 10 GHz.

When measurement frequencies lower than 1 GHz are of interest, dimensions of the resonant structures described thus far become impractically large. One of the commonly used resonant structures when dielectric measurements are required in the frequency range between 100 MHz and 1 GHz is the reentrant cavity, which is shown in Fig. 11. Advantages in the use of the reentrant cavity are relatively small sample size and, because of the axially directed electric field, a permittivity measurement normal to the substrate laminar surfaces. However, the reentrant cavity technique is sensitive to the presence of air gaps between the sample and the central metal posts, so that without metallization of the top and bottom surfaces of the specimen, it is only useful for measurements of low- to medium-permittivity materials.

At frequencies above 50 GHz, the dimensions of the resonant structures become impractically small. At these frequencies, a typical method (see Fig. 12) for complex permittivity measurements is the semiconfocal Fabry–Perot type resonator (56), although for very-low-loss materials, the whispering-gallery mode technique could also be used (57). Free-space transmission and reflection techniques previously described can also

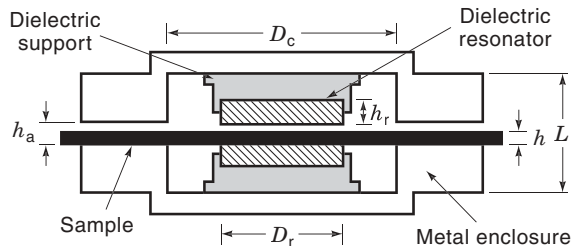


Figure 10. Split post-dielectric resonator.

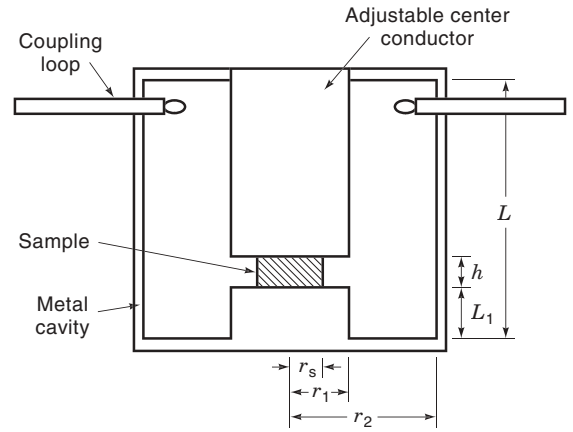


Figure 11. Doubly reentrant cavity.

be used to characterize materials from 50 GHz to visible-light frequencies.

Complex Permittivity Measurements of Anisotropic Materials.

Many materials, including some single crystals, exhibit dielectric anisotropy. For dielectrically anisotropic materials the complex permittivity is not independent of direction. The general (diagonalized) form of the electric susceptibility tensor is given in Eq. (7).

Anisotropy is usually related to the internal structure of the material. It can also be enforced by external factors, such as an external static magnetic field or mechanical stress. For anisotropic materials the relationship between the electric field intensity vector $\mathbf{E}$ and electric flux density vector $\mathbf{D}$ takes the form

$$\mathbf{D} = \bar{\epsilon} \cdot \mathbf{E} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \quad (66)$$

where the complex permittivity $\bar{\epsilon}$ is a second-rank tensor. The permittivity tensor may generally be expressed as a sum of symmetric and antisymmetric tensors

$$\bar{\epsilon} = \bar{\epsilon}^{(S)} + \bar{\epsilon}^{(A)} \quad (67)$$

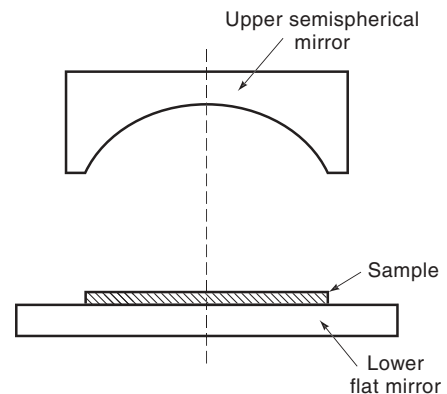


Figure 12. Fabry–Perot semiconfocal resonator.

For the symmetric tensor, $\epsilon^{(S)}_{ij} = \epsilon^{(S)}_{ji}$ and for the antisymmetric tensor, $\epsilon^{(A)}_{ij} = -\epsilon^{(A)}_{ji}$. The Hermitian and anti-Hermitian parts of the permittivity tensor are

$$\bar{\epsilon}' = \frac{1}{2}(\bar{\epsilon} + \bar{\epsilon}_t^*) \quad (68)$$

$$\bar{\epsilon}'' = \frac{1}{2}j(\bar{\epsilon} - \bar{\epsilon}_t^*) \quad (69)$$

where $\bar{\epsilon}_t$ represents the complex conjugate of the transposed tensor. The Hermitian part of the permittivity tensor is associated with electric energy stored in the dielectric while the anti-Hermitian part represents power dissipation in the dielectric. Most dielectric materials are described by a symmetric tensor only and dielectric losses by the imaginary components of a symmetric tensor. For a symmetric tensor there is always a coordinate system in which the permittivity tensor may take the diagonalized form,

$$\bar{\epsilon} = \begin{bmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{bmatrix} \quad (70)$$

When single crystals are measured, the coordinate axes are associated with the principal axes of the crystal lattice. When two of the three components of the permittivity tensor are equal, the material is said to be uniaxially anisotropic.

The easiest approach to measuring complex permittivity of anisotropic materials is to use measurement fixtures in which there is only one component of the electric field in the sample under test. For example, in the reentrant cavity and TM_{010} cylindrical cavity, there is only an axial electric field component in the specimen. In coaxial transmission line, only a radial electric field component is present when operating in dominant TEM mode field structure and in a TE_{01n} cavity, there is only an azimuthal electric field component in the specimen. If the specimen is properly oriented using one of these measurement techniques, measurement data can be obtained for evaluation of all tensor permittivity components.

For uniaxially anisotropic materials, a cylindrical specimen is first oriented so that the cylinder axis is parallel to the anisotropy axis of the material. Then a reentrant cavity can be used to evaluate the complex permittivity parallel

to the anisotropy axis. If the sample specimen is measured in a TE_{01n} mode cavity, the permittivity perpendicular to the anisotropy axis can be determined. For general (orthogonal) three-dimensional anisotropy, a rectangular waveguide sample may be fabricated so that the principal anisotropy axes are parallel to the Cartesian axes, or three cylindrical specimens can be fabricated the cylindrical axes of which are oriented parallel to each of the different anisotropy axes. In general, three independent measurements are required for evaluation of the permittivity tensor. This can be accomplished with two or three differently oriented samples or with the use of two or three different modes using only one sample. It is preferable, but not necessary, that only one permittivity tensor component be excited by the electric field structure in the measurement fixture. However, this is not always possible, since other factors, such as sensitivity and measurement accuracy, must also be considered.

Many single crystals have extremely low dielectric loss at low temperature (58–62). It is possible to evaluate dielectric losses in these extremely low-loss materials only by using resonant fixtures in which conductor losses are very small. For these measurements the whispering-gallery mode technique is appropriate. Whispering-gallery modes can always be classified into modes symmetric (S) or antisymmetric (N) to a plane of symmetry in a measurement fixture. In fact, since for these modes the electromagnetic fields are well confined to the dielectric specimen, this classification can still be made relative to the specimen's equatorial plane even if the resonant fixture does not have a symmetry plane. To evaluate the principal permittivity components of a uniaxially anisotropic dielectric resonator, a specimen whose cylindrical axis is along a principal direction of anisotropy is first obtained. Then two whispering-gallery mode resonances that exhibit a quasi-TE (H mode) and quasi-TM (E mode) field structure are identified and measured. Finally, a system of two nonlinear determinant equations are solved with respect to the two unknown permittivity tensor components,

$$\begin{aligned} F_1(f^{(H)}, \epsilon_{\perp}, \epsilon_{\parallel}) &= 0 \\ F_2(f^{(E)}, \epsilon_{\perp}, \epsilon_{\parallel}) &= 0 \end{aligned} \quad (71)$$

where $f^{(H)}$ and $f^{(E)}$ are the measured resonant frequencies for the quasi-TE (H) and the quasi-TM (E) whispering gallery

Table 1. Electric energy filling factors, geometric factors and measured loss tangents using several quasi-TM (N1) and quasi-TE (S2) modes of the sapphire resonator. Sapphire specimen, having diameter of 50.02 mm and height of 30.03 mm, in metallic shield having 80 mm diameter and 50 mm height. Computed permittivity tensor components: $\epsilon_{\perp} = 9.27$ and $\epsilon_{\parallel} = 11.35$ at 4.2 K and $\epsilon_{\perp} = 9.40$ and $\epsilon_{\parallel} = 11.59$ at 296.5 K.

Type	m	$p_{e\perp}$	$p_{e\parallel}$	$G(\Omega)$	tan δ	
					4.2 K	296.5 K
N1	11	0.0470	0.9341	6.77×10^6		5.0×10^{-6}
	12	0.0402	0.9423	1.63×10^7	5.0×10^{-9}	4.8×10^{-6}
	13	0.0350	0.9488	3.92×10^7	1.0×10^{-9}	4.8×10^{-6}
	14	0.0303	0.9538	9.42×10^7	7.0×10^{-10}	5.0×10^{-6}
S2	10	0.9548	0.0103	6.04×10^6	5.0×10^{-9}	9.0×10^{-6}
	11	0.9607	0.0064	1.60×10^7	4.0×10^{-9}	9.0×10^{-6}
	12	0.9620	0.0081	4.20×10^7	2.0×10^{-9}	7.0×10^{-6}
	13	0.9585	0.0106	1.10×10^8	9.0×10^{-10}	7.0×10^{-6}

modes and $\epsilon_{\perp}$ and $\epsilon_{\parallel}$ are the real parts of the permittivity tensor components normal and parallel to the anisotropy axis. The eigenvalue equations represented by F_1 and F_2 result from application of variational or mode-matching methods. Once the real permittivities are evaluated from Eq. (71), dielectric-loss tangents can be computed as solutions to

$$\begin{aligned} Q_E^{-1} &= p_{e\perp}^{(E)} \tan \delta_{\perp} + p_{e\parallel}^{(E)} \tan \delta_{\parallel} + R_s/G^{(E)} \\ Q_{(H)}^{-1} &= p_{e\perp}^{(H)} \tan \delta_{\perp} + p_{e\parallel}^{(H)} \tan \delta_{\parallel} + R_s/G^{(H)} \end{aligned} \quad (72)$$

where $\tan \delta_{\perp}$ and $\tan \delta_{\parallel}$ are the dielectric-loss tangents perpendicular and parallel to the anisotropy axis; $p_{e\perp}^{(H)}$, $p_{e\parallel}^{(H)}$, $p_{e\perp}^{(E)}$, $p_{e\parallel}^{(E)}$ are the electric energy filling factors perpendicular and parallel to the anisotropy axis of the dielectric resonator for quasi-TM whispering-gallery modes (superscript E) and quasi-TE whispering gallery modes (superscript H); and $G^{(E)}$ and $G^{(H)}$ are the geometric factors for quasi-TM and quasi-TE whispering-gallery modes.

Table 1 shows permittivity and dielectric-loss-tangent-measurement results for single-crystal sapphire at room temperature and at 4.2 K using whispering-gallery modes. The electric energy filling factors in Table 1 show that both N1 and S2 mode families can be treated as quasi-TM and quasi-TE modes. The geometric factor calculations demonstrate that conductor losses are considerably smaller than dielectric losses for large azimuthal mode numbers and can be neglected in most cases. Hence the dielectric-loss tangents can be approximated as the reciprocals of the unloaded Q factors for quasi-TM and quasi-TE whispering-gallery modes. Permittivities were evaluated accounting for thermal expansion coefficients (63,64). At 4.2 K, conductor losses for the copper metal shield used in experiments can be neglected only for azimuthal mode numbers $m > 13$. Conductor losses for whispering-gallery mode dielectric resonators can be made negligible by either performing measurements on higher-order azimuthal modes or by simply enlarging the enclosing metal shield.

BIBLIOGRAPHY

1. J. D. Jackson, *Classical Electrodynamics*, New York: Wiley, 1975.
2. A. Nussbaum, *Electromagnetic and Quantum Properties of Materials*, Englewood Cliffs, NJ: Prentice-Hall, 1966.
3. J. C. Maxwell, *A Treatise on Electricity and Magnetism*, New York: Dover, 1891.
4. K. W. Wagner, Erklärung der dielektrischen Nachwirkungswerte auf Grund Maxwellscher, *Arch. Electrotechnik*, **20**: 371, 1914.
5. R. W. P. King and G. S. Smith, *Antennas in Matter*, Cambridge, MA: MIT Press, 1981.
6. S. S. Dukhin, Dielectric properties of disperse systems, in E. Matijevic (ed.), *Surface and Colloid Science*, Vol. 3, New York: Wiley-Interscience, 1969.
7. R. G. Geyer, J. Mantese, and J. Baker-Jarvis, Effective medium theory for ferrite-loaded materials, *Natl. Inst. Stand. Technol. Tech. Note 1371*, 1994.
8. J. B. Hasted, *Aqueous Dielectrics*, London: Chapman & Hall, 1973.
9. P. Debye, *Polar Molecules*, New York: Chemical Catalog, 1929.
10. C. P. Smyth, *Dielectric Relaxation and Molecular Correlation in Dielectric and Related Molecular Processes*, London: Chemical Soc., 1966.
11. R. G. Geyer, Dielectric characterization and reference materials, *Natl. Inst. Stand. Technol. Tech. Note 1338*, 1990.
12. K. S. Cole and R. H. Cole, Dispersion and absorption in dielectrics, *J. Chem. Phys.*, **9**: 341–351, 1941.
13. G. Williams and D. C. Watts, Non-symmetrical dielectric relaxation behavior arising from a simple empirical decay function, *Trans. Faraday Soc.*, **66**: 80–85, 1970.
14. A. K. Jonscher, *The universal dielectric response, a review of data and their new interpretation*, Chelsea Dielectric Group, Univ. London, 1979.
15. A. Von Hippel, *Dielectrics and Waves*, New York: Wiley, 1954.
16. C. J. Böttcher, *Theory of Electric Polarization*, Vols. 1 and 2, New York: Elsevier, 1978.
17. J. Musil and F. Zacek, *Microwave Measurements of Complex Permittivity by Free Space Methods and Their Applications*, New York: Elsevier, 1986.
18. S. Roberts and A. Von Hippel, A new method for measuring dielectric constant and loss in the range of centimeter waves, *J. Appl. Phys.*, **7**: 610–616, 1946.
19. N. Marcuvitz, *Waveguide Handbook*, New York: Dover, 1951.
20. G. A. Deschamps, Determination of reflection coefficients and insertion loss of a waveguide junction, *J. Appl. Phys.*, **2**: 1046–1050, 1953.
21. D. M. Bowie and K. S. Kelleher, Rapid measurement of dielectric constant and loss tangent, *IEEE Trans. Microwave Theory Tech.*, **MTT-4**: 137–140, 1956.
22. M. Sucher and J. Fox (eds.), *Handbook of Microwave Measurements*, Polytechnic Inst. Brooklyn Series, New York: Wiley, 1963.
23. H. E. Bussey and J. E. Gray, Measurement and standardization of dielectric samples, *IRE Trans. Instrum.*, **I-11**: 162–165, 1962.
24. G. M. Brydon and D. J. Hepplestone, Microwave measurements of permittivity and $\tan \delta$ over the temperature range 20–700°C, *Proc. Inst. Elec. Eng.*, **112**: 421–425, 1965.
25. G. Franceschetti, A complete analysis of the reflection and transmission methods for measuring the complex permeability and permittivity of materials at microwave frequencies, *Alta Frequenzia*, **36**: 757–764, 1967.
26. A. M. Nicolson and G. F. Ross, Measurement of the intrinsic properties of materials by time domain techniques, *IEEE Trans. Instrum. Meas.*, **IM-19**: 377–382, 1970.
27. W. B. Weir, Automatic measurement of complex dielectric constant and permeability at microwave frequencies, *Proc. IEEE*, **62**: 33–36, 1974.
28. S. Stuchly and M. Matuszewski, A combined total reflection transmission method in application to dielectric spectroscopy, *IEEE Trans. Instrum. Meas.*, **IM-27**: 285–288, 1978.
29. M. S. Freeman, R. N. Nottenburg, and J. B. DuBow, An automated frequency domain technique for dielectric spectroscopy of materials, *J. Phys. E*, **12**: 899–903, 1979.
30. L. P. Ligthardt, A fast computational technique for accurate permittivity determination using transmission line methods, *IEEE Trans. Microw. Theory Tech.*, **MTT-31**: 249–254, 1983.
31. Hewlett Packard Product Note 8510-3, *Measuring dielectric constant with the HP 8510 network analyzer*, Palo Alto, CA: Hewlett-Packard, 1985.
32. L. Solymar and D. Walsh, *Lectures on Electrical Properties of Materials*, London: Oxford Univ. Press, 1988.
33. N. Belhadj-Tahar, A. Fourier-Lamer, and H. de Chanterac, Broadband simultaneous measurement of complex permittivity and permeability using a coaxial discontinuity, *IEEE Trans. Microw. Theory Tech.*, **2**: 1–7, 1990.
34. K. E. Mattar and M. E. Brodwin, A variable frequency method

- for wide-band microwave material characterization, *IEEE Trans. Instrum. Meas.*, **39**: 609–614, 1990.
35. H. B. Sequeira, Extracting μ_r and ϵ_r from one-port phasor network analyzer measurements, *IEEE Trans. Instrum. Meas.*, **39**: 621–627, 1990.
 36. G. Maze, J. L. Bonnefoy, and M. Kamarei, Microwave measurement of the dielectric constant using a sliding short-circuited waveguide method, *Microw. J.*, **33** (10): 77–88, 1990.
 37. J. Baker-Jarvis et al., Transmission/reflection and short-circuit line methods for measuring permittivity and permeability, *Natl. Inst. Stand. Technol. Note 1355-R*, 1993.
 38. K. Ikuta, Y. Umeda, and Y. Ishii, Measurement of high-frequency characteristics in the mm-wave band for dielectric thin films on semiconductor substrates, *Jpn. J. Appl. Phys.*, **34**, part 2 (9B): 1211–1213, 1995.
 39. J. D. Banieki et al., Dielectric relaxation of $\text{Ba}_{0.7}\text{Sr}_{0.3}\text{TiO}_3$ thin films from 1 MHz to 20 GHz, *Appl. Phys. Lett.*, **72** (4), 498–500, 1998.
 40. W. Williamson III et al., High frequency dielectric properties of PLZT thin films, *Integrated Ferroelectrics*, **17**, 197–203, 1997.
 41. E. J. Vanzura, R. G. Geyer, and M. D. Janezic, The NIST 60-millimeter cylindrical cavity resonator: performance evaluation for permittivity measurements, *Natl. Inst. Stand. Technol. Note 1354*, 1993.
 42. R. G. Geyer, L. Sengupta, and J. Krupka, Microwave properties of composite ceramic phase shifter materials, *IEEE Proc. 10th Int. Symp. Appl. Ferroelectrics*, 1996, pp. 851–854.
 43. R. G. Geyer, C. Jones, and J. Krupka, Microwave characterization of dielectric ceramics for wireless communications, in *Advances in Dielectric Ceramic Materials*, Am. Ceram. Soc. Trans., **88**: 75–91, 1998.
 44. Sz. Maj and M. Pospieszalski, A composite multilayered cylindrical dielectric resonator, *IEEE MTT-S Int. Microw. Symp. Dig.*, 1984, pp. 190–192.
 45. D. Kajfez and P. Guillon, *Dielectric Resonators*, Chap. 6, Dedham, MA: Artech House, 1986.
 46. W. E. Courtney, Analysis and evaluation of a method of measuring the complex permittivity and permeability of microwave insulators, *IEEE Trans. Microw. Theory Tech.*, **18**: 476–485, 1970.
 47. Y. Kobayashi, N. Fukuoka, and S. Yoshida, Resonant modes for a shielded dielectric rod resonator, *Electron. and Commun. (Japanese)*, **64-B**: 46–51, 1981.
 48. Y. Kobayashi, Y. Aoki, and Y. Kabe, Influence of conductor shields on the Q-factors of a TE_0 dielectric resonator, *IEEE MTT-S Int. Microw. Symp. Dig.*, 1985, pp. 281–284.
 49. J. Krupka, Resonant modes in shielded cylindrical and single-crystal dielectric resonators, *IEEE Trans. Microw. Theory Tech.*, **37**, 691–697, 1989.
 50. *Dielectric Resonators—A Designer Guide to Microwave Dielectric Ceramics*, Trans-Tech Inc. Pub. No. 50080040, Rev. 2, 1990.
 51. Y. Kobayashi and T. Senju, Resonant modes in shielded uniaxial-anisotropic dielectric rod resonator, *IEEE Trans. Microw. Theory Tech.*, **41**: 2198–2205, 1993.
 52. J. Krupka and A. Kedzior, Optimization of the complex permittivity measurement of low loss dielectrics in a cylindrical TE_{01n} mode cavities, *Electron Technol.*, **14**: 67–79, 1981.
 53. J. Krupka et al., Dielectric properties of single crystals of Al_2O_3 , LaAlO_3 , NdGaO_3 , SrTiO_3 , and MgO at cryogenic temperatures, *IEEE Trans. Microw. Theory Tech.*, **42**: 1886–1890, 1993.
 54. R. G. Geyer and J. Krupka, Microwave dielectric properties of anisotropic materials at cryogenic temperatures, *IEEE Trans. Instrum. Meas.*, **44**: 329–331, 1995.
 55. H. Takamura, H. Matsumoto, and K. Wakino, Low temperature properties of microwave dielectrics, *Proc. 7th Meeting Ferroelectric Materials Their Applications, Jpn. J. Appl. Phys., Suppl.* 28-2, **28**: 21–23, 1989.
 56. A. L. Cullen and P. K. Yu, The accurate measurement of permittivity by means of an open resonator, *Proc. R. Soc. London, Ser. A*, **325**: 493–509, 1971.
 57. J. Krupka et al., Study of whispering gallery modes in anisotropic single-crystal dielectric resonators, *IEEE Trans. Microw. Theory Tech.*, **42**: 56–61, 1994.
 58. V. Braginsky, V. S. Ilchenko, and Kh. S. Bagdassarov, Experimental observation of fundamental microwave absorption in high quality dielectric crystals, *Phys. Lett. A*, **120**: 300–305, 1987.
 59. R. Shelby and J. Fontanella, The low temperature electrical properties of some anisotropic crystals, *J. Phys. Chem. Solids*, **41**: 69–74, 1980.
 60. M. E. Tobar and A. G. Mann, Resonant frequencies of higher order modes in cylindrical anisotropic resonators, *IEEE Trans. Microw. Theory Tech.*, **39**: 2077–2083, 1991.
 61. A. N. Luiten, A. G. Mann, and D. G. Blair, Paramagnetic susceptibility and permittivity measurements at microwave frequencies in cryogenic sapphire resonators, *J. Phys. D*, **29**: 2082–2090, 1996.
 62. R. G. Geyer, J. Krupka, and M. Tobar, Microwave dielectric properties of low-loss materials at low temperature, *Proc. Mater. Res. Soc. Hybrid Mater.*, 1998.
 63. C. A. Swenson, R. B. Roberts, and G. K. White, Thermophysical properties of some key solids, in G. K. White and M. L. Mingos (eds.), *CODATA Bulletin 59*, Chap. 4, New York: Oxford Pergamon, 1985.
 64. G. K. White, Reference materials for thermal expansion: Certified or not, *Thermochim. Acta*, **218**: 83–99, 1993.

JERZY KRUPKA
Instytut Mikroelektroniki i
Optoelektroniki Politechniki
Warszawskiej

RICHARD G. GEYER
National Institute of Standards and
Technology

WAVELENGTH METER

BASIC INFORMATION

Definition of Wavelength for a Monochromatic Source

Wavelength meters are instruments designed to measure the wavelength of an electromagnetic wave. Electromagnetic waves are the classical way to represent modifications of the space surrounding a moving electrically charged body. Time-varying electric and magnetic fields are produced and can be sensed. Wavelength is a well-established parameter when considering a monochromatic (single-frequency) plane wave propagating in vacuum. It is the distance between two planes, measured along the direction of propagation of energy, where the phase and its derivatives have the same values. For the purpose of electromagnetic wave propagation, vacuum is a homogeneous, linear, isotropic, and nondispersive medium. This implies that the reaction of the medium to the electromagnetic wave's presence is instantaneous and does not depend on the location of the observation or on the amplitude and direction of propagation of the wave.

Nonconducting materials, like gases, ceramics, or polymers, are called dielectrics. They have electric charges but the charges are not allowed to move freely within the medium.

When an electromagnetic wave moves through such a medium, it stimulates the charges to oscillate around their equilibrium positions. These oscillating charges radiate an electromagnetic field that adds up to the impinging field. The resulting field is the field propagating in the medium. The reaction of the medium will then change the conditions of propagation when compared to those of vacuum. In general, a real medium is nonhomogeneous, nonlinear, nonisotropic, and dispersive. This implies that the reaction will be dependent on location in the medium and the amplitude and direction of propagation of the field, that it will not be instantaneous, and that some of the wave energy will be transferred to other spectrum bands (harmonic generation). In a first approximation, it is useful to reduce the complexity of the medium reaction to something close to that of the vacuum, taking only into account a reduction of the propagation speed associated with the summing action. The speed of the wave in a perfect dielectric material is less than that in vacuum, and the ratio of the speed value in vacuum to the speed value in the medium is the index of refraction. It is then possible to reconsider the propagation of infinite plane waves and use the same definition for the wavelength. Since the speed of propagation is less in the medium than in vacuum, the distance between the planes will be less, which implies that the wavelength in the material is less than that for vacuum. Consequently, the wavelength value is dependent on the medium in which the wave is propagating.

Wavelength or Frequency?

Wavelength is measured along the space (position) axis. But if we measure the field associated with a monochromatic electromagnetic wave at a fixed position, we observe that this field oscillates regularly with time. The rate of this oscillation per unit of time is the frequency. The frequency is then measured along the time axis. The wavelength (λ) times the frequency (ν) is equal to the speed of light in the medium (v).

The two measurements, wavelength and frequency, are dual quantities. They reveal two aspects of the same physical interaction. The wavelength is related to the “wave” aspect of electromagnetic energy while the frequency is related to the “corpuscular” aspect, the photon or quantum of energy. The wavelength is helpful for evaluating the probability of finding photons, which are the unit of exchange of electromagnetic energy with the surroundings during an interaction. The energy E of each photon is $E = h\nu$, where h is the Planck constant. The frequency ν is not dependent on the medium of propagation. The measurement of frequency is preferred by signal engineers, while wavelength is the preferred choice for opticians.

Spectral Distribution

Another important aspect of wavelength measurement is the distribution of optical energy as a function of the wavelength. When a pure monochromatic wave is considered, there is no problem with the definition of wavelength because all the energy is concentrated on a single value. But this is not realistic. Such a distribution would be generated by a noise-free source that evidently does not exist. This brings up a point related to the spectral distribution of the electromagnetic energy for which wavelength has to be measured.

Generally speaking the energy emitted by a source is distributed over a given range of wavelengths or frequencies. This is called the *spectral distribution* of the source. If the

spectral distribution is a symmetrically decreasing function around a central wavelength value, this value is defined as the nominal wavelength. If the distribution is not symmetrical and/or reveals many intense peaks, the wavelength definition is more complex. Giving a specific value of nominal wavelength is then somewhat arbitrary and must be justified by stating the prevailing conditions for the selection. In the next section, wavelength (or frequency) definitions will be given more in detail.

Particular cases that will require specifications are modulated and pulsed sources. According to the Fourier analysis of such signals, a periodically modulated source will produce a number of sidelobes, each being the replica of the unmodulated original source while a pulse has a broad spectrum whose width is related to the reciprocal of its time duration.

Instruments Used for Wavelength Measurement

In the subsequent sections, we describe different instruments performing wavelength measurements. First, we present a brief overview of the principle of operation of the two most commonly used instruments, which are the spectrometer-based and the interferometer-based wavelength meters.

If the source has a broad spectrum, spectrometers are the instruments commonly used. The spectrum of the source is first analyzed using a dispersive element such as a prism or a grating. The intensity and central wavelength are then recorded by an appropriate detector for a number of dispersive element settings. This method results in a spectral analysis from which a nominal wavelength can be identified if the instrument has been previously calibrated.

Since wavelength is defined in the space domain, its accurate measurement will be done essentially through interferometric phenomena. Let us first consider a monochromatic source. The source signal is first split into space, allowing the various parts to travel different trajectories and recombine to produce interference patterns. Calibrated measurement of the periodicity revealed in the pattern gives an indication on the wavelength. We will later describe in detail the principle of operation of the most commonly used type of interferometers used to measure wavelengths: Michelson, Fizeau, and Fabry–Perot interferometers.

Before we proceed with the description of the different wavelength meters encountered in research laboratories and in the industry, the next section gives some information regarding basic electromagnetic wave theory as well as useful wavelength-measurement definitions.

BASIC ELECTROMAGNETIC WAVE THEORY AND USEFUL WAVELENGTH-MEASUREMENT DEFINITIONS

Theoretical Background

The resolution of Maxwell's equations in vacuum gives rise to plane electromagnetic waves. Such waves are characterized by an electric field (units of volts per meter) and a magnetic field (units of amperes per meter), mutually oriented in perpendicular directions, which are themselves perpendicular to the direction of propagation of the plane wave, given by the direction of the Poynting vector (units of watts per square meter). The latter is defined as the cross-product between the electric and magnetic fields. The modulus of the Poynting vector represents

the instantaneous intensity of the light beam. Since both fields are perpendicular to the direction of propagation, such waves are called transverse electromagnetic (TEM) waves. Practical collimated laser beams are often well described by plane waves.

Fluctuating Electric Field

We define a Cartesian component of the electric field labeled $E^{(r)}(\mathbf{r}, t)$, which is a real field variable at a point represented by a position vector $\mathbf{r}$, at a time t . For any real light beam $E^{(r)}(\mathbf{r}, t)$, will be a fluctuating function of time. For example, laser fields exhibit intensity noise as well as phase noise.

In order to simplify the present analysis, we do not use the real field variable $E^{(r)}(\mathbf{r}, t)$. We rather consider its corresponding analytic signal $E(\mathbf{r}, t)$, which has only spectral components for the positive frequency part of the spectrum. This complex field is defined as

$$E(\mathbf{r}, t) = \frac{1}{2}E^{(r)}(\mathbf{r}, t) + j\frac{1}{2}H\{E^{(r)}(\mathbf{r}, t)\} \quad (1)$$

where $H\{E^{(r)}(\mathbf{r}, t)\}$ denotes the Hilbert transform of the real field.

The field intensity at the position $\mathbf{r}$ and at time t is defined as

$$I(\mathbf{r}, t) = E^*(\mathbf{r}, t)E(\mathbf{r}, t) \quad (2)$$

The field intensity gives the beam power distribution over the beam shape.

We are usually concerned with stationary fields, in which case the statistical properties are independent of the origin of time. Such a field could be characterized by the correlations of the field at two space-time points,

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle E^*(\mathbf{r}_1, t)E(\mathbf{r}_2, t + \tau) \rangle \quad (3)$$

where $\langle \rangle$ denotes an ensemble average. Since the field is stationary, this ensemble average does not depend on t . This function is also called the mutual coherence function (1, Ch. 4). When evaluated at the same position, it gives the autocorrelation function of the field $\Gamma(\mathbf{r}, \mathbf{r}, \tau)$ at position $\mathbf{r}$. Moreover, it follows that the autocorrelation function evaluated for a zero delay gives the ensemble average of the field intensity at position $\mathbf{r}$,

$$\Gamma(\mathbf{r}, \mathbf{r}, 0) = \langle I(\mathbf{r}, t) \rangle \quad (4)$$

The intensity is constant over time for stationary fields.

According to the generalized Wiener-Khintchine theorem, the cross-spectral density is defined as the Fourier transform of the mutual coherence function,

$$W(\mathbf{r}_1, \mathbf{r}_2, \nu) = \mathcal{F}\{\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)\} \quad (5)$$

The spectral density, also called the spectrum of the light beam at position $\mathbf{r}$, is then the Fourier transform of its autocorrelation function,

$$W(\mathbf{r}, \mathbf{r}, \nu) = \mathcal{F}\{\Gamma(\mathbf{r}, \mathbf{r}, \tau)\} \quad (6)$$

So the spread of the spectral density is directly related to the decorrelation time. Short decorrelation time gives rise to broad

spectrum and vice versa. Upon proper normalization, the field spectral density could be related to the measured power spectrum.

Usually the spectral density of the field has important components only over a frequency range much narrower than the frequencies of interest. This case is called quasimonochromatic light. For such situations the spectral density of the field is often characterized by two global parameters: one giving the position of the spectrum and the other one estimating its width.

Parameter Definitions

The center frequency could be defined in many ways. We could use the median frequency ν_m , which is the frequency for which half the power of the signal is distributed on each side, so

$$\int_0^{\nu_m} W(\mathbf{r}, \mathbf{r}, \nu) d\nu = \int_{\nu_m}^{\infty} W(\mathbf{r}, \mathbf{r}, \nu) d\nu \quad (7)$$

A common way of specifying the position of the spectrum is the most probable frequency ν_{mp} , which is the frequency at which the spectral density of the field is maximum. Finally the usual frequency value used and measured is the mean frequency, which is

$$\bar{\nu} = \frac{\int_0^{\infty} \nu W(\mathbf{r}, \mathbf{r}, \nu) d\nu}{\int_0^{\infty} W(\mathbf{r}, \mathbf{r}, \nu) d\nu} \quad (8)$$

The other parameter is the linewidth of the spectral density of the field. We could define the half-power bandwidth $\Delta\nu_{hp}$ as the symmetrical range around the median frequency containing half the power of the light,

$$\int_{\nu_m - \Delta\nu_{hp}/2}^{\nu_m + \Delta\nu_{hp}/2} W(\mathbf{r}, \mathbf{r}, \nu) d\nu = \frac{1}{2} \int_0^{\infty} W(\mathbf{r}, \mathbf{r}, \nu) d\nu \quad (9)$$

The root mean square (rms) width of the spectrum is also often encountered,

$$\Delta\nu_{rms}^2 = \frac{\int_0^{\infty} (\nu - \bar{\nu})^2 W(\mathbf{r}, \mathbf{r}, \nu) d\nu}{\int_0^{\infty} W(\mathbf{r}, \mathbf{r}, \nu) d\nu} \quad (10)$$

Usually we use a more experimentally convenient definition, which is the full width at half maximum (FWHM) width. This width gives the frequency range over which the spectral density exceeds half its maximum value,

$$\Delta\nu = \nu_2 - \nu_1, \quad W(\mathbf{r}, \mathbf{r}, \nu_{1,2}) = \frac{1}{2} W(\mathbf{r}, \mathbf{r}, \nu_{mp}) \quad (11)$$

For quasimonochromatic light, the effective width of its power spectrum is much smaller than the mean frequency,

$$\frac{\Delta\nu}{\bar{\nu}} \ll 1 \quad (12)$$

All these definitions also apply to wavelength and to wave number ($\sigma = \nu/\nu$).

Interference

Let us introduce the concept of interference through a simple experiment. Consider a light beam incident on an opaque screen having two distinct pinholes at positions r_1 and r_2 . The light emerging from these pinholes is observed on a second screen distant from the first one. The electric field at a position r on the second screen is the sum of two components,

$$E_s(r, t) = K_1 E(r_1, t - t_1) + K_2 E(r_2, t - t_2) \quad (13)$$

where K_1 and K_2 are constant factors and t_1 and t_2 are the propagation delays from the pinholes,

$$t_1 = |r_1 - r|/c \quad \text{and} \quad t_2 = |r_2 - r|/c \quad (14)$$

The average intensity observed on the second screen is

$$\begin{aligned} \langle I_s(r, t) \rangle &= |K_1|^2 \langle I(r_1, t - t_1) \rangle + |K_2|^2 \langle I(r_2, t - t_2) \rangle \\ &\quad + 2 \operatorname{Re}[K_1^* K_2 \Gamma(r_1, r_2, t_1 - t_2)] \end{aligned} \quad (15)$$

This intensity differs from the sum of the intensities of the distinct components. A supplementary term arises, which takes into account the “correlation” between the two beams. This phenomenon is called interference.

For quasi-monochromatic beams, the mutual coherence function could be expressed as

$$\Gamma(r_1, r_2, \tau) = g(r_1, r_2, \tau) e^{j2\pi\bar{\nu}\tau} \quad (16)$$

where $g(r_1, r_2, \tau)$ is a slowly varying function of τ , since its bandwidth is much smaller than the mean frequency of the field. In this case, the observed intensity on the second screen is

$$\begin{aligned} \langle I_s(r, t) \rangle &= |K_1|^2 \langle I(r_1, t - t_1) \rangle + |K_2|^2 \langle I(r_2, t - t_2) \rangle \\ &\quad + 2|K_1 K_2| |g(r_1, r_2, t_1 - t_2)| \cos[2\pi\bar{\nu}(t_1 - t_2)] \\ &\quad + \angle(K_1^*, K_2) + \angle g(r_1, r_2, t_1 - t_2) \end{aligned} \quad (17)$$

where $\angle$ is defined as the angle. Thus, as we move along the second screen, the intensity varies sinusoidally at the spatial frequency $\bar{\nu}/c$. This variation is the interference fringe pattern. Fringe contrast is measured through their visibility, defined as

$$V(r) = \frac{\langle I_s(r, t) \rangle_{\max} - \langle I_s(r, t) \rangle_{\min}}{\langle I_s(r, t) \rangle_{\max} + \langle I_s(r, t) \rangle_{\min}} \quad (18)$$

where the extreme values are evaluated in the neighborhood of the position r .

When the mutual coherence function vanishes, the visibility is 0. This represents complete incoherence. The other extreme case is called complete coherence. In this case,

$$|\Gamma(r_1, r_2, \tau)| = \sqrt{\langle I(r_1, t) \rangle \langle I(r_2, t) \rangle} \quad (19)$$

which corresponds to the following visibility:

$$V(r) = \frac{2|K_1 K_2| \sqrt{\langle I(r_1, t) \rangle \langle I(r_2, t) \rangle}}{|K_1|^2 \langle I(r_1, t) \rangle + |K_2|^2 \langle I(r_2, t) \rangle} \quad (20)$$

The visibility is 1 if the intensities associated with each beam are equal. Otherwise, perfect contrast is not achieved for complete coherence. In the general case, we experiment partial coherence, so the resulting visibility does not reach 1.

SPECTROMETER-BASED WAVELENGTH METERS

In the next section, we describe instruments based on spectrometers using prism or diffraction gratings. We discuss their performances and limitations and explain their use as wavelength meters.

Basic Properties of Spectrometers

Spectrographs or Monochromators?. The difference between the two types of instrument is related to the detection of diffracted/dispersed light: in a spectrograph, a photographic plate or a charge-coupled device (CCD) array is placed in the output focal plane and records the entire spectrum. In a monochromator, a slit and a photodetector are used so that only a portion of the output spectrum is recorded. It is therefore necessary to move the detection system in the output focal plane, or to rotate the dispersive/diffractive element, to obtain the complete spectrum. Both instruments are often referred to as spectrometers in the literature.

Spectral Resolving Power and Rayleigh's Criterion. The spectral resolving power R is defined by

$$R = \frac{\lambda}{\Delta\lambda} = \frac{\nu}{\Delta\nu} \quad (21)$$

where $\Delta\lambda = \lambda_1 - \lambda_2$ is the smallest separation between two wavelengths λ_1 and λ_2 that the instrument is able to resolve. According to Rayleigh's criterion, illustrated in Fig. 1, two lines are resolved when the intensity profile corresponding to λ_1 has its maximum in coincidence with the minimum of the intensity profile corresponding to λ_2 . The interval $\Delta\lambda$ between two resolved wavelengths, known as the resolution limit, can also be expressed in terms of frequency $\Delta\nu$ or wave number $\Delta\sigma$.

The achievable resolving power of a spectrometer is in direct relation with the slit width and its linear dispersion $dx/d\lambda$,

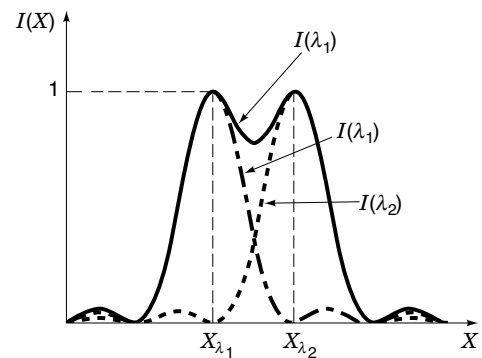


Figure 1. Rayleigh criterion for the definition of the spectral resolving power: two lines are resolved when the intensity profile corresponding to λ_1 has its maximum in coincidence with the minimum of the intensity profile corresponding to λ_2 . The interval $\Delta\lambda$ between two resolved wavelengths is the resolution limit.

which is given by

$$\frac{dx}{d\lambda} = f \frac{d\theta}{d\lambda} \quad (22)$$

where f is the lens focal length and $d\theta/d\lambda$ the angular dispersion. If the entrance slit width is ϵ_1 , the slit image in the focal plane of the focusing lens will be

$$\epsilon_2 = \frac{f_2}{f_1} \epsilon_1 \quad (23)$$

where f_1 and f_2 are, respectively, the focal length of the collimating and focusing lenses of the spectrometer. If we want to be able to resolve two lines at λ_1 and λ_2 , the separation of their image

$$\Delta x = f_2 \frac{d\theta}{d\lambda} \Delta\lambda = \frac{dx}{d\lambda} \Delta\lambda \quad (24)$$

must be greater than $2\epsilon_2$. It is therefore possible to increase the resolving power by decreasing the input slit width (at the expense of decreasing the amount of light to be detected). However, it will be limited because of diffraction due to the finite aperture a of the dispersive element. When the collimated beam is incident on the prism or grating, it experiences Fraunhofer diffraction resulting in an intensity distribution in $\sin^2(\gamma)(\sin(\pi\gamma)/\pi\gamma)$ with γ the diffraction angle. The central lobe of the distribution has a width equal to $2\lambda/a$ and contains 90% of the incident intensity. The distance between the two maxima in the output plane will be $f_2\lambda/a$ so that the condition for resolving λ_1 and λ_2 becomes

$$\Delta x \geq f_2 \frac{\lambda}{a} \quad \text{or} \quad \frac{\lambda}{\Delta\lambda} \leq a \frac{d\theta}{d\lambda} \quad (25)$$

giving the corresponding resolving power. One must notice that the spectral resolving power is limited by diffraction effects due to the finite aperture a (determined by the size of the prism or diffraction grating) and not by the entrance slit width. When taking into account the entrance finite slit width ϵ_1 , the condition is

$$\Delta x \geq \left(f_2 \frac{\lambda}{a} + \epsilon_1 \frac{f_2}{f_1} \right) \quad \text{or} \quad \Delta\lambda \geq \left(\frac{\lambda}{a} + \frac{\epsilon_1}{f_1} \right) \left(\frac{d\theta}{d\lambda} \right)^{-1} \quad (26)$$

However, there is a lower limit for ϵ_1 (2) resulting in a practical resolving power

$$R = \frac{\lambda}{\Delta\lambda} = \frac{1}{3} a \frac{d\theta}{d\lambda} \quad (27)$$

Free Spectral Range. In the case of a spectrometer, the free spectral range is the wavelength region where the instrument has a one-valued relation between the wavelength and the position $x(\lambda)$ in the focal plane of the focusing lens. For prism spectrometers, it corresponds to the whole wavelength range of the instrument, while for grating spectrometers it is related to the diffraction order m .

Prism Spectrometers

Principle of Operation. A typical arrangement for a prism spectrometer is given in Fig. 2. The entrance slit S illuminated

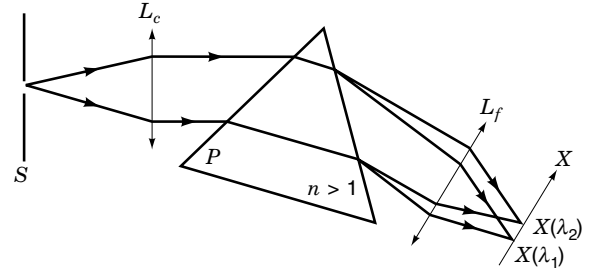


Figure 2. Typical configuration for a prism spectrometer: the light source illuminates the entrance slit S which is placed in the focal plane of a collimating lens L_c . The prism P diffracts the incident collimated beam and a focusing lens L_f images the entrance slit so that the position $x(\lambda)$ of the focal point in the x plane is wavelength dependent.

by the light source is placed in the focal plane of a collimating lens L_c . The prism P diffracts the incident collimated beam with an angle dependent on the wavelength. As shown in Fig. 3(a), the angle θ , relative to the incident direction, depends on the prism angle α , the angle of incidence of the beam θ_1 , and the refractive index of the prism material $n(\lambda)$. A focusing lens L_f is used after the prism to image the entrance slit so that the position $x(\lambda)$ of the focal point in the x plane is a function of λ . A wavelength scan can be done either by rotating the prism while keeping the output slit fixed, or the contrary.

Angular and Linear Dispersion. When the prism is used at the minimum deviation [Fig. 3(a)], we have

$$\frac{1}{2} \sin(\theta + \alpha) = n \sin(\alpha/2) \quad (28)$$

Then, it is possible to derive the angular dispersion $d\theta/d\lambda$ by first evaluating $d\theta/dn$. We have

$$\frac{d\theta}{dn} = \frac{2 \sin(\alpha/2)}{\cos[(\theta + \alpha)/2]} = \frac{2 \sin(\alpha/2)}{\sqrt{1 - n^2 \sin^2(\alpha/2)}} \quad (29)$$

The angular dispersion is given by

$$\frac{d\theta}{d\lambda} = \frac{2 \sin(\alpha/2)}{\sqrt{1 - n^2 \sin^2(\alpha/2)}} \frac{dn}{d\lambda} \quad (30)$$

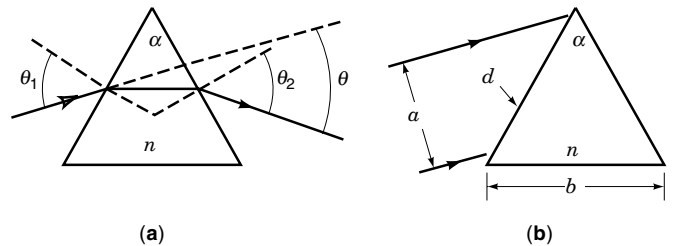


Figure 3. (a) Illustration of a diffraction prism used at minimum deviation: the beam exits the prism with an angle θ verifying the equation $\frac{1}{2} \sin(\theta + \alpha) = n \sin(\alpha/2)$, where α is the prism angle and n is its refractive index. (b) The limiting aperture a of the prism d and on the incident angle θ_1 following the equation $a = d \cos \theta_1 = b[(\cos \theta_1)/2 \sin(\alpha/2)]$.

We see that dispersion increases with the prism angle α (and so does the size of the prism) but does not depend on the prism size. Therefore, small prisms can be used for small laser beams while keeping the same angular dispersion. However, the prism must be chosen large enough to prevent diffraction problems and to achieve a large spectral resolving power. An equilateral prism with $\alpha = 60^\circ$ is usually chosen as the best compromise. In that case

$$\frac{d\theta}{d\lambda} = \frac{1}{\sqrt{1 - (n/2)^2}} \frac{dn}{d\lambda} \quad (31)$$

For values of n around 1.4 to 1.6, the angular dispersion reduces to

$$\frac{d\theta}{d\lambda} \cong n \frac{dn}{d\lambda} \quad (32)$$

The linear dispersion $dx/d\lambda$ depends directly on the prism material dispersion $dn/d\lambda$ and on the focal length f of L_f . It is given by

$$\frac{dx}{d\lambda} = f \frac{d\theta}{d\lambda} \quad (33)$$

Spectral Transmission and Dispersion of Prism Materials. The transparent spectral range for fused silica prisms is 180 nm to 3000 nm. For shorter wavelengths (VUV region), lithium or calcium fluoride can be used, while for longer wavelengths CaF_2 , KBr, or NaCl are transparent up to 30 μm .

The spectral dispersion $dn/d\lambda$ increases greatly near a region of absorption, making glass (e.g., BK7) an attractive material in the visible and near-ultraviolet regions while quartz is more interesting for the ultraviolet region.

Performance and Limitations. If the collimating and focusing lenses L_c and L_f are not achromatic, their focal length will decrease with wavelength. This effect can be compensated by tilting the image plane so that it matches the focal point for every wavelength. If the plane is tilted by an angle ϕ , the linear dispersion will be increased by a factor $1/\cos \phi$.

The resolving power is

$$R = \frac{1}{\lambda} a \frac{d\theta}{d\lambda} \quad (34)$$

where the limiting aperture of the prism a is given by [see Fig. 3(b)]

$$a = d \cos \theta_1 = b \frac{\cos \theta_1}{2 \sin(\alpha/2)} \quad (35)$$

with d the height of the input face of the prism and b the length of its base. At minimum deviation, we have

$$n \sin(\alpha/2) = \sin \theta_1 \quad (36)$$

so that

$$\frac{\lambda}{d\lambda} = b \frac{dn}{d\lambda} \quad (37)$$

is only limited by the size of the prism base and the dispersion. For example, if the prism material has a dispersion $dn/d\lambda$ of

500 cm^{-1} at 633 nm and the prism has a base $b = 50 \text{ mm}$, then $R \leq \frac{1}{3} \times 2500$. A spectrometer with such a prism is then able to resolve two lines separated by $\Delta\lambda = 0.75 \text{ nm}$ at 633 nm.

A prism spectrometer is interesting because it allows an unambiguous determination of wavelengths due to the nature of the equation $x(\lambda)$. Its cost can be low when using a small prism and it is simple to make and to adjust. However, as its spectral resolution is limited, it is mainly used for wavelengths scans over large spectral regions as a preliminary survey work.

Grating Spectrometers

Principle of Operation. Many configurations have been demonstrated for grating spectrometers. One possibility is to use an arrangement similar to the prism spectrometer (Fig. 2) and replace the prism by a diffraction grating used in transmission. Another way is to use the diffraction grating in reflection. Plane diffraction gratings can be used in a Littrow mounting [Fig. 4(a)] or in the Ebert mounting with a concave mirror [Fig. 4(b)]. A variation of the Ebert mounting is the Czerny–Turner mounting where two adjacent concave mirrors are used. Many configurations have also been demonstrated using concave diffraction gratings (3).

The input light is first collimated by a lens or a concave mirror. It is then diffracted by the grating that separates the different wavelengths. The refracted light is focused in the detection plane. If an aperture is used, its width determines the wavelength resolution of the spectrometer.

In every case, the diffraction grating acts as a wavelength-selective reflector, by reflecting light into well-defined directions for each incident wavelength. The different reflected beams correspond to the orders of the grating. The general equation for the grating is

$$d(\sin \alpha \pm \sin \beta) = m\lambda \quad (38)$$

where d is the groove dimension, α and β are the incidence and reflection angles relative to the grating normal, m is the grating order, and λ the wavelength [see Fig. 5(a)]. The amplitude of the wave reflected in the detection β is the superposition of the amplitudes reflected by all the grooves illuminated. The

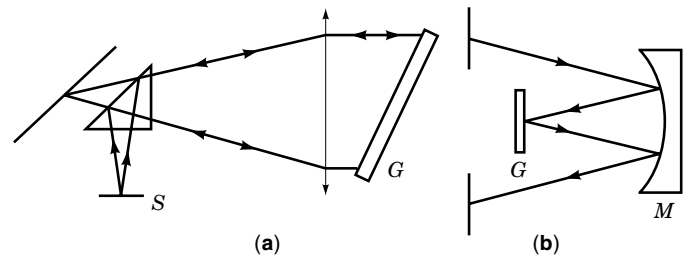
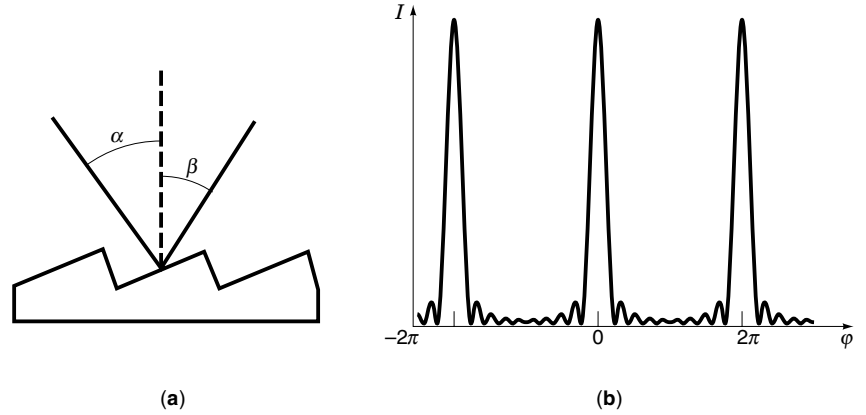


Figure 4. (a) Grating spectrometer using a Littrow mounted grating: the light exiting the input slit is first reflected by the prism and diffracted by the grating. The diffracted light is imaged on the output plane. Note that the slit and the prism are placed in a different plane from the grating so that the light reflected by the grating does not pass twice in the prism. (b) Grating spectrometer using the Ebert mounting configuration: the light exiting the slit is collimated by a concave mirror, diffracted by the grating and imaged in the output plane after a second reflection on the concave mirror.

Figure 5. (a) The diffraction of a beam by a diffraction grating follows the equation $d(\sin \alpha \pm \sin \beta) = m\lambda$, where d is the groove dimension, α and β are the incidence and reflection angles relative to the grating normal, m is the grating order and λ the wavelength. (b) Intensity distribution of the reflected beam for $N = 10$. The amplitude of the wave reflected in the direction β results from the interference of the beams reflected by all the grooves illuminated. The corresponding intensity is $I = R_g I_0 \sin^2(N\phi/2)/\sin^2(\phi/2)$.



corresponding intensity is

$$I = R_g I_0 \frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)} \quad (39)$$

where R_g is the reflectivity of the grating, I_0 the intensity of the incident wave on each groove, N the number of grooves illuminated, and

$$\phi = \frac{2\pi}{\lambda} d(\sin \alpha \pm \sin \beta) \quad (40)$$

is the phase difference between two adjacent grooves. Figure 5(b) shows I as a function of ϕ for $N = 10$. Each maximum occurs when $\phi = 2k\pi$. In terms of β , the line profile has a base half width of

$$\Delta\beta = \frac{\lambda}{Nd \cos \beta} \quad (41)$$

We can evaluate the angular dispersion from the grating equation to obtain

$$\frac{d\beta}{d\lambda} = \frac{m}{d \cos \beta} = \frac{\sin \alpha \pm \sin \beta}{\lambda \cos \beta} \quad (42)$$

which only depends on α and β and not on the number of grooves. The angular dispersion is reduced to

$$2 \tan \alpha / \lambda \quad (43)$$

in the case of a Littrow mounted grating ($\alpha = \beta$).

Possible Monochromator Configurations

A double monochromator is a combination of two single monochromators placed in cascade. This configuration improves the dynamic range, but the wavelength span is reduced and the losses are increased, degrading the sensitivity. Another possible configuration is the double-pass monochromator, which provides the dynamic range of a double monochromator while keeping the sensitivity and compactness of a single one.

Performance and Limitations. The resolving power R can be expressed from

$$\Delta\lambda = \Delta\beta \left(\frac{d\beta}{d\lambda} \right)^{-1} \quad (44)$$

so that

$$R = \frac{Nd(\sin \alpha \pm \sin \beta)}{\lambda} = mN \quad (45)$$

R depends on the number of illuminated grooves N and on the diffraction order m used. If the finite slit width and diffraction effects are considered we have:

$$R = \frac{1}{3}mN \quad (46)$$

R is limited physically by the dimensions of the grating. It is advantageous to use a high-order m but the grating reflectivity might decrease drastically. For example, a 5 cm $\times$ 5 cm grating with 1200 grooves/mm used to first order has a practical resolving power of 2×10^4 . When used at 1550 nm, it is capable of resolving $\Delta\lambda = 0.075$ nm.

The main disadvantage of grating spectrometers is the possible ambiguity in interpreting the output spectrum as the different orders will overlap for different wavelengths. A solution consists in using wavelength-selective filters in order to isolate one particular order of diffraction. However, the resolving power is much better than that obtained with a prism spectrometer. It can be increased by using a double-pass configuration where the beam is diffracted once by the grating, then passes into an aperture and diffracted again before being detected.

Wavelength Measurement Using Spectrometers. Both prism and grating spectrometers can be used as wavelength meters. To do so, it is necessary to tune the instrument over the unknown light-source spectrum. Another important aspect is to perform an accurate calibration for the wavelength determination.

One possibility for tuning monochromators is to rotate either the prism or the grating, which causes a different wavelength to be focused on the output slit, placed in front of a detector. Another possibility is to keep the prism or grating fixed and move the slit and photodetector in the output plane. In both cases,

Table 1. Comparative Table of Commercial Instruments

Model Company	MS9710A Anritsu	HP 71451B Hewlett Packard	WA-1500 Burleigh	HP 86120B Hewlett Packard	7711 New Focus	LMW-6500B ILX Lightwave
Configuration	Grating spectrometer	Double-pass grating spectrometer	Michelson interferometer	Michelson interferometer	Fizeau interferometer	Colored glass filter
Wavelength range (nm)	600–1750 nm	600–1700 nm	200–650 nm (UV) 600–1800 nm (NIR) 1500–4000 nm (IR)	700–1650 nm	400–1000 nm	400–1100 nm (OMH6722B) 950–1650 nm (OMH6727B)
Absolute accuracy (nm)	±0.05 nm	±0.3 nm	±1 × 10 ⁻⁷	± 0.005 nm @ 1550 nm	± 0.01 nm	± 0.5 nm
Resolving power (nm)	0.07 nm	0.08 nm	NA	0.16 nm @ 1550 nm	NA	NA
Sensitivity	–90 dBm (1.25–1.6 μm)	–90 dBm (1.2–1.6 μm)	–10 dBm (UV) –17 dBm (other)	–25 dBm (0.8–1.2 μm) –40 dBm (1.2–1.6 μm)	–24 dBm (400 nm) –7 dBm (1 μm)	–20 dBm
Maximum input power (dBm)	20	30	20	Not available	11	30
Dynamic range (dB)	70 Δλ = ±1 nm	50 Δλ > ±1 nm	NA	25 Δλ = 0.8 nm	NA	NA
Meas. rate	0.5 s/500 nm	≥50 ms/40 nm	1 Hz	1 Hz	1–10 Hz	1, 4, or 16 Hz
Input laser requirements	CW or pulsed	CW or pulsed	CW single mode	CW up to 100 wavelengths	CW or pulsed (pulse length ≥30 ps)	CW single-mode or pulsed
Calibration source	Internal wavelength Reference (C ₂ H ₂)	External wavelength references	Internal stabilized He-Ne laser	Internal He-Ne laser	Plant calibration	Plant calibration

NA: not applicable.

the prism or grating angle or the slit–photodetector movement must be controlled precisely and in a repeatable way to allow accurate tuning from measurement to measurement (4).

In the case of spectrographs (no moving parts), the spectrum is recorded at once using a photodetector array, and it is only necessary to know the wavelength calibration along the photodetector array. In order to make sure that the wavelength is correctly assigned, a calibration procedure is required for spectrometers as they do not have an internal wavelength reference. It is necessary to have external reference sources such as spectral lamps or lasers stabilized to atomic or molecular absorption lines. For an accurate calibration, the use of several wavelength references (at least three) is recommended and they should be distributed over the whole spectral range of the instrument. Some commercial instruments now include internal references based on the use of molecular resonances, such as acetylene, to perform internal calibrations (see Table 1).

INTERFEROMETER-BASED WAVELENGTH METERS

The next section describes wavelength meters using interference phenomenon, including Michelson, Fizeau, and Fabry–Perot interferometers.

Classification of Interferometers

Optical interferometers can be used to perform a wide variety of precision measurements such as measurement of length, studies of surface structure, measurement of temperature, pressure, and particle velocities. Those measurements are based on the interference phenomena produced by light waves. Wavelength measurement can also be realized with very high precision. In fact, interferometers are the most accurate instruments for measuring the wavelength of an optical source.

Interferometers can be classified as two-beam interferometers or multiple-beam interferometers according to the number

of interfering beams. The most common type of two-beam interferometers used to measure wavelength are the Michelson and Fizeau interferometers. The Fabry–Perot interferometer is the most usual type of multiple-beam interferometers.

Wavelength meters based on interference phenomena can be subdivided in two categories; dynamic and static wavelength meters. The former relies on the displacement of an element. It is the most accurate type of instrument but it can only perform wavelength measurement on continuous wave (CW) sources. On the other hand, static wavelength meters have no moving parts and can be used to measure wavelength of CW or pulsed sources.

Michelson-Based Wavelength Meters

The principle of operation of the Michelson interferometer is presented in Fig. 6. The incident beam is first divided by a 50-50 beam splitter that can be either a partially reflecting metal mirror or dielectric film on a transparent substrate. The two beams are next recombined by the same beam splitter on a detector (or a screen). Since the second beam passes only one time through the beam splitter, a compensating plate may be used in the second arm to equalize the optical path followed by the two beams, but is not always necessary for small collimated beams.

If the mirror M_2 is moved by Δx , the intensity of the interference pattern on the detector changes sinusoidally. A complete cycle is observed when the optical path difference between the two arms of the interferometer corresponds to one wavelength of the incident light beam. Thus, when a mirror displacement of $\lambda/2$ is performed, a complete cycle is observed.

The wavelength of the incident light beam can be deduced if the interference maxima frequency is measured when the mirror M_2 is displaced at constant speed

$$\lambda = \frac{2v_m}{f} \quad (47)$$

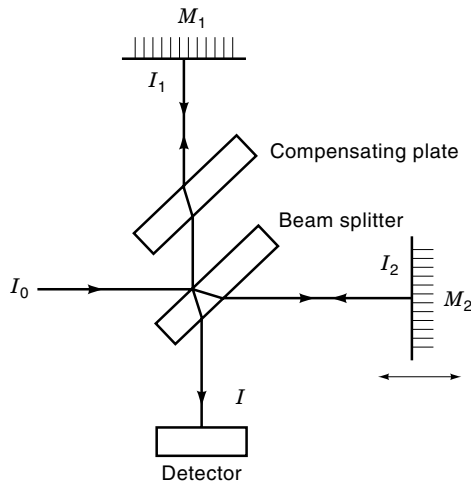


Figure 6. Principle of operation of a Michelson interferometer. An incident beam is divided by a 50–50 beam splitter. The two beams are next recombined by the same beam splitter on a detector. If the mirror M_2 is moved, the intensity of the interference pattern on the detector changes sinusoidally. The wavelength of the incident light beam is deduced if the interference maxima frequency is measured when the mirror M_2 is displaced at constant speed.

where v_m is the displacement speed of moving mirror M_2 and f is the frequency of apparition of interference maxima.

The main problem with the previously described Michelson-based wavelength meter resides in the determination and the accuracy of the displacement speed of the moving mirror. For that reason, Michelson interferometers are rarely used in that configuration unless low accuracy ($\sim 1 \times 10^{-4}$) is sufficient.

The most common type of Michelson-based wavelength meter uses a two-beam interferometry process (5). In that configuration, a two-beam scanning Michelson interferometer compares the number of fringes obtained by both a reference laser for which the wavelength is known very accurately and an unknown laser. Absolute wavelength measurement of CW sources can then be performed. For the two beams, the optical path d is changed by the same and known amount. A typical

optical configuration for that type of interferometer is presented in Fig. 7.

With this type of wavelength meter, the wavelength of the unknown laser can be determined by the following relation:

$$\lambda_U = \frac{N_R n_U \lambda_R}{N_U n_R} \quad (48)$$

where N_R and N_U are the number of fringes counted on the reference and input photodetectors (corresponding to the reference and unknown laser), n_R and n_U are the refractive index of the media at the reference and unknown wavelengths, and λ_R is the wavelength of the reference laser. From that equation, it is clear that all parameters in Eq. (48) must be known very accurately. The maximum relative uncertainty of the unknown wavelength λ_U is the sum of the relative uncertainty of each relevant parameter, which are the wavelength reference, number of fringes, the ratio of the refractive indexes, beam misalignment, and wave-front distortion (2, Ch. 4).

The measurement accuracy can be greatly enhanced by improving the wavelength reference's own accuracy. Wavelength references found in current laboratory instruments are relatively simple He–Ne gas lasers with about 10^{-7} absolute accuracy. A more accurate I_2 -stabilized He–Ne laser is often used in laboratory installations where size is of minor concern but is somewhat difficult to implement in a portable instrument. Those type of lasers are calibrated with an uncertainty smaller than 10^{-10} . Recently, some progress has been made in replacing He–Ne lasers at 633 nm by semiconductor lasers, providing a more compact and reliable source for the reference laser (6).

Other sources of error for wavelength determination come from the number of fringes seen by the input and reference detectors. One way to improve the accuracy is to count as many fringes as possible; this can be done by increasing the retroreflector mirror displacement. Unfortunately, this also increases the size of the instrument which for a commercial instrument is not suitable. Usually, this technique is implemented in a laboratory environment where space is not a concern. The maximum retroreflector mirror displacement is limited by the coherence length l_c of the reference and unknown laser sources.

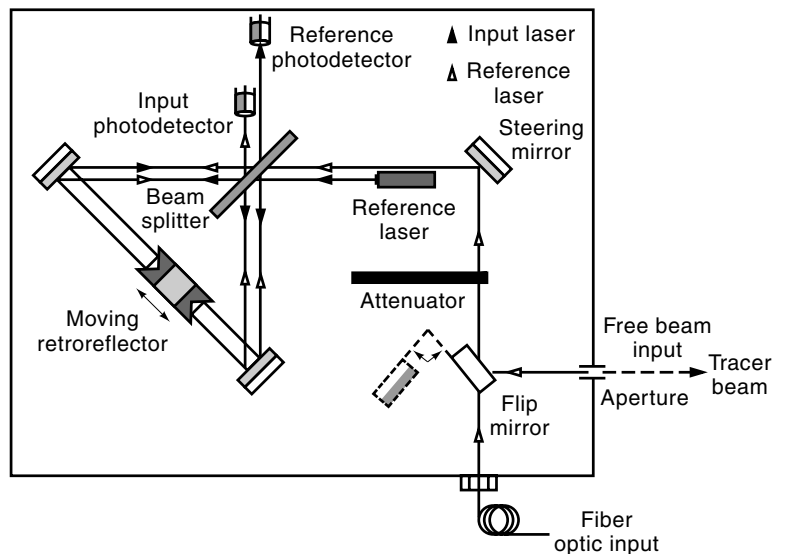


Figure 7. Commercially available two-beam scanning Michelson-based wavemeterTM (courtesy of Burleigh).

After a displacement corresponding to one coherence length, no interference pattern can be observed (7). The coherence length is related to the spectral width of the source and is given by

$$l_c = c/\Delta\nu \quad (49)$$

where c is the speed of light in vacuum and $\Delta\nu$ is the spectral width of the source. From that relation, it is clear that the frequency of a source emitting in a very narrow band of frequency can be determined with greater precision than a broad-band source.

Another way to improve the wavelength-measurement accuracy without increasing the path displacement of the retroreflector mirror is to determine a fractional order number of interference fringes. Techniques such as phase-locking an oscillator to an exact multiple of the frequency of the ac signal from the reference laser (8) or using a vernier method in which the counting cycle starts and stops when the two signals coincide (9) have been proposed. With these techniques, fringe fractions can be determined with an uncertainty of $1/500$ of a fringe (10). Moreover, vibrations during the measurement must be reduced to a minimum in order to eliminate the frequency jitter on the fringe signal (11).

To eliminate the dispersion of air, the wavelength meter is often operated in a vacuum chamber. If the instrument is operated in air, the index of refraction depends on the wavelength, the total air pressure, the partial pressures of H_2O and CO_2 , and the temperature. All those parameters must be controlled and/or recorded with great precision. Moreover, the ratio of the refractive indices depends on the wavelength difference $\Delta\lambda = \lambda_R - \lambda_U$. The relative uncertainty on the index of refraction depends on the wavelength range. Typically for visible wavelengths, the uncertainty ranges from 10^{-11} ($\Delta\lambda = 1 \text{ nm}$) to 5×10^{-9} ($\Delta\lambda = 200 \text{ nm}$).

There are other sources of systematic errors that can influence the achievable accuracy of a two-beam scanning Michelson-based wavelength meter. One of them is the misalignment of the two beams that causes them to travel slightly different path lengths. As an example, if the two beams are slightly tilted against each other by 10^{-4} rad, the systematic relative error becomes 5×10^{-9} . In the wavelength meter proposed in Fig. 7, the corner cube retroreflector guarantees that the incoming light beam is reflected exactly parallel to its incident direction regardless of a slight misalignment.

Finally, the quality of the optical components can also limit the measurement accuracy. With a surface quality of $\lambda/10$, wavefront distortions are already visible in the interference pattern (2). Moreover, to minimize diffraction effects particularly important in the infrared region, a large beam diameter should be used. The uncertainty due to diffraction is inversely proportional to the square of the beam diameter (11).

Laboratory and Commercial Instrument. In a laboratory experiment, Ishikawa, Ito, and Morinaga demonstrated a wavelength meter with an accuracy of 4×10^{-10} (10). In that experiment, the main limitation was caused by a slight optical misalignment between the two beams. An improved version of that wavelength meter was later designed (11). That time the wavelength uncertainty was evaluated at 7×10^{-11} and was limited by the accuracy of the I₂-stabilized He–Ne laser.

For the most accurate currently available commercial instrument (Burleigh WA-1500), the wavelength of the unknown laser source can usually be determined to an accuracy of 10^{-7} . In that case, the accuracy is limited by the reference laser, which is a He-Ne gas laser stabilized on its gain curve. Temperature and pressure sensors are used to evaluate the index of refraction of air as the interferometer is not evacuated. Figure 8 presents a picture of that instrument.



Figure 8. Front view of the Burleigh WA-1500 Michelson-based wavemeter (courtesy of Burleigh).

Multiwavelength Measurement. One of the disadvantages of the previously described wavelength meter is that it can only perform a single wavelength measurement. However, the previous configuration can be modified to perform multiwavelength measurement. This task is particularly suited to dense-wavelength division multiplexing (DWDM) communication systems. The fringe pattern produced by the unknown laser can be sampled at regular intervals corresponding to the period of the reference-laser interference pattern. A Fourier transform on those data yields the complete spectrum. The resolution or the minimum frequency spacing $\Delta\lambda$ that can be measured by the instrument is directly related to the number of counted fringes (12)

$$\frac{\lambda}{\Delta\lambda} = 2N \quad (50)$$

As for the accuracy, the resolution can be improved by counting as many fringes as possible; this can be done by increasing the retroreflector mirror displacement.

A motionless Michelson interferometer with a fixed path difference Δs can also be used to measure the wavelength of pulsed sources (13). In such a wavelength meter, the incident signal enters the interferometer polarized at 45° . A phase difference $\Delta\phi = \pi/2$ is next introduced between the two polarized components. The interference signal at the exit of the interferometer is recorded separately for both polarizations. From the two interference signals obtained, it is possible to deduce the wave number $\sigma = 1/\lambda$ modulo $1/\Delta s$ since all wave numbers $\sigma_m = \sigma_0 + m/\Delta s$ ($m=1, 2, 3, \dots$) give the same interference signals. If similar interferometers having a path difference in geometric ratios are used, the wave number can be deduced without ambiguity. Since such an instrument measures the wave number it is called a *sigmometer*.

Fizeau-Based Wavelength Meters

Fizeau interferometers can be used to perform wavelength measurement (14). Those static wavelength meters have no moving parts and can measure the wavelength of CW and pulsed sources.

Fizeau interferometers consist of two plates inclined to form a wedge of angle α (see Fig. 9). When plane monochromatic light is incident on that type of interferometer, reflected light forms a series of uniformly spaced interference fringes for which the intensity will vary as

$$I(x) = \frac{1}{2} \left[1 + \cos \left(\frac{2\pi x}{\Lambda} + \phi \right) \right] \quad (51)$$

where $\Lambda = \lambda/(2\alpha)$, x is the distance along the detector array, and ϕ is the phase at $x = 0$. The fringes produced are imaged on a detector array. With a typical interferometer plate spacing of 1 mm and a wedge of 3 min, an accurate measurement of the fringe period can give a wavelength measurement accuracy of 10^{-4} . Moreover, if the phase of the interference pattern on the detector array is determined accurately (15), the accuracy can go as high as 10^{-7} .

Such an accurate measurement can only be accomplished if the wedge angle and spacing are previously calibrated. Also, the fringe pattern on the detector must approximate as much as possible the sinusoidal pattern of Eq. (51). This is done by ensuring that the incident light has plane-wave fronts. This

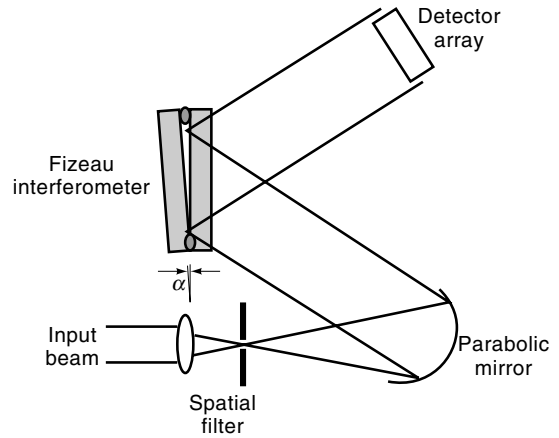


Figure 9. Principle of operation of a Fizeau-based wavelength meter. Fizeau interferometers consist of two plates inclined to form a wedge of angle α . When plane monochromatic light is incident on that type of interferometer, reflected light forms a series of uniformly spaced interference fringes for which the intensity will vary. The fringes produced are imaged on a detector array. The fringe period is computed and gives a wavelength measurement.

is accomplished by spatially filtering the incident light and by using interferometer plates with excellent flatness. Dispersion effects must also be taken into account for very accurate wavelength measurement. The advantage of the Fizeau wavelength meter is its simple, compact, and robust design. However, the accuracy of such a wavelength meter is usually less than a scanning Michelson-based wavelength meter.

Figure 10 presents a commercially available Fizeau-based wavelength meter. It is worth mentioning that this instrument contains a thin fused silica etalon, a precision input aperture, and a silicon CCD array. Due to the type of detector array used, this instrument can perform wavelength measurement from 400 nm to 1000 nm. The accuracy of the instrument is 10^{-5} . The instrument is wavelength calibrated at the factory using five different stabilized lasers. The calibration process allows for effective determination of the etalon thickness and to linearize the detector array signal. All the components are housed in



Figure 10. Front view of the New Focus Fizeau-based wavelength (courtesy of New Focus).

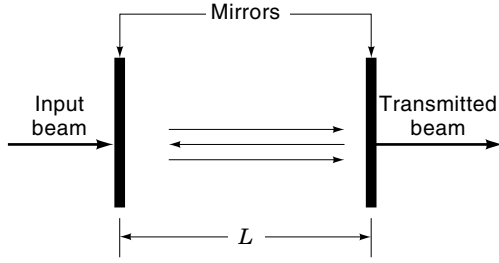


Figure 11. Principle of operation of a Fabry-Perot interferometer. This type of interferometer is made of two highly reflecting mirrors (flat or concave) uniformly spaced by air or a dielectric material. When coherent light is injected on one side of the interferometer, an interference phenomenon is created by the multiple reflections inside the cavity.

a thermally isolated enclosure. The index of refraction of air is mathematically compensated for temperature and pressure variations.

Fabry-Perot-Based Wavelength Meters

Scanning Fabry-Perot interferometers can be used to perform wavelength measurement although those instruments are usually used as optical spectrum analyzers. Figure 11 depicts the principle of operation of a Fabry-Perot interferometer. This type of interferometer is made of two highly reflecting mirrors (flat or concave) uniformly spaced by air or a dielectric material (16).

When coherent light is injected on one side of the interferometer, an interference phenomenon is created by the multiple reflections inside the cavity. For that reason, Fabry-Perot interferometers are called multiple-beam interferometers. The interference is constructive and a stationary wave is built when the distance between the two mirrors corresponds to an integer multiple of half wavelengths of the input beam. This condition is given by

$$L = \frac{m\lambda}{2n} \quad (52)$$

where L is the distance between the mirrors, m is an integer, and n is the refractive index of the material inside the cavity. When the condition in Eq. (52) is respected, light transmission through the interferometer is maximum. The frequency spacing between the transmission peaks is called the free spectral range (FSR) and is determined by the spacing between the mirrors and by the refractive index of the material inside the cavity. The FSR (in hertz) is given by

$$\text{FSR} = \frac{c}{2nL} \quad (53)$$

The transmission coefficient of an ideal Fabry-Perot interferometer is described by an Airy function (16)

$$\tau(\lambda) = \frac{I_t}{I_0} = \left[1 + \frac{4R}{(1-R)^2} \sin^2 \left(\frac{2\pi nL \cos(\theta)}{\lambda} \right) \right]^{-1} \quad (54)$$

where I_t and I_0 are the input and transmitted intensities, R and T are the mirror reflection and transmission intensity coefficients, and θ is the angle (relative to the mirror plane

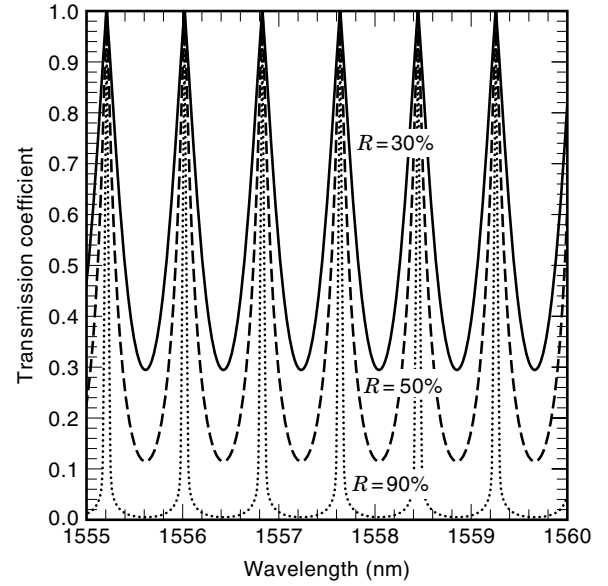


Figure 12. Transmission of an ideal Fabry-Perot interferometer for different mirror reflectivities ($\theta = 0^\circ$, $L = 1.5$ mm, $n = 1$, FSR = 99.9 GHz).

inside the cavity) of the input beam (usually the input beam is perpendicular to the mirrors and this angle is 0°). Figure 12 presents the transmission of an ideal Fabry-Perot interferometer for various mirror reflectivities at $\theta = 0^\circ$.

From Fig. 12, we see that another parameter, the finesse (F), gives information about the spectral width of each transmission peak. This parameter depends on the mirror reflectivities, optical quality, and alignment. The finesse is defined by

$$F = \frac{\text{FSR}}{\Delta\nu} \quad (55)$$

where $\Delta\nu$ is the full width at half maximum of the transmission peaks. If the finesse is much larger than 1 (which is usually the case), F can be approximated by

$$F = \frac{\pi R^{1/2}}{1-R} \quad (56)$$

Let us now return to the principle of operation of a scanning Fabry-Perot interferometer. If two signals are incident on the interferometer and if the distance L between the two mirrors is changed, the interference patterns of both signals appear one after the other. For a given FSR, a high finesse allows the resolution of very closely spaced signals. The resolution is given by (12)

$$\frac{\lambda}{\Delta\lambda} = mF \quad (57)$$

where m is the mode number. The resolution can be increased if a high-order mode is used. This can be done by using a large distance L between the mirrors. In that case, the FSR is reduced; this means that the spectral range is also reduced. Depending on the application, trade-offs have to be made. As we mentioned previously, that type of instrument is mainly used to perform

high-resolution optical spectrum analysis over a relatively narrow spectral range.

However, by counting the fringes obtained by one reference and one unknown laser while the mirrors are scanned over a relatively long displacement and by using a coincidence technique similar to the vernier technique, wavelength measurement with an accuracy of 1×10^{-7} can be obtained (17). Also, by carefully calibrating the position of the mirrors against wavelength and by carefully controlling parameters such as temperature, pressure and alignment, the instrument could be turned into a wavelength meter. Today, Fabry–Perot-based wavelength meters are widely used for measuring the wavelength of pulsed lasers. In that case, a plane-mirror Fabry–Perot interferometer is operated in a static configuration (etalons). Such etalons, when illuminating with diverging light, produce a characteristic bull’s-eye fringe pattern corresponding to different angles of incidence. A CCD camera positioned across the fringe pattern is then used along with a computer to reconstruct the interferometer spectrum. The fringes of a Fabry–Perot etalon are unequally spaced; for that reason, Fizeau etalons, which produce equally spaced fringes, can be used instead of the Fabry–Perot etalon. However, Fabry–Perot etalons have better shaped fringes. Let us also mention that high-accuracy (1 ppm) wavelength measurement of pulsed laser sources can also be performed using two different fixed Fabry–Perot etalons and a reference laser for continuous calibration. In that case, the interference patterns generated by both etalons are recorded by two CCD cameras. The patterns are finally analyzed in a computer and the wavelength is deduced.

APPLICATIONS AND ADVANCED TOPICS

Applications

Scientific Measurements. With the now generalized use of the laser as a scientific investigative tool, wavelength meters have established a firm presence in research and industrial environments. Laser-based measurement techniques reach well beyond the optics field into general physics, chemistry, biology, and engineering. The extremely coherent signal emitted from a laser, whether semiconductor, solid-state, gas, or dye-based, is ideal for probing materials in order to determine their composition or properties, measure dimensions or distances accurately, and much more. In such measurements, the laser wavelength must generally be known precisely, and often the parameter of interest can actually be transformed into a reading of wavelength or wavelength offset. This is where the accuracy of modern wavelength meters as instruments plays a vital role. Wavelength meters in general can also be used for spectral analysis, which expands even further their applicability.

Optical spectroscopy, in its various forms, is useful in a vast number of applications, since it deals with the precise measurement of atomic and molecular absorption or emission features. These features reveal fundamental characteristics of virtually any material, e.g. metals, gases, organic tissue, and crystals (18). Possible wavelength ranges go from visible to infrared, which will guide the choice of a particular tunable laser, for example, semiconductor or dye (19), and of a particular wavelength meter. In a basic measurement, the frequency of an absorption feature is readily obtained by tuning the laser to it and measuring its wavelength with a wavelength meter. Simple spectral analysis can then be achieved by scanning the laser wavelength through the absorption feature and recording

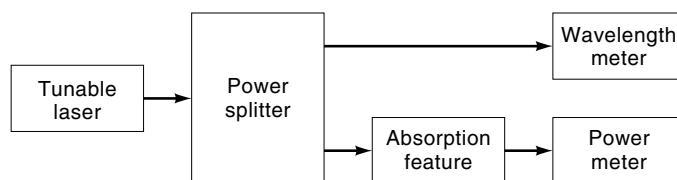


Figure 13. Use of the wavelength meter in a simple absorption spectroscopy experiment. Simple spectral analysis can be achieved by scanning the laser wavelength through the absorption feature and recording the amplitude versus wavelength for each step.

the amplitude versus wavelength for each step, as depicted in Fig. 13.

A different application of spectroscopy is the light detection and ranging (LIDAR), used for monitoring aerosols in the atmosphere. A pulsed (nanosecond time scale) laser signal is sent to a target zone, from which backscattered light is captured. Detailed optical frequency/spectrum measurement of this signal yields information about the target atmosphere. Since this is a pulsed application, the typical scanned Michelson interferometer is not suitable and instruments with faster response times using Fabry–Perot or Fizeau interferometers are likely to be used. Calibrated with a reference He–Ne laser, such an instrument gives information about the absolute wavelength and spectral properties of the backscattered light, which in turn give information about the scattering medium properties.

DWDM Optical Communication. The laser has also enabled an area of significance for electrical engineering, that of optical communication, where wavelength meters find widespread use. After years of steady growth with single-wavelength systems at $0.8 \mu\text{m}$, then $1.3 \mu\text{m}$, the advent of the erbium-doped fiber amplifier (EDFA) has enabled a massive deployment of systems in the low-loss region at $1.5 \mu\text{m}$ (20). Accurate knowledge of the wavelength is generally not required in a single-wavelength system. However, technology advances coupled to the ever-increasing demand for bandwidth fueled by the Internet explosion have recently led to industrial implementation of dense wavelength division multiplexing (DWDM). In such systems, up to tens of optical channels are transmitted on a single fiber with frequency spacing currently as low as 100 GHz (about 0.8 nm at $1.55 \mu\text{m}$). Typical bit rates for individual channels in current high-capacity transport applications are 2.488 Gbit/s (Sonet OC-48) and 9.952 Gbit/s (Sonet OC-192). (Sonet is a multiplexing format and is the acronym for Synchronous Optical NETWORK.)

The 100 GHz standard channel spacing has been endorsed globally through a recommendation of the International Telecommunication Union (ITU). This recommendation actually defines precisely each channel frequency to be used: at this time, 40 channels from $\nu = 192\,100 \text{ GHz}$ to $\nu = 196\,100 \text{ GHz}$ applying to the individual transmitters and receivers. Corresponding vacuum wavelengths can be computed with $\lambda = c/\nu$, where $c = 299\,792\,458 \text{ m/s}$, the speed of light. Each frequency (or wavelength) value is in essence absolute: no particular technique for obtaining it is specified. This allows various equipment manufacturers and users worldwide to work with a common set of specifications, easing interoperability while employing potentially different technologies. The frequency tolerance around the absolute value is not standardized rigidly and is evolving as manufacturers and users refine their designs.

Current observed tolerances are in the ± 10 GHz to ± 20 GHz range for a 100 GHz channel spacing.

Since every transmitter and receiver must be designed, constructed, verified, and maintained in accordance with such tolerances, the need for accurate and easy wavelength measurements is clear. DWDM transmitters typically use distributed feedback (DFB) semiconductor lasers (21), which present a significant frequency dependence on operational parameters like temperature and current. Since typical coefficients are 10 GHz/°C and 1.5 GHz/mA, and given the tight-frequency tolerance required, it follows that transmitter emission wavelengths need to be calibrated in terms of temperature and current. These parameters also influence markedly the laser's output power, a critical parameter that must also be mapped out.

Wavelength meters as instruments now find widespread industrial use in optical telecommunication manufacturers production facilities at various instances. At the DFB fabrication and testing stage, they are used for selecting a laser operating point (temperature and current within an acceptable range) resulting in a wavelength situated within a desired channel tolerance, this at a predefined power level. Devices for which such a setpoint cannot be found are rejected. At the transmitter subsystem assembly stage, they are used for setting up and verifying the laser initial operating conditions, always ensuring an output wavelength within prescribed tolerances. At the DWDM transmission system setup stage, they are used for final verification of the multiple channel wavelengths. Meters based on FFT techniques providing a multiwavelength function are particularly useful at this point since their spectral analysis capability provides a quick view of the entire transmitted signal. See the HP 86120B instrument in Table 1.

DWDM systems must be designed for very high reliability over many years, and laser frequency aging must be taken into account in order to guarantee that transmitter wavelengths will remain within the specified tolerances. Frequency stabilization external to the laser itself is then required and various schemes have been developed (22–24). The basic approach is to compare the emitted wavelength with some active (reference laser) or passive (reference filter) wavelength standard, determine if the transmitter wavelength is adequate, and take corrective action if not. This will counteract the effect of laser wavelength aging and maintain a precise emission frequency over long periods of time. In a way, this comparison with a wavelength reference implements the wavelength meter function, although it does not necessarily return a reading of the wavelength itself. Of course, one possibility is to monitor the transmitter output with an actual wavelength meter instrument, and instruct a controller unit to slightly modify the laser operating conditions over time so as to maintain a prescribed wavelength. As represented in Fig. 14, a multiwavelength meter allows this concept to be extended simultaneously to every channel in the system (25).

DWDM receivers generally use demultiplexing filters with well-defined transmission responses that must be measured precisely. As in the electrical domain, a simple way to achieve this measurement is to send the signal from a broadband noise source through the filter and look at the transmitted shape on an Optical Spectrum Analyzer (OSA). For the 1.55 μm communication band, an easily available noise source is the amplified spontaneous emission (ASE) found at the output of an EDFA (20). The OSA, typically a grating-based monochromator-type

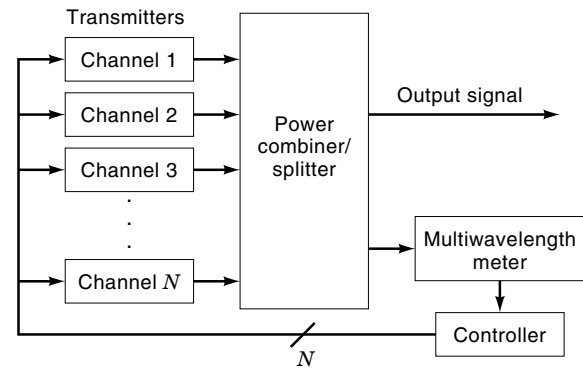


Figure 14. Use of a multiwavelength meter to control simultaneously the frequency of multiple laser transmitters in a DWDM communication system.

instrument, can measure wavelengths with less accuracy than a typical interferometer-based wavelength meter but offers more advanced spectral analysis capability. Again, wavelength meters with FFT-based spectral analysis capability can be used advantageously depending on the application. However, for very precise transmission versus wavelength measurements, one can revert to the fundamental spectroscopic technique of scanning the wavelength of a tunable laser (typically a laboratory instrument based on an external-cavity laser design (19, Ch. 8), recording simultaneously the transmitted power and the wavelength with, respectively, a power meter and a wavelength meter, exactly as depicted in Fig. 13.

Advanced Topics

Let us now address some foreseen or desirable avenues for improvements of the wavelength meter instrument, as well as discuss the current state-of-the art in wavelength measurement. In stand-alone instruments, one may wish to improve the resolution and accuracy of present units, in order to enable yet more precise measurements of all kinds. Improving the resolution means being able to sense a smaller wavelength change, and for the Michelson interferometer this requires a longer mirror travel. While this can be achieved easily in a laboratory interferometer setup, it is harder to realize in an actual instrument because of size, reliability and cost constraints. For example, a wavelength ratio uncertainty of 7×10^{-11} has been demonstrated in a laboratory configuration presenting a 60 cm mirror travel in vacuum and a 30 mm beam size (11). Mechanically scanned interferometers tend also to have slow update rates, which will likely be slowed down further by longer travels. Simultaneous improvements in resolution and measurement speed appear difficult, although compromises toward either parameter can certainly be achieved.

Improving the measurement accuracy rests first on improving the wavelength reference's own accuracy. Semiconductor lasers stabilized to atomic or molecular features are a foreseen development in future instruments, featuring reduced size, increased reliability, and an improvement in reference accuracy to the 10^{-10} level (23,24). Given an improved reference wavelength, one can then refine the interferometer's intrinsic design. This implies a minimization and a precise characterization of the sources of error inherent in the design, its tolerance to misalignment, and the ultimate effect on the wavelength reading. Again, in the case of a commercial instrument, cost and

reliability issues must be addressed at the same time as these refinements.

For the DWDM field, improvement paths may be oriented slightly differently, since the resolution and accuracy of commercial wavelength meters are basically sufficient for the needs of current and near-future systems. Emphasis should then be put on realizing wavelength meters which are smaller, cheaper, and more robust, particularly by developing solid-state designs with no moving parts. Resolution and accuracy may well be traded off for these considerations, depending on the particular application. Such developments might result in the ability to integrate an accurate wavelength-meter function (as well as spectrum analysis) into optical transmission subsystems. Various possibilities for solid-state wavelength meters have been investigated (25,26).

Finally, let us touch on absolute wavelength measurement. Here, the term *absolute* means that the measurement must be related directly to the universal reference for frequency and time, the cesium primary standard. The SI unit of time, that is, the second, is defined with respect to a microwave transition from the ground state of cesium, around 9.192 GHz. Cesium primary standards, or "atomic clocks," exist as standalone instruments, but the most accurate and stable ones are fairly complex installations found at various national laboratories around the world. Their output usually consists of an RF signal in the 10 MHz range. One approach for absolute optical frequency measurement is to synthesize from the cesium reference, through frequency multiplication and heterodyning, phase-locked signals at progressively higher frequencies until the unknown wavelength has been reached (24, Ch. 5). This unknown wavelength can of course be measured with less accuracy prior to the absolute measurement with the help of a standard laboratory wavelength/optical frequency meter. The final determination of frequency is done by measuring a specific RF or microwave transfer frequency at a particular point in the chain. Such an absolute optical frequency measurement can then be transformed into an absolute vacuum wavelength measurement through the basic relation $\lambda_0 = c_0/\nu$. This was enabled by the 1983 redefinition of a meter by the Bureau International des Poids et Mesures (BIPM) (27), which effectively defined the speed of light in vacuum as $c_0 = 299\,792\,458$ m/s exactly. Measurement accuracies of 10^{-12} have been obtained in the visible range and also at the well-established $3.39\,\mu\text{m}$ and $10\,\mu\text{m}$ wavelengths of metrological He-Ne and CO₂ lasers. This is to be compared with the 10^{-7} basic accuracy of the most accurate currently available commercial instrument, the Burleigh WA-1500 depicted in Fig. 8.

Optical comb generation (28), that is the generation of precisely controlled optical sidebands at RF or microwave intervals from a precise reference wavelength, can also be used for creating signals in the vicinity of the unknown signal. This technique has recently been employed for measuring accurately a series of molecular resonances of acetylene in the $1.5\,\mu\text{m}$ communication band (29). An accuracy in the 10^{-9} range was achieved.

CONCLUSION

Wavelength measurement is a fundamental and versatile function in many aspects of science and technology, as optical waves are prevalent in our everyday lives, in the scientific laboratory

and in technology-based commercial ventures. As we have seen, wavelength measurement can take various shapes depending on the particular application, and it calls upon diverse aspects of optics and both electrical and mechanical engineering.

One can use calibrated spectrometers based on prisms or gratings, or for better precision, Michelson, Fizeau, or Fabry-Perot interferometers. Each type of instrument presents advantages and limitations, so specific applications will dictate the choice of a particular configuration. Refined laboratory setups built for absolute measurements demand very accurate modeling and optical/mechanical design as well as construction, but size, complexity, and cost constraints are relaxed. Stand-alone laboratory instruments for scientific applications must be engineered so that a suitable level of performance is packaged in an easy to operate, robust, and cost-effective instrument. Finally, further improvements in the design of wavelength meters is leading the way toward the integration of this function into more and more compact subsystems, where size and cost are paramount.

BIBLIOGRAPHY

1. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge, UK: Cambridge Univ. Press, 1995.
2. W. Demtröder, *Laser Spectroscopy*, 2nd enlarged ed., Berlin: Springer-Verlag, 1996.
3. A. P. Thorne, *Electrophysics*, London: Chapman Hall & Science Paperbacks, 1974.
4. J. N. West and J. D. Knight, A high-resolution direct-drive diffraction grating rotation system, *Hewlett-Packard J.*, **44**(6): 75-79, 1993.
5. F. V. Kowalski, R. T. Hawkins, and A. L. Schawlow, Digital wavemeter for cw lasers, *J. Opt. Soc. Amer.*, **66**: 965-966, 1976.
6. A. Zarka et al., Intracavity iodine cell spectroscopy with an extended-cavity laser diode around 633 nm, *IEEE Trans. Instrum. Meas.*, **46**: 145-148, 1997.
7. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Oxford, UK: Pergamon, 1980.
8. J. L. Hall and S. A. Lee, Interferometric real-time display of cw dye laser wavelength with sub-Doppler accuracy, *Appl. Phys. Lett.*, **29**: 367-369, 1976.
9. A. Kahane et al., Vernier fringe counting device for laser wavelength measurements, *Rev. Sci. Instrum.*, **54**: 1138-142, 1983.
10. J. Ishikawa, N. Ito, and K. Tanaka, Accurate wavelength meter for cw lasers, *Appl. Opt.*, **25**: 639-643, 1986.
11. J. Ishikawa, N. Ito, and A. Morinaga, Accurate wavelength measurement of the intercombination line of calcium, *Jpn. J. Appl. Phys.*, **33**: 1652-1654, 1994.
12. K. D. Moller, *Optics*, Mill Valley, PA: University Science Books, 1988.
13. P. Jacquinet, P. Juncar, and J. Pinard, Motionless Michelson for high precision laser frequency measurement, in J. L. Hall and J. L. Carlsten (eds.), *Laser Spectroscopy III*, Berlin: Springer-Verlag, 1977.
14. J. J. Snyder, Fizeau wavemeter, *Proc. SPIE*, **288**: 258-262, 1981.
15. M. B. Morris, T. J. McIlrath, and J. J. Snyder, Fizeau wavemeter for pulsed laser wavelength measurement, *Appl. Opt.*, **23**: 3862-3868, 1984.
16. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991.
17. R. Salimbeni and R. V. Pole, Compact high-accuracy wavemeter, *Opt. Lett.*, **5**: 39-41, 1980.

18. L. J. Radziemski, R. W. Solarz, and J. A. Paisner (eds.), *Laser Spectroscopy and Its Applications*, New York: Dekker, 1987.
19. F. J. Duarte (ed.), *Tunable Lasers Handbook*, San Diego, CA: Academic Press, 1995.
20. I. P. Kaminow and T. L. Koch (eds.), *Optical Fiber Telecommunications III*, San Diego, CA: Academic, 1997.
21. G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*, 2nd ed., New York: Van Nostrand-Reinhold, 1993.
22. M. de Labachellerie et al., The frequency control of laser diodes, *J. Phys. (France), Sect. 3*, **2**: 1557–1589, 1992.
23. T. Ikegami, S. Sudo, and Y. Sakai, *Frequency Stabilization of Semiconductor Laser Diodes*, Norwood, MA: Artech House, 1995.
24. M. Ohtsu (ed.), *Frequency Control of Semiconductor Lasers*, New York: Wiley, 1996.
25. M. Guy et al., Simultaneous absolute frequency control of laser transmitters in both 1.3 and 1.55 μm bands for multiwavelength communication systems, *J. Lightw. Technol.*, **14**: 1136–1143, 1996.
26. J. C. Braasch, W. Holzapfel, and S. Neuschaefer-Rube, Wavelength determination of semiconductor lasers: Precise but inexpensive, *Opt. Eng.*, **34**: 1417–1420, 1995.
27. Documents concerning the new definition of the metre, *Metrologia*, **19** (4): 163–178, 1984.
28. M. Kourogi, K. Nakagawa, and M. Ohtsu, Wide-span optical frequency comb generator for accurate optical frequency difference measurement, *IEEE J. Quantum Electron.*, **29**: 2693–2701, 1993.
29. K. Nakagawa et al., Accurate optical frequency atlas of the 1.5- μm bands of acetylene, *J. Opt. Soc. Amer. B*, **13**: 2708–2714, 1996.

MARTIN GUY
Institut National d'Optique
CHRISTINE LATRASSE
Laval University
BERNARD VILLENEUVE
Nortel
MICHEL TÊTU
PIERRE TREMBLAY
Laval University

REFLECTOMETERS, TIME-DOMAIN

Reflectometers are typically applied in high-frequency, microwave, millimeter, and optical circuits and systems to test discontinuities of the signal transmission channels. The first type, the frequency-domain reflectometer (FDR), determines the incident and reflected powers to compute reflection coefficients in terms of frequency. The frequency-domain measurements do not directly distinguish the location or the type of components introducing discontinuities. Some frequency-domain reflectometers measure the ratios of reflected and incident waves for known termination and for components under test within prescribed frequency ranges (1). The amplitude data points are registered and processed for each frequency to yield computed magnitudes and phases of individual reflections related to the distances along the tested components. The frequency information is then converted into time-domain pulses. The second type, the time-domain reflectometer (TDR), can measure the location and evaluate the type of the discontinuity, and also disclose multiple defects that would appear only through their accumulated effects if the time-domain relations were not available. The principles of classical time-domain reflectometry are based on observations of the reflections from the transmission

pathway superimposed on the incident wave, which is usually a fast steplike signal. Another class of time-domain reflectometry utilizes synthetic pulses forming short sinusoidal bursts (2). The reflectometers based on this idea analyze networks within desired frequency bands. The waves, incident and reflected, are acquired by high-frequency samplers, analyzed, and processed.

A typical measurement setup is composed of a fast pulse generator and a fast operating oscilloscope, usually a digitizing sampling oscilloscope to register the signals. Such a setup delivers broadband information about the behavior of the tested system. Due to the low level of high-frequency components of the pulse-step spectrum used in time-domain reflectometers certain narrowband discontinuities may not be observed. In those cases, frequency-domain reflectometry can give more reliable answers, thanks to the even distribution of frequency components in a sweeping signal.

In this article, only time-domain reflectometry (TDR) and reflectometers are discussed.

PRINCIPLES OF TIME-DOMAIN REFLECTOMETRY

Step-Pulse Time-Domain Reflectometry

Figure 1 shows the basic structure of a TDR with a high-input-impedance oscilloscope. A step generator sends its pulse into the transmission system under test. The step generator's impedance matches that of the transmission-line connection, TL, so the only reflections expected are from the circuit under test. The incident wave, which is one-half the open-circuit step voltage, is transmitted toward the circuit under test. When this signal encounters any discontinuity, including unmatched impedance of the circuit under test, a reflected voltage wave is produced, and it is propagated back to the generator and added to the incident wave. The reflected wave can be positive or negative.

The waves shown in Fig. 2 include the incident wave and the reflected wave that comes from the pure resistive load. The delay time, t_D is related to the distance between the oscilloscope connection to the transmission channel and the single reflecting discontinuity of the channel. The amplitude character of the reflection and the location of the reflecting object can be determined from the following expressions:

$$R = \frac{E_r}{E_i} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (1)$$

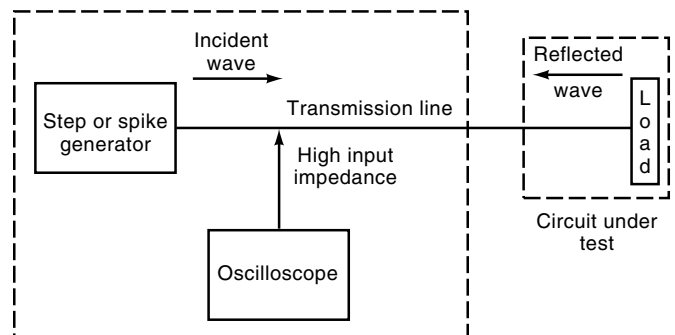


Figure 1. Basic structure of TDR.

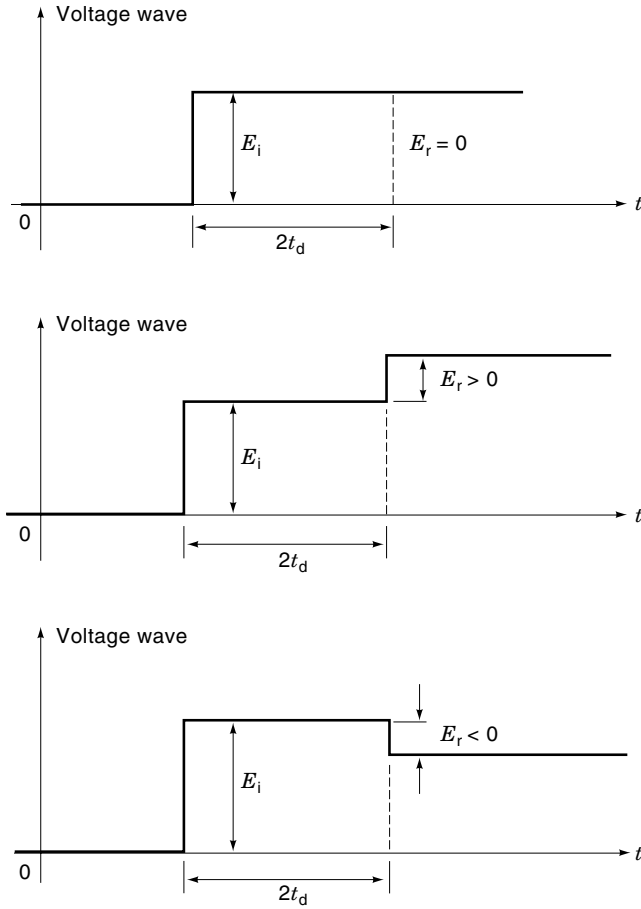


Figure 2. Simplified TDR displays for nonreflective and reflective loads.

where R is the reflection coefficient, and Z_L and Z_0 are the load impedance and the transmission-line characteristic impedance respectively;

$$d = 0.5v_d t_{d0} \quad (2)$$

where d is the distance, v_d is the wave velocity, and t_{d0} is observed delay time. For a transmission line, the velocity of the wave is

$$v_d = \frac{c}{\sqrt{\mu\epsilon}} \quad (3)$$

where c is the light velocity in vacuum, μ is the magnetic permeability of the medium, and ϵ is the electric permittivity of the medium.

The waves of Fig. 2 are related to single, pure resistive discontinuities. The transmission-line reflections produced by complex loads are shown in Fig. 3.

In both cases considered so far to illustrate reflections from single discontinuities, an ideal step wave has been applied. The general shape of the waves is preserved when real step signals with finite risetimes are applied. The reflected wave transitions are smoother and the step-like sections are shorter, due to the transition parts.

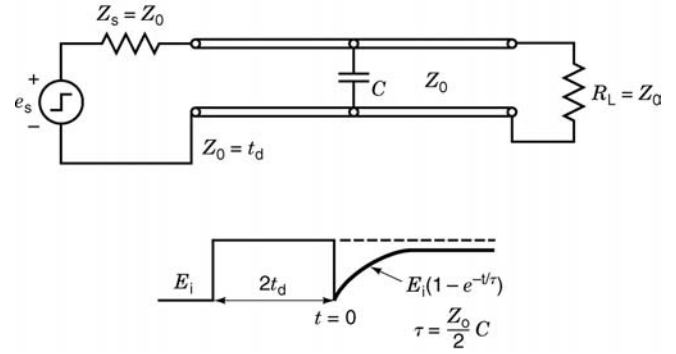


Figure 3. Displays for ideal TDR and various complex loads.

Inductance effects are observed in the response of an open circuit to very sharp changes of the input signal. The response of inductance to the slow signal changes is revealed in exponential signal traces whose time constants are given by the ratios of the inductance to the resistance observed at the terminals of the inductor. Capacitance, on the other hand, creates a short-circuit connection for very sharp signal changes, while the slow changes are controlled by a time constant that is the product of the capacitance and the resistance observed at the terminals of the capacitor.

In all the time-domain test examples mentioned, it is important to define a reference plane so all distances can be measured in relation to this plane. The reference planes chosen in Figs. 2 and 3 are at the points where the oscilloscope-input terminals are connected. For simplicity of presentation the exponential parts of the waves are referred to their own origins ($t = 0$).

The information about reflections in circuits with transmission lines helps detect discontinuities in communication networks as well as determine characteristic impedances of the transmission lines. Figure 4 shows typical reflections for the case when two sections of the transmission lines are connected in such a way that the first section matches the standard impedance of the time-domain reflectometer, and the second, open-ended, produces reflections allowing one to determine its characteristic impedance. The characteristic impedance of the second section can be calculated from the second-step voltage level, which is

$$V_s = E_i \left(1 + \frac{Z_{02} - Z_{01}}{Z_{01} + Z_{02}} \right) \quad (4)$$

For example, if $E_i = 0.5$ V, $V_s = 0.6$ V, and $Z_{01} = 50 \Omega$, then $Z_{02} = 75 \Omega$ ($Z_{02} > Z_{01}$).

Usually TDRs compute desired quantities, including the discontinuity location, by analyzing the character of the reflected waves.

Short-Pulse Time-Domain Reflectometry

Short-pulse time-domain reflectometry uses very narrow pulses, and reflections are observed as waves occurring after the incident pulses are observed. The principles of the operation resemble those of radar. This type of reflectometry has found its main applications in long-distance cable tests,

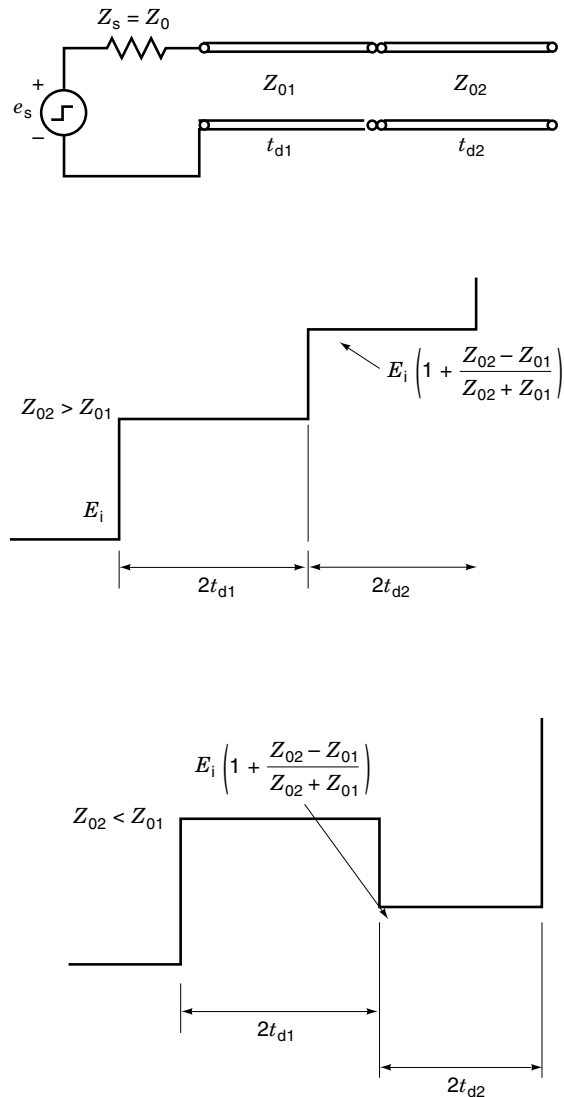


Figure 4. TDR used to determine cable impedance.

including optical fiber probing, which can measure distances from millimeters to tenths of kilometers with resolution better than 1 mm. The reflectometers can also measure the losses and other parameters of the transmission communication channels.

Figure 5 shows typical waveforms with reflections when the incident and reflected pulses are close one to each other but the pulses are separated. When the two pulses are close and overlap, signal processing in the reflectometer can still separate the waves and determine the required parameters. In optical TDRs the optical signals are converted into corresponding electrical waves, which are coherently sampled and processed in a digitizing sampling oscilloscope (3). An example of waves that overlap is shown in Fig. 6. The wave, which is sent into an optical channel, is memorized before the test starts, and it is used as a reference in signal processing.

Applications of narrow test pulses are very convenient in finding faults on long transmission lines designed for telephone, television, and computer networks. High resolution in time is achieved by applying very short pulses with duration

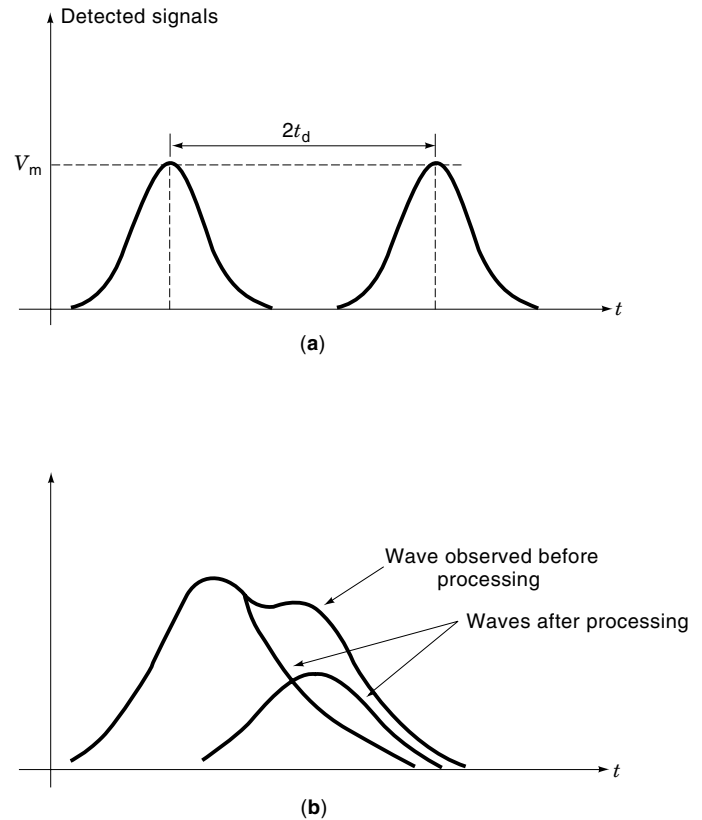


Figure 5. Signals received and registered by an optical TDR: (a) incident and reflected waves are separated; (b) incident and reflected waves overlap.

of several nanoseconds. High resolution to detect small amplitudes of reflected waves is secured by using high-sensitivity and low-noise digitizing sampling oscilloscopes with averaging to reduce amplitude noise and jitter (4).

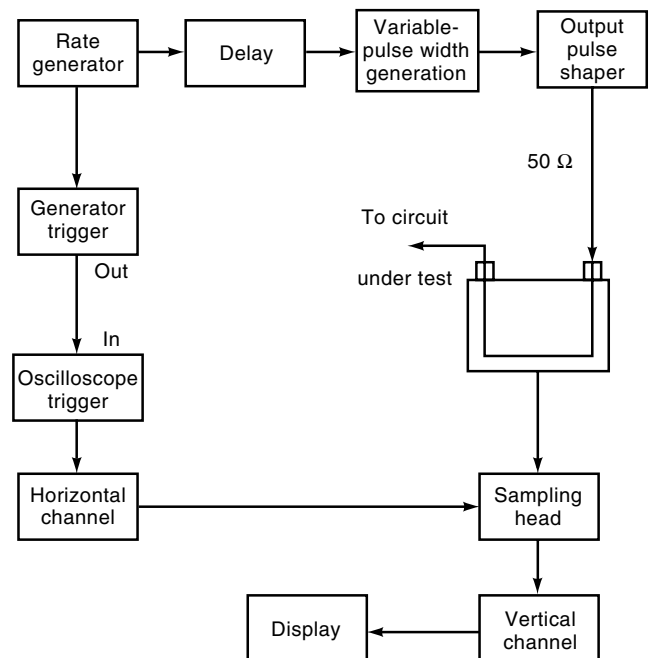


Figure 6. Block diagram of a TDR.

Differential Time-Domain Reflectometry

Differential time-domain reflectometry and reflectometers have been developed to investigate properties of transmission channels that are symmetric in relation to the ground plane. Two separate pulses are generated and applied to the system under test.

Two identical pulses of the same polarity are used to test reflections related to the common mode of the signal transmission. The same pulses but of opposite polarity are used to test reflections of the differential mode of the signal transmission. Both pulses must either coincide or be staggered with great precision, and then the two responses are processed separately and the results referred to the same time origin as if the time-delayed pulses were synchronous (5).

Synthetic Pulse-Domain Reflectometry

Synthetic pulse time-domain reflectometry employing controllable-band synthesized bursts has been developed to characterize microwave bandpass components. The spectrum of the testing signal is adjusted to cover the spectrum of the devices under test. This technique, supported by software signal processing, brings sufficient information to determine magnitude and phase of reflections. The synthesized signals are composed of sinusoidal signals whose amplitudes and phases can be controlled to achieve the best resolution (1).

EXAMPLES OF TIME-DOMAIN REFLECTOMETERS

TDRs can be categorized into two major groups according to their applications. The first group includes field time-domain

reflectometers applied to test cables as well as fiber-optic links. The second group includes laboratory time-domain reflectometers used to test microwave system components.

The first type, with low resolution (down to about 30 cm), uses a cosine-shaped pulse to detect discontinuities located at large distances. The pulse widths are adjustable from nanoseconds to about 1 μ s, and the distances covered can reach 17 km. The system internal impedances are 50, 75, and 100 Ω , including cable impedances of symmetric cables for television, telephone, and computers. The pulse amplitudes are about 10 V. The time resolution in such systems is usually limited by the narrowest pulse width.

A typical laboratory time-domain reflectometer system applied to test short-distance transmission line channels can be configured by connecting a fast pulse generator and a broadband digitizing sampling oscilloscope to a high-precision transmission line, which is connected to a device under test. The maximum bandwidth of the available sampling oscilloscopes reaches 50 GHz (4,6). Flat-top pulse generators, such as the HP 4015B (4) have risetimes better than 15 ps and low overshoot (less than 1%). The overall risetime of the system is close to 17 ps. The pulses, of amplitude close to 10 V are shaped by the step-recovery diodes. Some fast pulse generators generating small-amplitude pulses with the amplitudes of several hundred millivolts employ tunnel diodes. Two-channel operation allows single-ended and differential time-domain reflectometric tests. The noise levels are close to 1 mV, which determines the minimum value of the reflection coefficient that can be detected: about 1% for small pulse amplitude units, and about 0.1% for large pulse amplitudes if other pulse distortions are ignored or corrected for by digital processing.

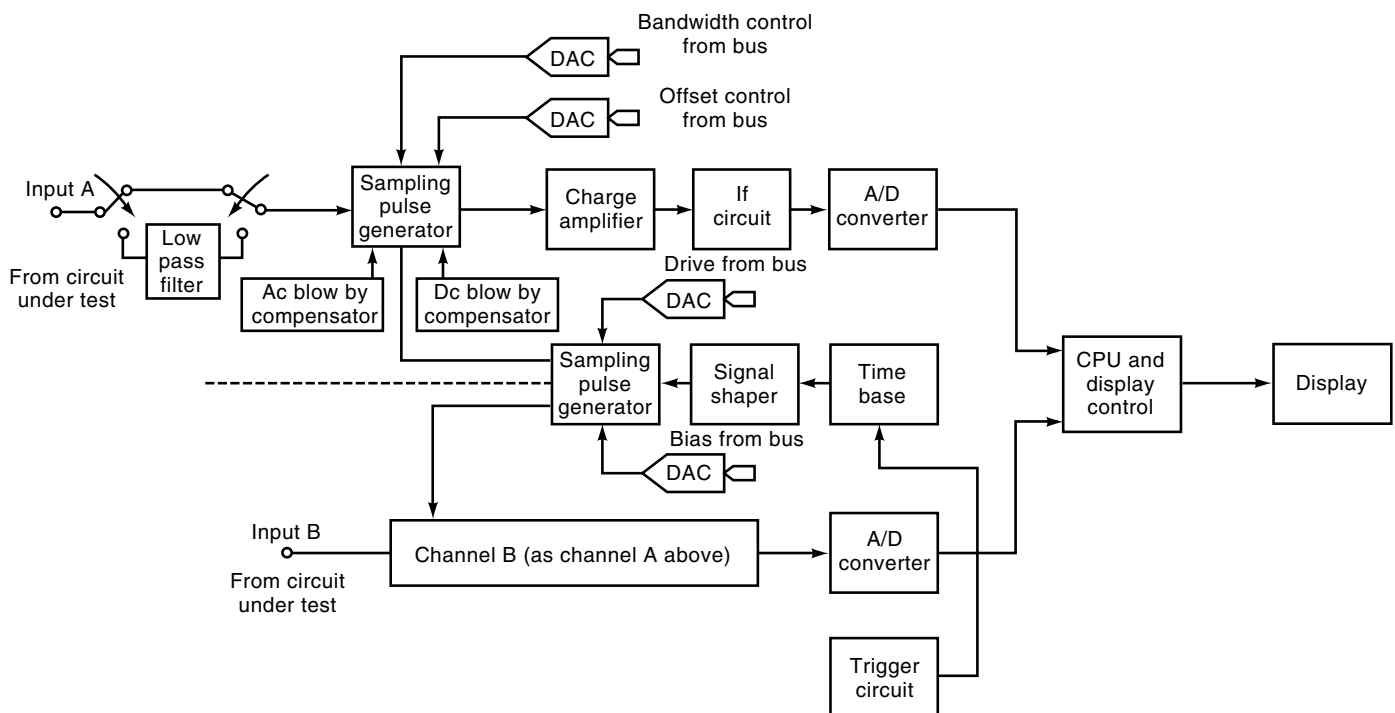


Figure 7. Simplified diagram of a two-channel digitizing sampling oscilloscope used in a TDR system.

The digitizing sampling oscilloscopes used in the TDRs resemble the typical sampling oscilloscopes described in the article Oscilloscopes. A more detailed block diagram of a typical TDR is shown in Fig. 7. The reflectometers briefly described here can be also used as optical time-domain reflectometers if a fast-operating laser transmitter and a fast optical-to-electrical converter are supplied.

A block diagram of the converter applied in the HP 83481A unit is shown in Fig. 8 (3,5). The photodiode receives the light beam from the fiber through a lens. The light signal, converted into an electrical current, is translated into the input voltage of the low-output-impedance driver. The driver output with an added resistor matches the typical $50\ \Omega$ transmission channel and input of the sampling oscilloscope. Optical time-domain reflectometers, such as the HP 8146A (6) include lasers operating at wavelengths of 850 nm, 1300 nm, or 1550 nm. The light pulses from the laser diode are sent to the fiber optic link under test through a 3 dB optical coupler. The light passing through the link is partly reflected and scattered towards the transmitter-receiver unit (Fig. 9). The return light passes to the receiver through the coupler and is converted into electrical signal for further processing in a sampling oscilloscope (3).

A typical image observed on the screen of an optical TDR is shown in Fig. 10 for a case without reflections. The straight-line section represents continuous fiber backscatter. The slope of the line shows the amount of attenuation. Fiber reflections due to the nonideal connectors, splices, and cracks in the fiber are shown as peaks and steps modifying the straight-line part of the response, as shown in Fig. 11.

Figure 12 shows SPICE simulations illustrating operation of a synthetic pulse time-domain reflectometer. In the circuit, the input waves are synthesized using two sinusoidal signals of two different frequencies, 1.25 GHz and 2.5 GHz gated by 2 ns pulses. The shape of the signal resembles the waves applied in Ref. 1. The reflected waves, separated from the incident waves in a simple signal processor, are observed for broad and narrow band loads connected to a coaxial cable. In Figure 12, the reflected waves are delayed and inverted in phase due to low resistance of a broadband load. The narrow band loads introduce oscillatory responses when the resonance frequencies of the loads are centered within the band of the input waves.

In the reflectometer described in Ref. 1, the Hewlett-Packard Automatic Network Analyzer was used to generate harmonically related signals to 40 MHz within the range of 400 MHz to 3560 MHz. The analyzer also served as a receiver processing the received information in frequency domain, and converted it into time-domain through Fourier transformations.

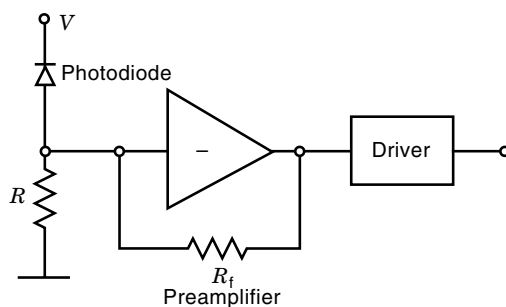


Figure 8. Photo-optical converter with amplifiers.

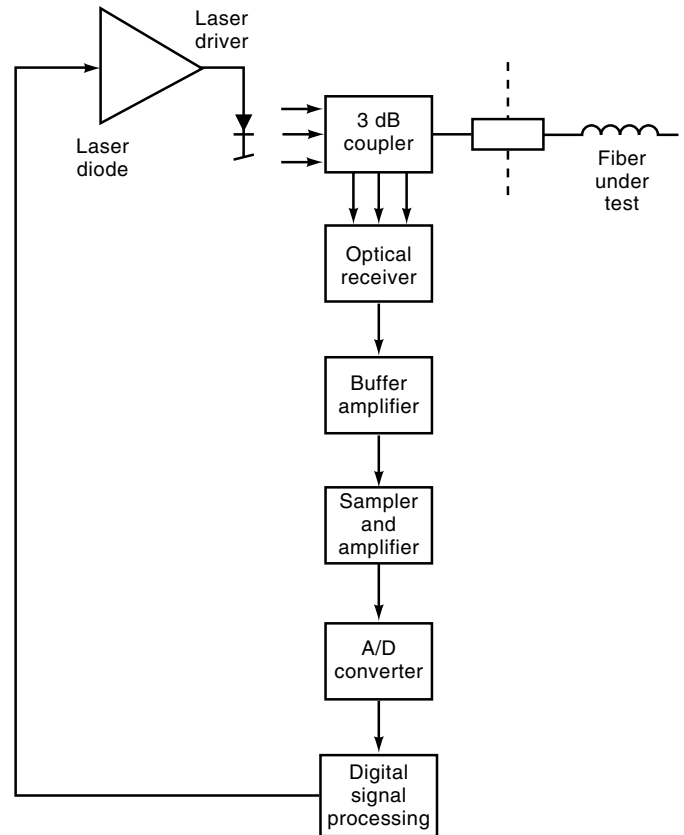


Figure 9. Simplified block diagram of HP 8146A optical TDR.

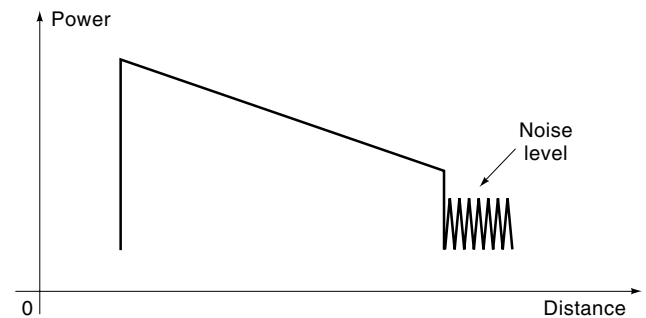


Figure 10. Display of an optical TDR for nonreflective loads.

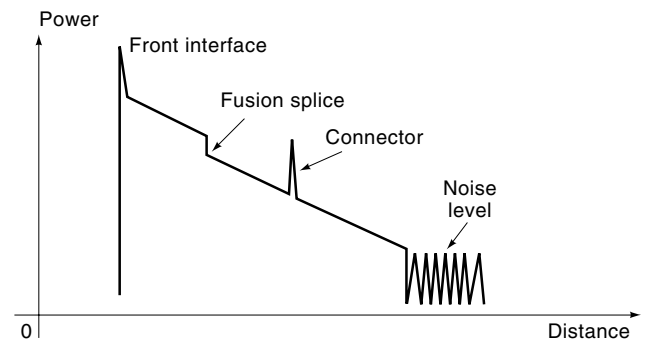
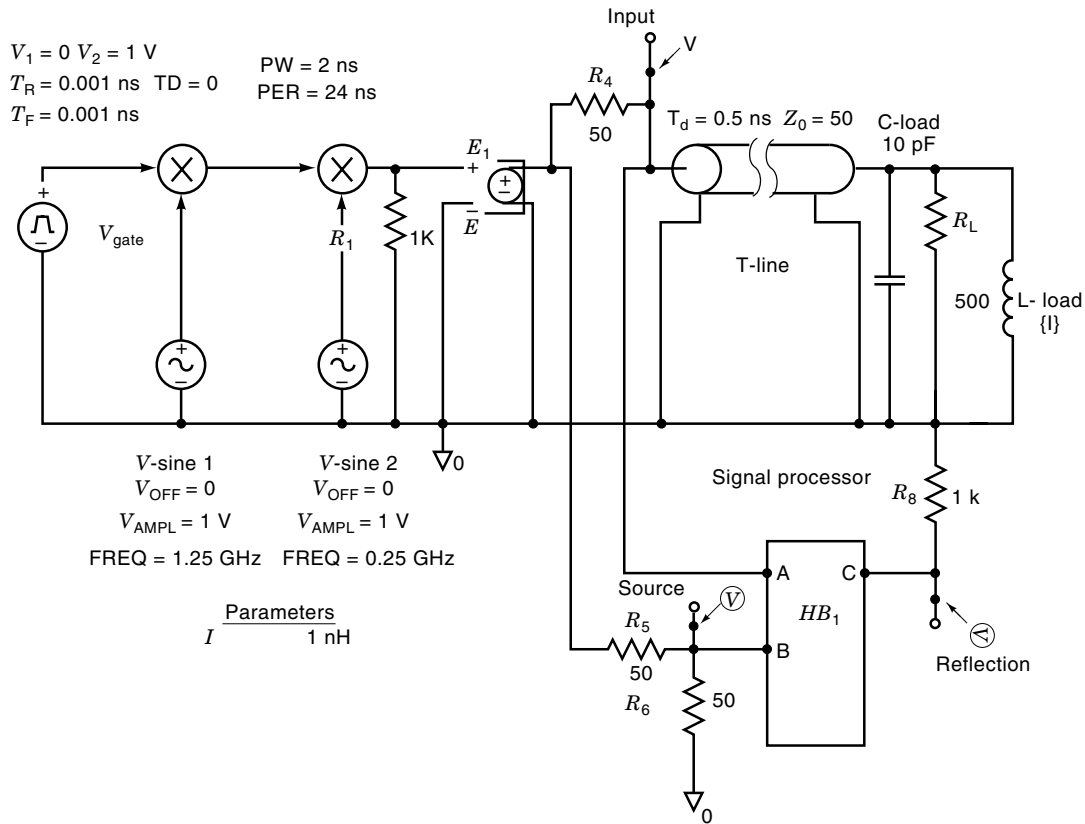
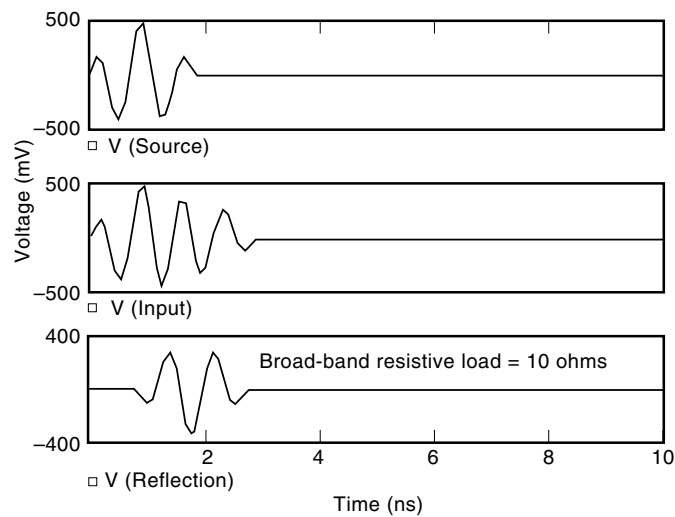


Figure 11. Display of an optical TDR for reflective loads.



(a)



(b)

Figure 12. Simulations of a synthetic pulse time-domain reflectometer. (a) Diagram, (b) waves for broadband load, (c) waves for narrow band loads, (d) spectrum of the source wave.

BASIC CHARACTERISTICS OF THE TIME-DOMAIN REFLECTOMETERS

As mentioned before, there are two major characteristics of TDRs. First is the ability to distinguish two consecutive reflections, representing the time resolution, which is a direct function of the system risetime. Two reflected waves merge together

when they are located at a distance corresponding to less than one-half the system risetime. The computer signal processing available in most modern reflectometer systems makes it possible to incorporate the system imperfections into the compensating process to correct overall system characteristics. In this way, the measurement results approach the behavior of an ideal system, thus improving the time resolution. This characteristic

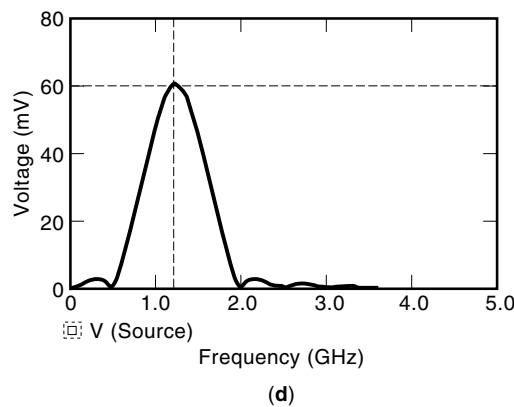
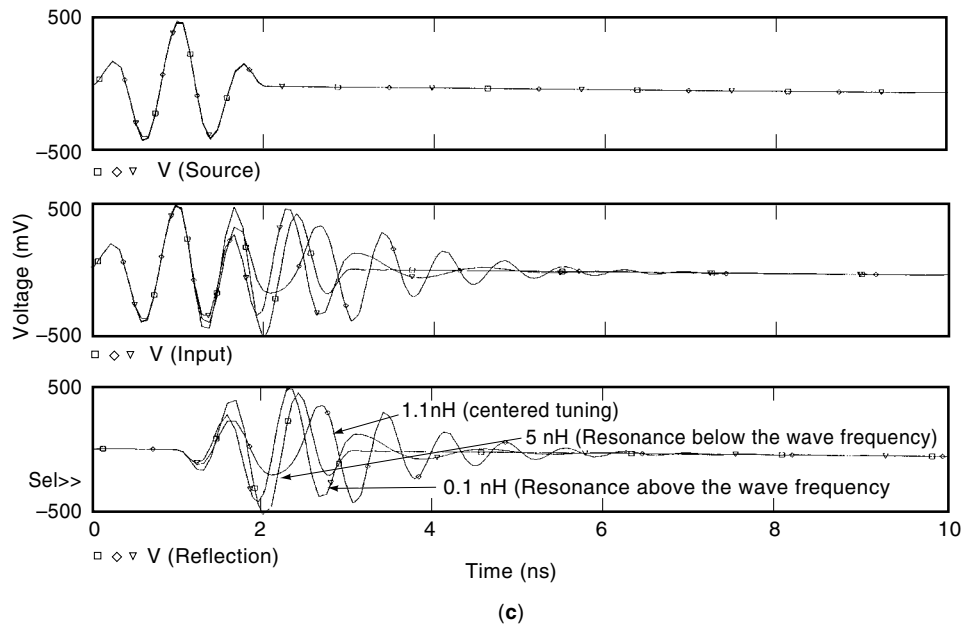


Figure 12. (Continued)

is strongly dependent upon the system noise level, observed in as jitter.

The second characteristic is the amplitude resolution, which expresses the ability to register small changes in the reflection coefficient. The amplitude resolution is limited by the signal-to-noise ratio of the system. The noise level registered in the system can be greatly reduced by analog or digital averaging or other signal processing. Obviously, the signal processing requires longer times of data collection and higher densities of sampling.

EXAMPLES OF APPLICATIONS OF TIME-DOMAIN REFLECTOMETERS

A number of examples described below illustrate how the character and the location of various discontinuities can be found. Investigation of the reflected waves yields additional information about transmission channel losses, which are more visible in systems with longer cables.

A typical example of a response of a transmission line is shown in Fig. 13. The first two plots show the response of

a lossless transmission line with a short-circuit termination. The third shows the pulse distortions introduced by the line losses. The fast changes are attenuated more by the skin effect than slower changes, which indicate the effects of low-frequency losses. The response does not allow evaluating the losses related to specific frequencies, which is possible by using FDRs. Fig. 14 shows the effects of the termination resistor leads (7). A microstrip $50\ \Omega$ transmission line is terminated with a $50\ \Omega$ resistor with long leads (2.5 cm). From the amount of reflected amplitude and the speed of the wave decay, it is possible to evaluate the lead inductance, which is about 6 nH when the estimated decay time is 0.3 ns.

The next example consists of the results of simulations and measurements of the T connection of three transmission lines, each having $50\ \Omega$ characteristic impedance. The reflections shown in Fig. 15 help locate the connection and find the length of each section. The reflection coefficient found from the ratio of the voltages yields a $25\ \Omega$ impedance terminating the center node of the transmission lines. The signals observed at the input terminals show an added reflection from the end of transmission line T_3 when the load of this line does not match

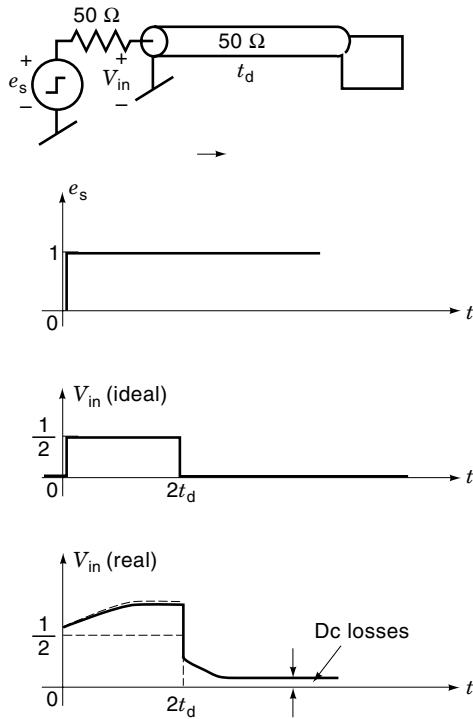


Figure 13. Effects of transmission line losses on reflected signals.

the line impedance. The second reflection reaches the input after 140 ns and represents the doubled transmission time of T_2 and T_3 . The plots of Fig. 15(b) were obtained with the help of HPVEE software (8).

The example of Fig. 16(a) illustrates reflections from different discontinuities of a strip line with a short section of narrow

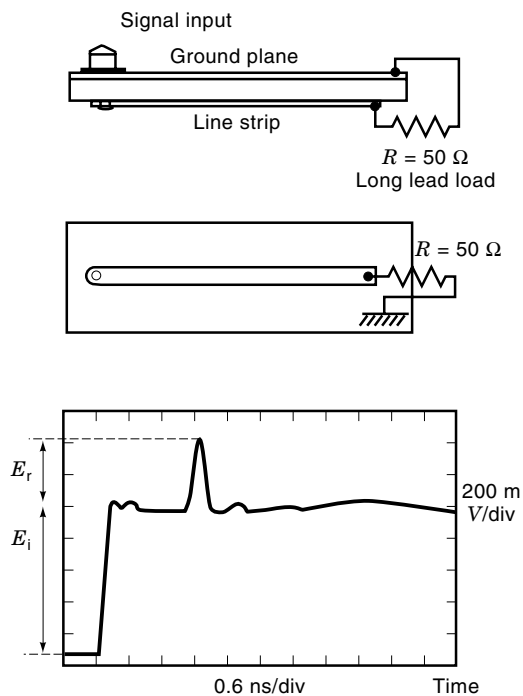


Figure 14. Effects of long resistor leads.

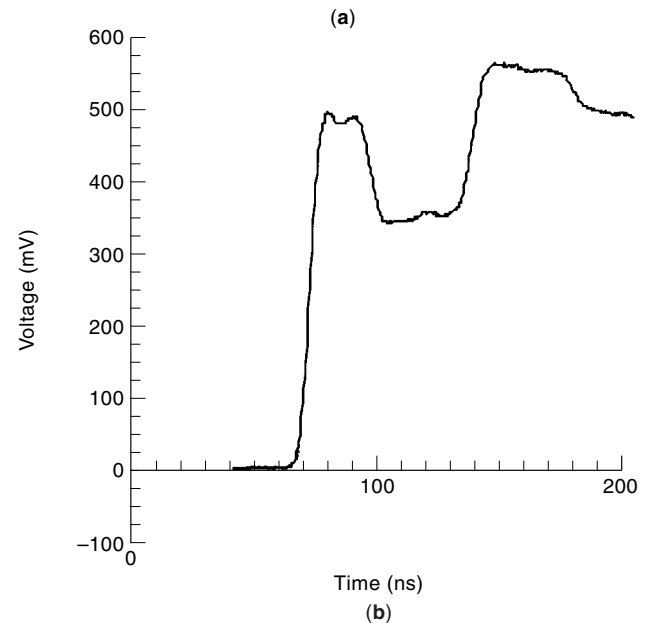
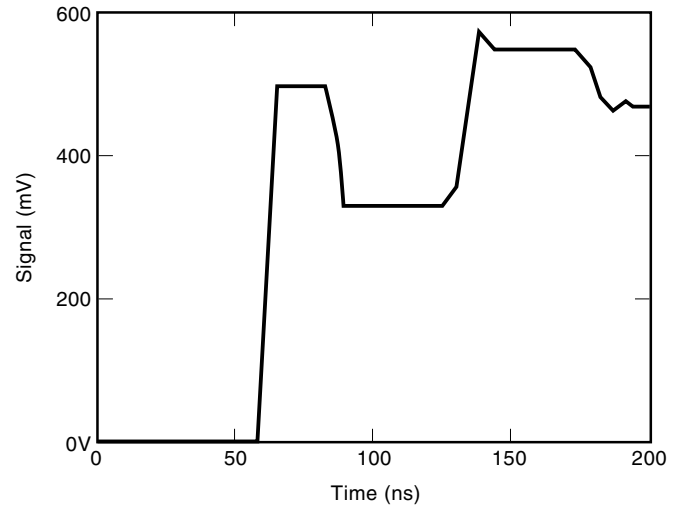
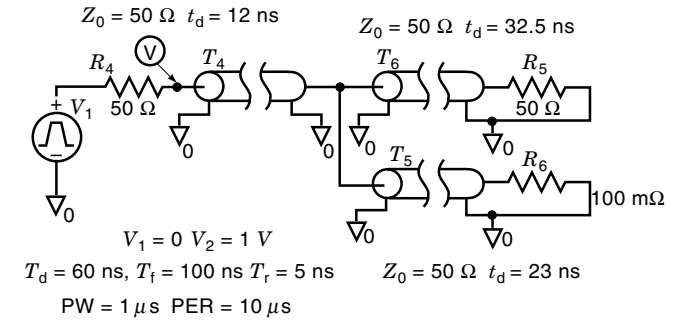


Figure 15. Reflections from T connection of three cables: (a) simulated, (b) measured.

strip and a short section of broad strip (7). If the discontinuity length is smaller than the shortest wavelength component in the spectrum of the testing pulses, then the discontinuity can be treated as a lumped element and the effects of the sections

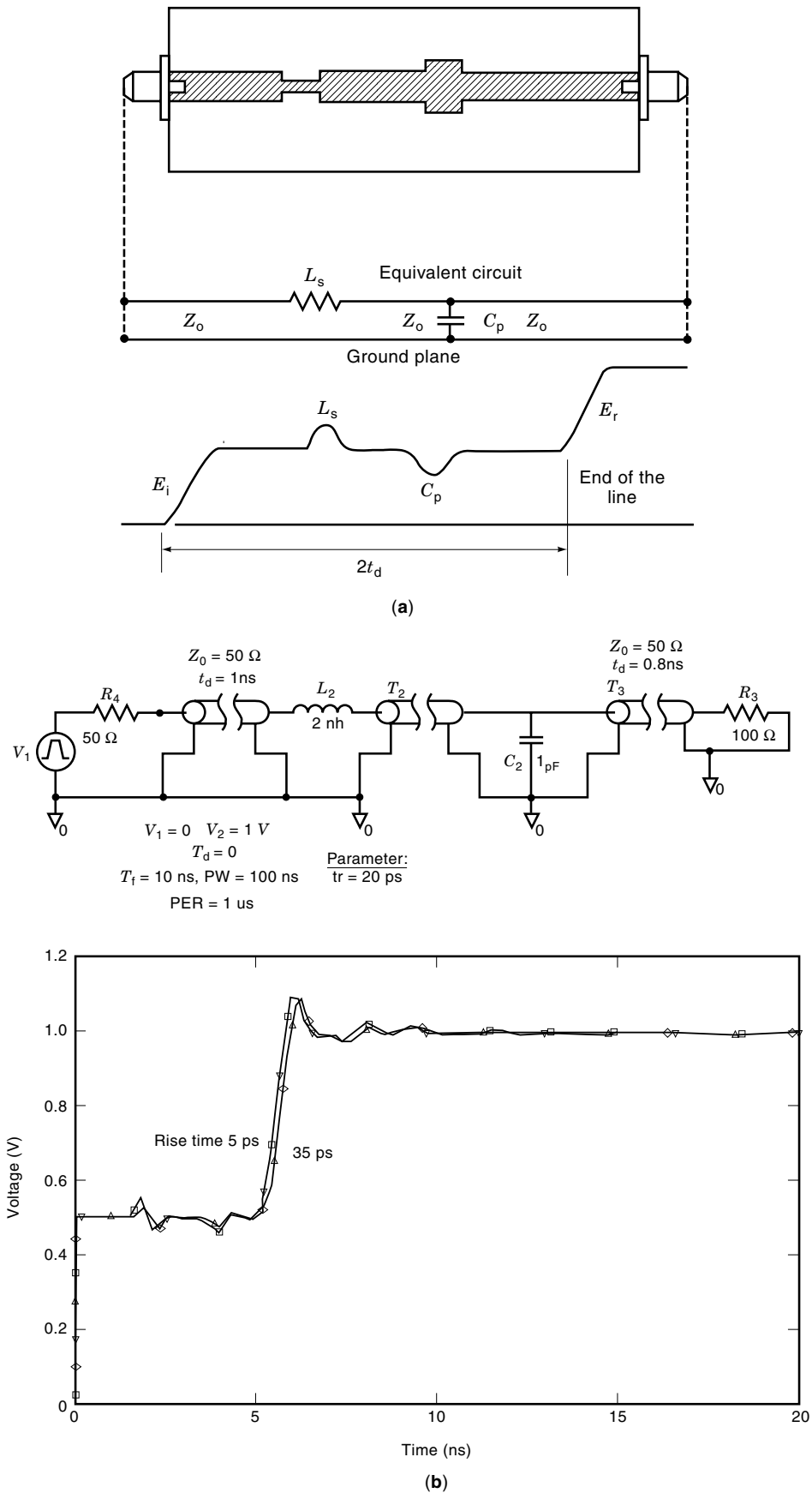


Figure 16. Reflections in the circuit with a strip line: (a) measured, (b) simulated.

interpreted as inductive and capacitive, respectively, as shown in the simulated model of Fig. 16(b) using SPICE (9).

BIBLIOGRAPHY

1. D. L. Holloway, The Comparison Reflectometer, *IEEE Trans. Microw. Theory Tech.* **MTT-15**: 250–259, 1967.
2. L. A. Robinson, W. B. Weir, and L. Young, An RF Time-Domain Reflectometer Not in Real Time, *IEEE Trans. Microw. Theory Tech.* **MTT-20**: 855–857, 1972.
3. J. Beller and W. Pless, A modular all-haul optical time-domain reflectometer for characterizing fiber links, *Hewlett Packard J.*, **44** (1): 60–62, 1993.
4. *Tektronix Instrumentation and Measurement Catalog*, 1998.
5. M. M. McTigue and C. P. Duff, Differential time-domain reflectometry for a digital oscilloscope and communications analyzer, *Hewlett Packard J.*, **47** (1): 1996.
6. *Hewlett Packard Instrumentation and Measurement Catalog*, 1998.
7. TDR fundamentals for use with HP 54120T digitizing oscilloscope and TDR, Hewlett Packard Application Note 62, April 1988.
8. R. Helsel, *Visual Programming with HPVEE*, Englewood Cliffs, NJ: Prentice-Hall, 1977.
9. Design Center, Microsim, 1997.

Reading List

- P. I. Somlo and J. D. Hunter, *Microwave Impedance Measurement*, Oxford, UK: Peter Peregrinus, 1985.
- D. J. Dascher, Measuring parasitic capacitance and inductance using TDR, *Hewlett Packard J.*, **47** (2): 83–96, 1996.
- Advanced TDR techniques, Hewlett Packard Application Note 62-3, May 1990.
- TDR techniques for differential system, Hewlett Packard Application Note 62-2.
- Characterizing IC Packages with TDR/TDT and the UTP-3000 Text Fixture, Hewlett Packard Application Note 1210-5.
- Beginner's guide to using the HP 8147 optical time-domain reflectometer, Product Note, Hewlett Packard, 1996.
- G. Metzger and J.-P. Vabre, *Electronique des Impulsions, Tome II. Circuits à Constantes Réparties*, Paris: Masson, 1966.
- Millimeter resolution OTDR system manual, Optoelectronics Inc., 1988.
- Tektronix Corp., User's Manual of 1S2 Sampling Unit, Section 2.
- M. D. Tilden, Measuring controlled-impedance board with TDR, *Tektronix/Test and Meas.*, February 1992.
- J.-P. Vabre, *Electronique des Impulsions, Tome IV. Lignes Couplées en Régime Transitoire*, Paris: Masson, 1972.

ANDREW RUSEK
Oakland University

STANDING WAVE METERS AND NETWORK ANALYZERS

MEASUREMENT OF NETWORKS AT MICROWAVE FREQUENCIES

Determining the characteristics of electrical networks (network analysis) is an important measurement tool both for

developers of electrical circuits and systems and in production testing. The measurement of network characteristics at microwave frequencies employs different concepts than are used at lower frequencies, where measuring port voltages and currents readily allows determination of impedance (Z) or admittance (Y) parameters. At higher frequencies, neither high-impedance voltage probes nor low-impedance current measurements are feasible, because parasitic capacitance and inductance of probes (L_s and C_p in Fig. 1) cannot be made small enough. Also, the physical dimensions of microwave circuits typically are no longer small with respect to the wavelength. A cable, connecting the port of a device under test with the measurement equipment, may easily transform the intended open circuit into a short, if its length happens to be an odd multiple of $\lambda/4$. Furthermore, short- or open-circuiting network ports, invariably linked to voltage and current measurements, may alter the behavior of microwave circuits (e.g., an amplifier, operated at those impedance extremes, may oscillate). And in waveguides, voltage and current are not even defined.

The above difficulties are circumvented if the network is embedded between defined generator and load impedances Z_0 (typically real and $50\ \Omega$) and described in terms of its power transfer characteristics. The “scattering” of the incident power wave a_1 into the transmitted wave b_2 and the reflected wave b_1 fully characterizes the two-port network of Fig. 1 in the forward direction.

Conceptually, waves represent the propagation of energy along a transmission line. The incident (forward) wave $a = U_f/\sqrt{Z_0}$ travels along the line, with a portion of it ($b = U_r/\sqrt{Z_0}$) being reflected at the line's end. Adding both complex quantities along the transmission line yields a pattern that is referred to as the standing wave (Fig. 2). The ratio of its maximum to its minimum voltage is called the voltage standing wave ratio (VSWR) and is directly linked to the magnitude of the impedance Z_L , terminating the line.

$$\text{VSWR} = \frac{|U_f| + |U_r|}{|U_f| - |U_r|} = \frac{U_{\max}}{U_{\min}} = \begin{cases} \left| \frac{Z_L}{Z_0} \right|, & |Z_L| > Z_0 \\ \left| \frac{Z_0}{Z_L} \right|, & |Z_L| < Z_0 \end{cases} \quad (1)$$

The incident and reflected waves a_i and b_i at a network port (Fig. 1) are defined as linear combinations of the port currents and voltages:

$$a_i = \frac{U_i + I_i Z_0}{2\sqrt{Z_0}}, \quad b_i = \frac{U_i - I_i Z_0}{2\sqrt{Z_0}}$$

making it easy to compute the impedance Z_i at that port:

$$Z_i = \frac{U_i}{I_i} = Z_0 \frac{1 + \frac{b_i}{a_i}}{1 - \frac{b_i}{a_i}} = Z_0 \frac{1 + \Gamma_i}{1 - \Gamma_i}, \quad \Gamma_i = \frac{b_i}{a_i} \quad (2)$$

Γ is called the reflection coefficient. Similarly, the portion of the incident power transmitted from port i to port j is obtained by the transmission factor $t_{ji} = b_j/a_i$.

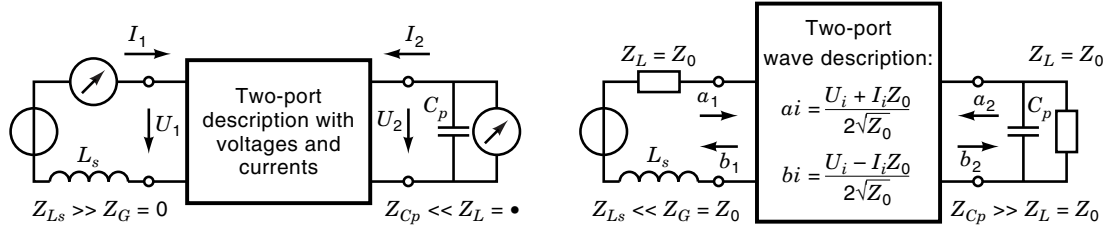


Figure 1. Drawbacks of voltage and current measurement versus the wave concept.

In general, a two-port network is fully characterized by its 2×2 scattering matrix,

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

which is composed of

$$\begin{aligned} \text{Reflection coefficients: } S_{11} &= \left. \frac{b_1}{a_1} \right|_{a_2=0}, \quad S_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0} \\ \text{Transmission coefficients: } S_{21} &= \left. \frac{b_2}{a_1} \right|_{a_2=0}, \quad S_{12} = \left. \frac{b_1}{a_2} \right|_{a_1=0} \end{aligned}$$

The exact determination of scattering parameters, however, still requires the locus, where they are valid, to be fixed. If the measurement port at $x = 0$ in Fig. 2 is shifted to the left, the phase of incident and reflected waves changes and so does the reflection coefficient, which therefore depends on the locus. Especially at small wavelengths, measurement of S parameters stipulates the exact definition of *where* the waves are recorded, leading to the concept of *measurement* or *reference planes*. These planes are thought to perpendicularly intersect the transmission lines leading to the ports of the network, such that the phase of the waves in each plane is constant. For scattering parameter S_{ij} , the incident wave in plane j is thus the reference for the emanating wave in plane i . S_{ij} therefore describes the transmission of energy between those planes.

Since scattering parameters are ratios of emanating and incident waves, their computation does not require absolute measurement of the wave quantities. This important advantage allows the measurement of scattering parameters (with network analyzers) to inherently exhibit much higher precision than can be expected from a spectrum analyzer that relies on absolute measurement of its input power.

The remainder of this article will present a number of measurement systems for the determination of reflection coefficients (one-port network analyzer or reflectometer) and for full

characterization of two-port networks. Although the concept of S -parameters is easily extended to N -ports, measurement of networks with more than two ports will not be covered, because most commercially available hardware is built for two-port measurements and can be used to characterize N -ports by simply terminating the unused $N-2$ ports in Z_0 .

REFLECTION MEASUREMENT TECHNIQUES

The Slotted Line for VSWR and Impedance Measurements

A very old technique for measuring the VSWR on a transmission line, to determine the impedance connected at its end, makes use of a slotted transmission line of characteristic impedance Z_0 (see, for example, Ref. 1, Chap. 2, for a very detailed outline; or see Ref. 2, Chap. 4).

The slot allows a small enough probe, as not to disturb the line's field, to be moved along the x axis of the transmission line and sample its electric field. The high-frequency voltage is converted to direct current (dc) by a diode detector and brought to an instrument for readout. Since the signal level of the probe must be kept low in order to avoid loading of the line, the detector operates in its square-law region and its output is proportional to the detected *power*, or the square of the voltage. Noting the maximum and minimum values of the detector output, while varying the position of the probe, allows computation of the VSWR, provided that the slot is long enough to reach both maximum and minimum.

As indicated by Eq. (1), measurement of the scalar VSWR is not sufficient to uniquely determine the complex impedance Z_L , terminating the line. In fact, since

$$\begin{aligned} \text{VSWR} &= \frac{|U_f| + |U_r|}{|U_f| - |U_r|} = \frac{1 + \left| \frac{U_r}{U_f} \right|}{1 - \left| \frac{U_r}{U_f} \right|} = \frac{1 + \left| \frac{b}{a} \right|}{1 - \left| \frac{b}{a} \right|} \\ &= \frac{1 + |\Gamma|}{1 - |\Gamma|} \rightarrow |\Gamma| = \frac{\text{VSWR} - 1}{\text{VSWR} + 1} \end{aligned}$$

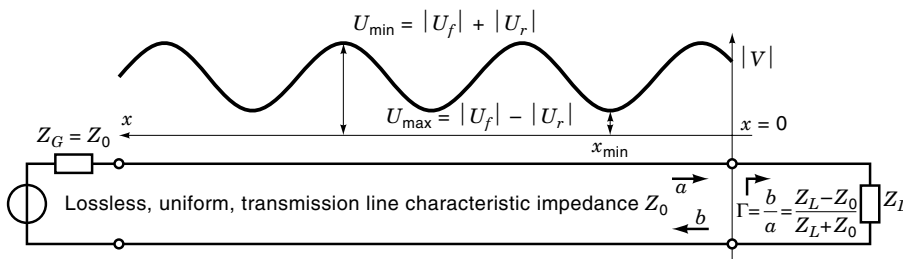


Figure 2. Standing waves along a transmission line, terminated in Z_L .

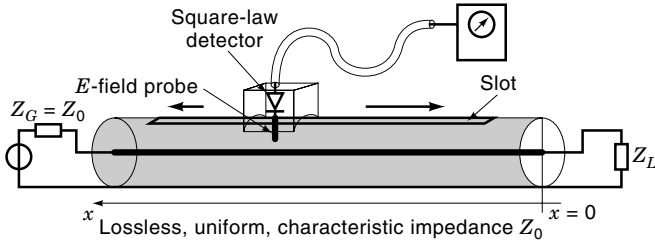


Figure 3. The slotted line measurement setup.

only the *magnitude* of the reflection coefficient Γ is available. The *phase* of Γ is obtained by noting the *position* $x_{\min}$ of the first minimum, appearing when the probe is moved away from Z_L . With the propagation constant

$$\beta = \frac{2\pi}{\lambda_{sl}}$$

of the slotted transmission line, the phase of Γ becomes

$$\varphi = 2\beta x - \pi$$

such that

$$\Gamma = |\Gamma|e^{j\varphi} = \frac{VSWR - 1}{VSWR + 1} e^{j(2\beta x - \pi)}$$

and the complex impedance Z_L may be computed from Γ using Eq. (2).

Though conceptionally simple, the slotted line technique suffers from a number of drawbacks and limitations:

1. The slotted transmission line itself is a costly precision device, exhibiting precise line impedance Z_0 , low losses, negligible loading due to the probe, and mechanical stability of the probe depth over the entire moving range.
2. Linearity errors and the limited dynamic range of the detectors reduce the accuracy when measuring large VSWRs. By modulating the RF source and selective reception of the modulation frequency, noise can be reduced, increasing sensitivity and thus measurement dynamic (the so-called "modulated frequency technique").
3. The measurement procedure requires manual interaction and, in its simple form of Fig. 3, does not allow for swept frequency impedance measurements. If, however, only the magnitude of the reflection coefficient is of

interest, Sorger (3) describes a setup that uses two slotted lines for swept frequency VSWR measurement.

Because slotted line reflection measurements are a very old technique, much research has been undertaken to alleviate the above limitations. Many ideas and extensions can be found in Ref. 2, covering the topic in more detail.

Using the Slotline with Multiple Probes

The biggest drawback of slotline measurements, the need for manual interaction, can be overcome if more than one probe taps to the field of the line in known and fixed positions. Figure 4 shows a system proposed by Caldecott (4) that uses three probes in arbitrary but known positions x_1 , x_2 , and x_3 . Caldecott proved that

$$\Gamma_L \approx \frac{P_1(e^{-j2\beta x_2} - e^{-j2\beta x_3}) + P_2(e^{-j2\beta x_3} - e^{-j2\beta x_1}) + P_3(e^{-j2\beta x_1} - e^{-j2\beta x_2})}{\sin[2\beta(x_2 - x_3)] \sin[2\beta(x_3 - x_1)] + \sin[2\beta(x_1 - x_2)]}$$

even providing complex measurability of the reflection coefficient, although only scalar power measurements ($P_1 \dots P_3$) are used.

Another interesting multiprobe reflectometer for swept frequency/swept power applications reported in Ref. 5 even provided an automatic Smith-chart display of the measured impedance.

The Six-Port Reflectometer

The idea of the foregoing section can be generalized to the so-called six-port reflectometer, developed by Hoer (6) and Engen (7) in the early 1970s. Figure 5 shows a general six-port junction, driven by a signal generator at port 5, ports 1 through 4 connected to power sensors (diode detectors, bolometers, thermocouple, etc.) and port 6 interfacing to the unknown impedance with reflection coefficient Γ_L .

As the mathematical derivation of the six-port theory is rather lengthy, the interested reader should refer to Refs. 2, 6, or 7 for details. In essence the 6×6 S -matrix of the six-port network is reduced to the form

$$b_i = Q_i a_6 + R_i b_6, \quad i = 1, \dots, 4 \quad (3)$$

by using $\Gamma_i = a_i/b_i$ to eliminate $a_1 \dots a_5$ (b_5 is also eliminated by omitting the 5th equation). The powers P_i are related to b_i by

$$P_i = |b_i|^2(1 - |\Gamma_i|^2)$$

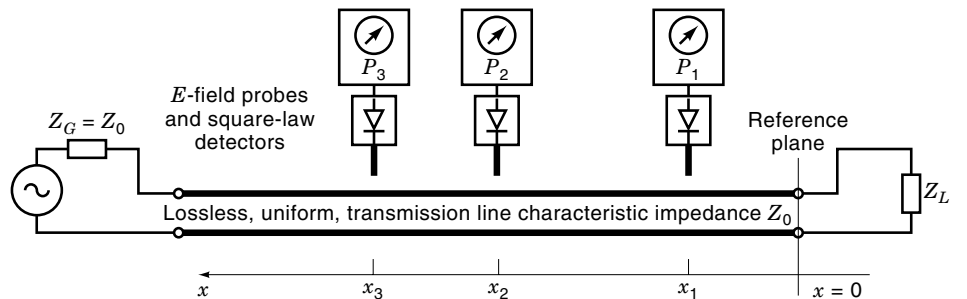


Figure 4. The multiprobe reflectometer.

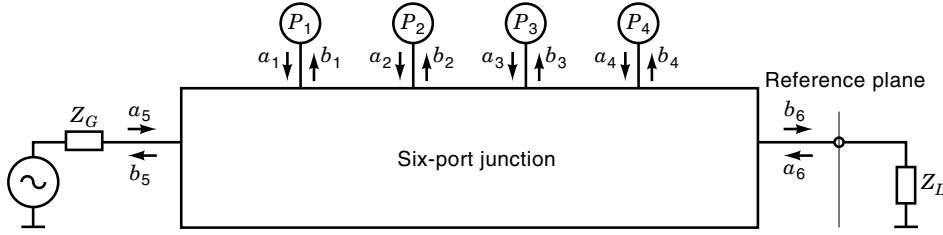


Figure 5. The six-port reflectometer.

and, using Eq. (3), can be expressed as

$$P_i = (1 - |\Gamma_i|^2) \{ |R_i|^2 |b_6|^2 + 2\text{Re}(Q_i^* R_i) \text{Re}(a_6 b_6^*) + 2\text{Im}(Q_i^* R_i) \text{Im}(a_6 b_6^*) + |Q_i|^2 |a_6|^2 \}$$

which may be written in matrix form as

$$(P_1 P_2 P_3 P_4)^T = [D] (|b_6|^2 \text{Re}(a_6 b_6^*) \text{Im}(a_6 b_6^*) |a_6|^2)^T \quad (4)$$

with $[D]$ being a real-valued 4×4 matrix. Inverting Eq. (4) gives access to its right-hand-side vector, provided that $[D]$ is known and invertible with $[C]$ being its inverse,

$$(|b_6|^2 \text{Re}(a_6 b_6^*) \text{Im}(a_6 b_6^*) |a_6|^2)^T = [C] (P_1 P_2 P_3 P_4)^T$$

such that Γ_L can be computed as

$$\begin{aligned} \Gamma_L &= \frac{a_6}{b_6} = \frac{a_6 b_6^*}{b_6 b_6^*} = \frac{\text{Re}(a_6 b_6^*) + j \text{Im}(a_6 b_6^*)}{|b_6|^2} \\ &= \frac{\sum_{n=1}^4 C_{2n} P_n + j \sum_{n=1}^4 C_{3n} P_n}{\sum_{n=1}^4 C_{1n} P_n} \end{aligned} \quad (5)$$

With just 12 of the 16 real-valued elements of $[C]$ appearing in Eq. (5), only 11 need to be computed, as Eq. (5) may be normalized to one of the 12 terms. They are determined by measuring *six known* reflection coefficients and using the complex equation [Eq. (5)] to establish a set of 12 linear, real-valued equations in the 12 unknowns.

Although such a calibration procedure to determine the elements of $[C]$ *must* be carried out before the six-port can be used for reflection measurements, it has the added advantage of providing error-corrected results. No a priori knowledge about the six-port is necessary, because the calibration sufficiently characterizes $[C]$, which reflects the properties of the six-port junction.

Further advantages of this kind of instrument are that neither the signal generator nor the power sensors need to be perfectly matched. The mismatch of the power sensors must however be constant. Because only ratios of measured powers enter the calculation of Γ_L , the power sensors do not need to be absolutely calibrated for power measurements.

However, the power sensors must be linear; and because of their broadband nature, care must be taken to suppress spurious emissions from the signal generator.

Evaluation of Eq. (5) and the determination of $[C]$ require some amount of numerical processing. Six-port reflectometers

are therefore almost always combined with microcontrollers to automate the process of calibration and measurement.

Internally, the six-port must be built such that the reduced matrix $[D]$ is invertible at all measurement frequencies. Details of the design process are outlined in Ref. 8 with a practical realization, using 0° and 90° couplers described in Ref. 9.

Reflection Measurement with the Directional Coupler

All of the impedance measurement methods described thus far evaluate the *sum* of incident and reflected waves, requiring at least two independent measurements for calculation of the reflection coefficient.

Alternatively, forward and reflected waves may be separated using a directional coupler. Such devices are common building blocks in microwave engineering and exist in a wide variety of designs, many of which are treated in Ref. 10, Chap. 8. The interested reader may also refer to Ref. 11 for some less common coupler structures and unusual applications. Figure 6 depicts the associated symbol of a directional coupler and its scattering matrix.

The important property of the directional coupler, used to separate incident and reflected wave is in its decoupled port. Figure 7 illustrates that feeding a wave a_1 into port 1 results in wave $b_2 = \tau a_1$ and $b_3 = \kappa a_1$ being emitted at ports 2 and 3 with no energy being fed to port 4, the decoupled port. Similarly, port 2 is decoupled from port 3. Driving the coupler at port 1, terminating port 3 in Z_0 , and connecting a matched power detector to port 4 allows measurement of unknown reflection coefficients connected to port 2.

Directional couplers, also called “VSWR-bridges” or “reflection coefficient bridges,” appear in various forms, with the microstrip coupler, the resistive coupler, and the Wilkinson divider (Fig. 8) being the variants most often found.

All practical realizations of directional couplers exhibit finite *directivity*, meaning that a small amount of the power fed to port 1 will leak to port 4, even if port 2 and 3 are perfectly terminated in Z_0 . The directivity of a coupler is defined as

$$d = \frac{S_{14}}{\kappa \tau}$$

Mismatch at port 2 and even at the driving port 1 will further add to erroneous reflections being indicated by the wave b_4 ,

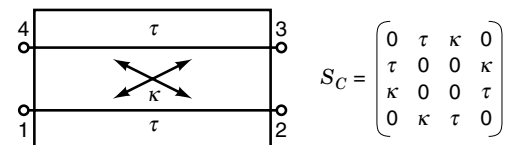


Figure 6. The directional coupler.

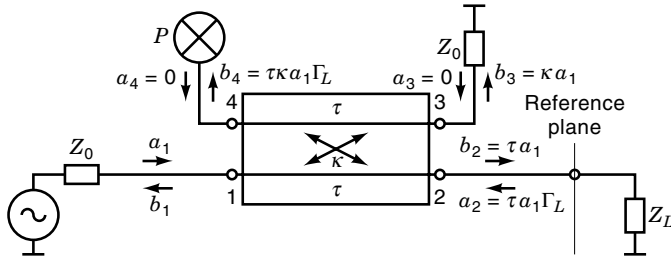


Figure 7. Reflection measurement using a directional coupler.

which ideally should be a measure of the reflection of Z_L only. All these errors must be kept low if the coupler is to be used for precision reflection measurement. Narrow-band precision couplers, exhibiting more than 50 dB directivity, have been built in waveguide and broad-band couplers, typically of the resistive bridge type, and exhibit more than 36 dB directivity over a bandwidth of more than three decades. As a rule of thumb, the directivity should be at least 10 dB better than the expected reflection coefficient of the measured device under test (DUT).

ARCHITECTURES OF NETWORK ANALYZERS

In order to measure the entire scattering matrix of a two-port, some means of measuring the transmitted waves as well as the reflected waves must be provided. Because reflection measurements, nowadays, are almost always carried out with directional couplers, we will use its symbol for denoting measurement of reflected waves. As in Fig. 7, the “mixer” symbol will be used to denote a measurement point, at which the wave to be measured may be further processed in one of the following ways:

1. With power detectors, leading to scalar measurement of the wave
2. With a homodyne receiver, potentially capable of providing amplitude and phase information
3. With a heterodyne receiver (hence the mixer symbol), providing amplitude and phase information

Depending on the type of detector used, network analyzers (NAs) are classified into one of the following types of analyzer:

Scalar Network Analyzer. The simplest and most economic approach to the measurement of scattering-parameters

employs power (diode) detectors. Because these provide no phase information, only the magnitude of the S -parameters can be measured, hence the term *scalar network analyzer* (SNA). For many applications, however, the magnitude may be all that is needed. Calibration techniques, as described in the section entitled “Error Models, Calibration, and Error Correction,” however, require the phase information and consequently cannot be applied to SNAs. Instead, normalization is all that can be employed with such an instrument.

Other drawbacks of SNAs are the limited dynamic range and the broad-band nature of the power detectors. Since they record the entire power of a wave, regardless of its spectral origin, these instruments are susceptible to spurs or harmonics of the generator. If, for example, the harmonics of the radio-frequency (RF) source are suppressed by 30 dB, and the DUT has high-pass or band-stop characteristic (return loss of a filter or an antenna), the SNA will show no less than -30 dB attenuation or return loss, regardless of the DUT’s real properties.

It should be emphasized that the mere use of power detectors does not necessarily yield a scalar network analyzer. The six-port reflectometer, as described in the section entitled “The Six-Port Reflectometer,” employs four scalar detectors to determine the complex reflection coefficient; a complete vector NA can be built by using two such reflectometers (see Refs. 8, 9, and 12), retaining all vector error-correction capabilities of those instruments (13,14), as discussed in the section entitled “Error Models, Calibration, and Error Correction.”

Heterodyne (Vector) Network Analyzer. If the wave quantities are processed by heterodyne receivers, the phase information can be preserved by phase-locking the heterodyne oscillator to the RF source driving the DUT (Fig. 9), and complex measurement of the scattering parameters is possible. Besides providing a much higher dynamic range (over 120 dB for recent industry products versus 70 dB for SNAs), the ability to measure the complex S -parameters has the added benefit of allowing calibration techniques to enhance measurement accuracy. On the other hand, the inherent higher complexity of heterodyne vector network analyzers (VNAs) results in more expensive hardware. For broad-band systems, the heterodyne oscillator and its Phase Locked Loop (PLL) add considerably to the cost, because they must cover the system’s entire bandwidth.

A cost reducing alternative exists by employing “harmonic mixing” (e.g., Ref. 35). In this case the local oscillator (LO) in Fig. 9 has to cover only one octave with all higher frequencies that are needed being supplied by the LO’s harmonics. Harmonic mixing, however, leads to higher mixer conversion loss and thus also reduces the instrument’s dynamic range.

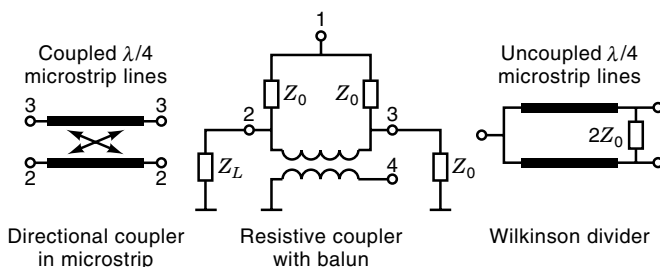


Figure 8. Some commonly used directional couplers.

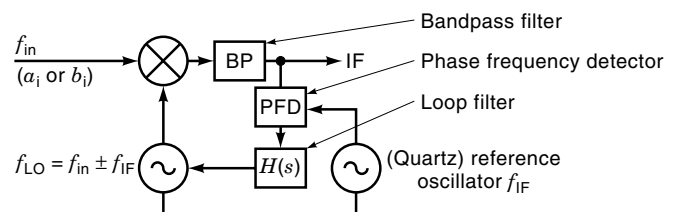


Figure 9. Heterodyne receiver (LO: Local Oscillator, IF: Intermediate Frequency, BP: Bandpass filter, PFD: Phase-Frequency Detector, $H(s)$: Loopfilter).

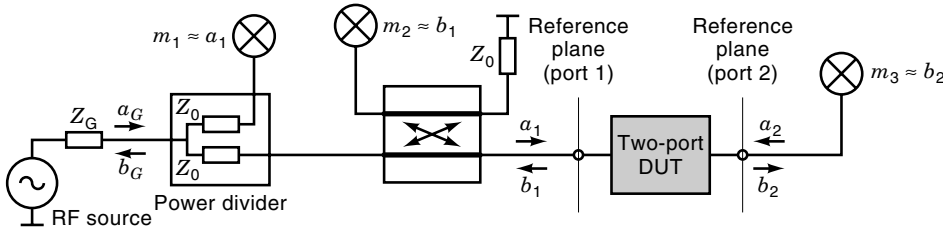


Figure 10. Unidirectional network analyzer.

Homodyne Network Analyzers. In order to circumvent the need for a second microwave source and phase-locking circuitry, homodyne systems use the same generator for down-conversion that drives the DUT. Because the homodyne mixing process provides only the real part of the wave quantity, homodyne NAs need additional hardware (phase-shifters) to extract the complex information from two or more measurements (e.g., Ref. 15). Though relatively simple hardware concepts have been developed (16), homodyne NAs also suffer from their sensitivity to generator harmonics, limiting dynamic range. It is for that reason that homodyne systems are rarely used in the laboratory, although some realizations exist for industrial measurements.

Regardless of the kind of detector used, all practical realizations of the NA architectures introduced in the next sections should be built to provide

- Good port match, properly terminating the DUT in the system impedance Z_0
- High directivity of the directional coupler for precise reflection measurement
- Low variation of transmission losses over the bandwidth of the system

Reflection and Transmission Measurement with the Unidirectional Network Analyzer

Unidirectional network analyzers (Fig. 10) extend the capability of a simple reflectometer to measure S_{11} and S_{21} of a two-port DUT.

With the measured quantity m_1 being proportional to a_1 , m_2 presenting a measure for the reflection b_1 , and m_3 presenting a measure for the transmission b_2 , S_{11} and S_{21} can be expressed as

$$S_{11} = c_{11} \frac{m_2}{m_1}, \quad S_{21} = c_{21} \frac{m_3}{m_1} \quad (6)$$

The proportionality constants are usually obtained through normalization measurements of known reflection/transmission coefficients. Connecting a “Short” with $S_{11} = -1$ to port 1 yields

$$c_{11} = -1 \frac{m_1^S}{m_2^S}; \quad \text{and} \quad c_{21} = 1 \frac{m_1^T}{m_3^T}$$

is obtained by connecting both ports (“Through”), such that $S_{21} = 1$.

For measurement of S_{22} and S_{12} (reverse direction), the DUT must be connected to the unidirectional NA with its ports reversed.

The Three-Receiver Network Analyzer Architecture for Full Two-Port Measurements

The need for manually reversing the DUT can be eliminated if a switch and another directional coupler is introduced into the system of Fig. 11. With the switch in position I, the RF source is connected to port 1 of the DUT, m_1 is proportional to a_1 , m_2 is a measure for b_1 , and m_3 is a measure for b_2 . The second contact of the switch terminates the coupler at port 2 in Z_0 , ensuring proper port match. With the switch in position II, port 2 of the DUT is excited, the coupler connected to port 1 of the DUT is terminated in Z_0 , and m_1 is now a measure for a_2 . Since the three receivers m_1 , m_2 , and m_3 now provide different readings, depending upon the position of the switch, their measurement values will from now on be referred to as m_i^I and m_i^{II} , the superscript denoting the position of the switch. The S -parameters of the DUT are therefore determined as

$$S_{11} = c_{11} \frac{m_2^I}{m_1^I}, \quad S_{21} = c_{21} \frac{m_3^I}{m_1^I}$$

$$S_{22} = c_{22} \frac{m_3^{II}}{m_1^{II}}, \quad \text{and} \quad S_{12} = c_{12} \frac{m_2^{II}}{m_1^{II}}$$

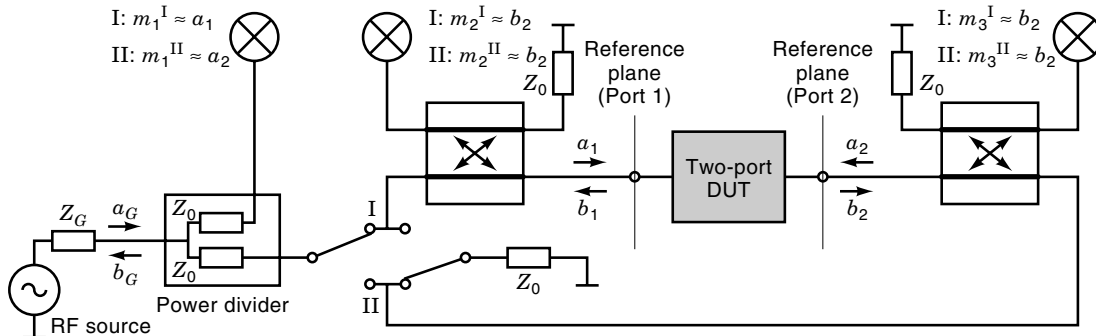


Figure 11. Bidirectional network analyzer with three receivers.

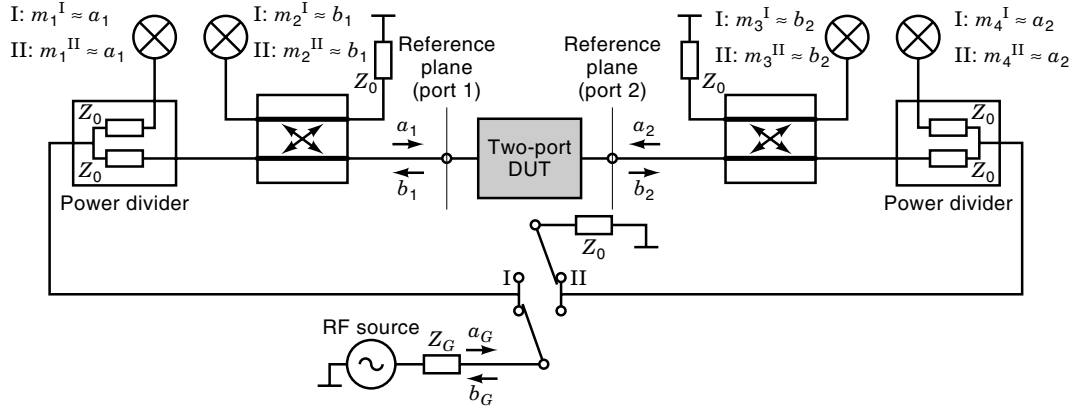


Figure 12. Bidirectional vector network analyzer with four receivers.

with the proportionality constants resulting from normalization measurements. Using a “Short” at port 1 and port 2 and a “Through” connecting both ports yields

$$c_{11} = -1 \frac{m_1^I}{m_2^I}, \quad c_{21} = 1 \frac{m_1^I}{m_3^I}$$

$$c_{22} = -1 \frac{m_1^{II}}{m_3^{II}}, \quad \text{and} \quad c_{12} = 1 \frac{m_1^{II}}{m_2^{II}}$$

The switch should be well-matched to keep measurement errors low, must be reproducible, and must provide sufficient isolation in order not to limit the dynamic range of the instrument.

The Four-Receiver Vector Network Analyzer

There seems to be little gained in extending the above-described three-receiver NA with a fourth receiver, as outlined in Fig. 12, since the complete S -matrix of a two-port can already be measured with three receivers. For *vector* network analyzers (VNA), however, the additional receiver provides simultaneous measurement capability of all waves and makes the architecture completely symmetrical around the DUT, resulting in interesting calibration capabilities, detailed in the sections entitled “The Seven-Term Error Model for Vector Network Analyzers with Four Receivers,” “The Cross-Talk Correcting 15-Term Error Model for Four-Receiver VNAs,” and “Modern Self-Calibration Techniques for Four-Receiver VNAs.” The subtle change of moving the switch from *between* two receivers to *in front* of the receiver pairs radically alters the error model, even reducing the requirements of the switch (see section entitled “The Seven-Term Error Model for Vector Network Analyzers with Four Receivers”). For measurements *without* error correction, however, the switch must, again, be as ideal as possible with respect to match, repeatability, and isolation.

A four-receiver VNA measures the S -matrix of a two-port DUT as

$$S_{11} = c_{11} \frac{m_2^I}{m_1^I}, \quad S_{21} = c_{21} \frac{m_3^I}{m_1^I}$$

$$S_{22} = c_{22} \frac{m_3^{II}}{m_4^{II}}, \quad \text{and} \quad S_{12} = c_{12} \frac{m_2^{II}}{m_4^{II}}$$

with the proportionality factors determined through normalization measurements as described before.

ERROR MODELS, CALIBRATION, AND ERROR CORRECTION

For a long time, precision NAs required carefully designed and expensive hardware, ensuring a good port match for embedding the DUT between proper terminations and high coupler directivity for precise reflection measurement. Losses in the hardware had to be accounted for by reference measurements of devices with known characteristic. With the advent of microprocessors, however, the focus has shifted toward calibration techniques that allow the imperfections of the hardware to be taken into account and corrected mathematically. An error model is needed for this purpose, and its parameters are determined through measurement of well-known one- or two-ports (calibration). This mathematical description of the non-ideal hardware is then used to correct the measurement of the DUT.

The error models, covered in subsequent sections, are used to model and correct all the linear errors of the VNA: transmission losses between the generator, the receivers, and the DUT; port match; and coupler directivity. Theoretically, the hardware of a VNA that provides these error-correction capabilities need no longer be built to high standards. Yet, commercially available systems still strive for good hardware performance for two reasons: first, in an effort to assist the calibration through good hardware performance and achieve even better accuracy; and, second, to provide an instrument that does not depend upon time consuming calibration procedures, allowing uncorrected measurements for all those applications that can tolerate the reduced accuracy. However, there are commercially available instruments that use the calibration capabilities for a considerable extension of their usable bandwidth. When the directivity of the directional couplers decreases to unsatisfactory values below a certain frequency, a built-in error correction mathematically enhances the directivity to usable values, extending the frequency range to as low as 9 kHz.

Since all error models presented herein are linear and deterministic, there remains a stringent requirement that the VNA hardware be linear (errors due to compression of receivers cannot be corrected) and stable.

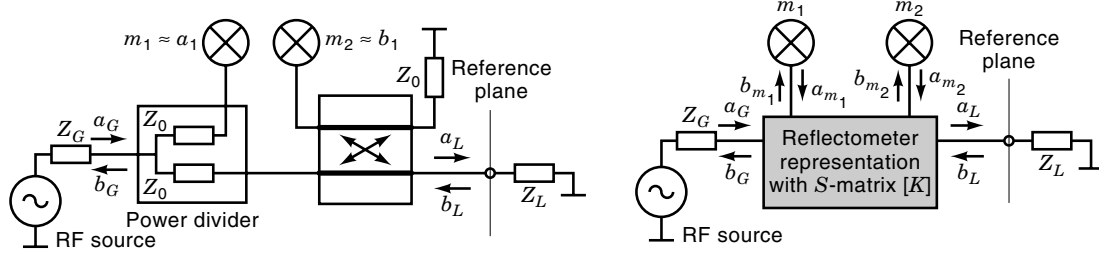


Figure 13. Four-port description of a reflectometer.

With modern equipment achieving a short-term stability in the millidecibel range, the most prominent source of error in many cases remains in the cabling between the VNA and the DUT and the associated interface contacts. These contacting errors are statistical in nature and must be minimized by proper cable and connector selection and maintenance. As the transmission phase of coaxial cables varies over temperature and may change if the cable is bent, VNAs are usually equipped with special test-port cables, optimized for phase-stability and mechanically protected against overbending.

The mathematical descriptions of the error models presented in the following sections make heavy use of matricial representations, leading to very compact expressions and similar structures for the different models. Alternatively, flow-graph techniques could be applied as in Refs. 17 and 18, leading to comparable results.

The Three-Term Error Model for Reflection Measurements

A reflectometer like the one of Fig. 13 can always be represented by a linear four-port, driven at the left port by the RF source, two ports connected to the receivers for measurement of the incident and reflected waves and the DUT placed at the port to the right. In this general form, the underlying hardware of the reflectometer becomes unimportant, much like the fact that internal structure of the six-port reflectometer was of no concern in the section entitled “The Six-Port Reflectometer.”

The waves incident to and reflected from this four-port relate to its S -matrix by

$$\begin{pmatrix} b_{m1} \\ b_{m2} \\ a_L \\ b_G \end{pmatrix} = [K] \begin{pmatrix} a_{m1} \\ a_{m2} \\ b_L \\ a_G \end{pmatrix} \quad (7)$$

The measured quantities m_1 and m_2 are proportional to the incident waves of the receivers ($m_1 = \eta_1 b_{m1}$, $m_2 = \eta_2 b_{m2}$), with the receiver constant η_i representing the entire transfer characteristics of the i th receiver, comprising mixer conversion, intermediate frequency (IF) amplification, and so on. Architectural details of the receiver are of no concern at this point. The only requirement is that η_i must be constant, implying stability of the receivers phase and amplitude response. Let r_i be the reflection coefficients of the receiver inputs such that the conditions

$$m_1 = \eta_1 b_{m1}, \quad m_2 = \eta_2 b_{m2}, \quad a_{m1} = r_1 b_{m1}, \quad a_{m2} = r_2 b_{m2} \quad (8)$$

result. Their use allows the elimination of the generator waves a_G and b_G in Eq. (7), according to the following scheme: The

third equation of Eq. (7) is solved for a_G and the result is inserted into the first two equations. In these, b_{m_i} is expressed through a_{m_i} and a_{m_i} in turn expressed by m_i , using Eq. (8). The resulting two equations relate the measured quantities m_1 and m_2 to the waves at the DUT and are arranged in matrix form as

$$\begin{pmatrix} b_L \\ a_L \end{pmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{pmatrix} m_2 \\ m_1 \end{pmatrix} = [G] \begin{pmatrix} m_2 \\ m_1 \end{pmatrix} \quad (9)$$

Every element of $[G]$ evolves as a function of K_i , r_i , and η_i , which need not be explicitly known. It suffices to determine $[G]$ through a process called calibration, in order to calculate the correct waves at the DUT from the raw measurements m_i :

$$\Gamma_L = \frac{b_L}{a_L} = \frac{G_{11}m_2 + G_{12}m_1}{G_{21}m_2 + G_{22}m_1} \quad (10)$$

Interestingly, this four-port/two-port reduction with its elimination of the generator waves has freed the error model [Eq. (9)] from any influence of the RF source. Power or impedance variations do not introduce measurement errors, provided that m_1 and m_2 are measured simultaneously.

Equation (10) can be further simplified by dividing numerator and denominator by m_1 and defining

$$m_{11} = \frac{m_2}{m_1} = \frac{\eta_2 b_{m2}}{\eta_1 b_{m1}} \quad (11)$$

as the uncorrected (raw) reflection measurement [compare with Eq. (6)] such that the error-corrected reflection coefficient becomes

$$\Gamma_L = \frac{b_L}{a_L} = \frac{G_{11}m_{11} + G_{12}}{G_{21}m_{11} + G_{22}} \quad (12)$$

The importance of Eq. (11) is twofold:

1. The ratio of two measurements enters the computation of the error-corrected result. Relative, rather than absolute, measurements are therefore still sufficient if an error model and error correction are used. And with the reflected wave measurement m_2 being referenced to the incident wave measurement m_1 , the theoretically derived insensitivity of the error model to power variations of the RF source becomes obvious.
2. The ratio of η_1/η_2 appearing in Eq. (11) indicates that amplitude and phase drifts of the receivers cancel if both receivers are built equal.

Both properties are essential for the high accuracy achievable with VNAs. The use of the latter property is especially important for accurate phase measurements at high frequencies.

Calibration of the Three-Term Error Model. Before error-corrected measurements using Eq. (12) can be carried out, the error matrix $[G]$ must be determined. For this calibration procedure, a number of one-ports (calibration standards) with known reflection coefficients are needed. Since numerator and denominator of Eq. (12) may be multiplied by any complex number without altering the result, one of the error terms may arbitrarily be set to 1, with the other three remaining to be computed. Without loss of generality, let $G_{11} = 1$ and Eq. (12) can be rearranged to yield a linear equation in the three remaining error terms:

$$-G_{12} + \Gamma^i m_{11}^i G_{21} + \Gamma^i G_{22} = m_{11}^i \quad (13)$$

The superscript i denotes the i th calibration measurement with known reflection coefficient Γ^i and the corresponding measured value m_{11}^i . Performing calibration measurements with three distinct and known reflection standards yields three linear inhomogeneous and independent forms of Eq. (13), sufficient to solve for the needed error terms.

In order to maximize the independence of the three equations, the three-term calibration is typically performed with those three impedances, which are furthest apart in the reflection plane:

$$\text{Open: } \Gamma^O = 1, \quad \text{Short: } \Gamma^S = -1, \quad \text{Match: } \Gamma^M = 0$$

Three-term reflectometer calibration is therefore also referred to as OSM calibration.

The Five-Term Error Model for Unidirectional Vector Network Analyzers

Extending a single reflectometer with a third receiver for transmission measurements yields the unidirectional NA of Fig. 10. Its error model comprises two parts: The reflectometer hardware to the left of the DUT which is again represented by an error two-port $[G]$, and the hardware to the right of the DUT which constitutes a physical two-port with a corresponding 2×2 S -matrix (Fig. 14).

As in the foregoing section, the reflection coefficient r_3 and the transfer characteristics η_3 of the receiver m_3 establish the conditions

$$a_{m_3} = r_3 b_{m_3} \quad \text{and} \quad m_3 = \eta_3 b_{m_3}$$

which are used to reduce $[T]$ to an error two-port that relates

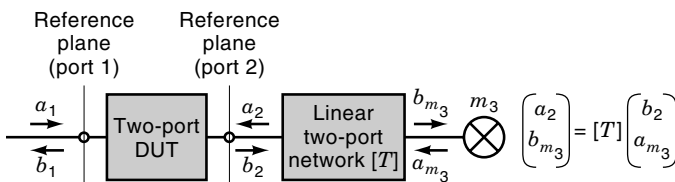


Figure 14. Hardware description of the transmission receiver m_3 .

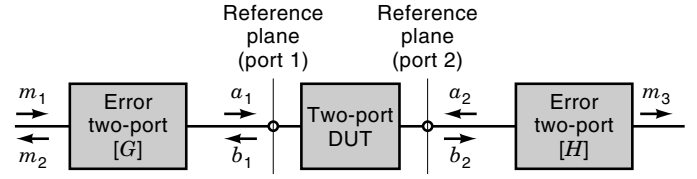


Figure 15. Error model of the unidirectional vector network analyzer.

the waves a_2 and b_2 of the DUT to the measured value m_3 :

$$\begin{pmatrix} b_2 \\ a_2 \end{pmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{pmatrix} m_3 \\ 0 \end{pmatrix} = [H] \begin{pmatrix} m_3 \\ 0 \end{pmatrix}$$

Together with the error two-port of the reflectometer

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{pmatrix} m_2 \\ m_1 \end{pmatrix} = [G] \begin{pmatrix} m_2 \\ m_1 \end{pmatrix}$$

and the definition of the DUT's S -matrix

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad (14)$$

the error model (Fig. 15) can be expressed as

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} G_{11}m_2 + G_{12}m_1 \\ H_{11}m_3 \end{pmatrix} = [S] \begin{pmatrix} G_{21}m_2 + G_{22}m_1 \\ H_{21}m_3 \end{pmatrix} = [S] \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad (15)$$

It is obvious that the two equations of Eq. (15) are not sufficient to solve for the entire S -matrix of the DUT. On the other hand, the "forward" S -parameters S_{11} and S_{21} cannot be error-corrected by themselves without the knowledge of S_{12} and S_{22} .

Error correction is therefore only possible if a second measurement with the DUT reversed is made. Denoting the measured values with superscripts f and r for forward and reverse measurement of the DUT, both vector equations of Eq. (15) can be combined into a matrix equation

$$\begin{bmatrix} G_{11}m_2^f + G_{12}m_1^f & H_{11}m_3^r \\ H_{11}m_3^f & G_{11}m_2^r + G_{12}m_1^r \end{bmatrix} = [S] \begin{bmatrix} G_{21}m_2^f + G_{22}m_1^f & H_{21}m_3^r \\ H_{21}m_3^f & G_{21}m_2^r + G_{22}m_1^r \end{bmatrix} \quad (16)$$

Again, Eq. (16) may be normalized to the incident waves m_1^f and m_1^r such that only the uncorrected measurement ratios

$$m_{11} = \frac{m_2^f}{m_1^f} = \frac{\eta_2 b_{m_2}^f}{\eta_1 b_{m_1}^f}, \quad m_{21} = \frac{m_3^f}{m_1^f} = \frac{\eta_3 b_{m_3}^f}{\eta_1 b_{m_1}^f}$$

$$m_{12} = \frac{m_3^r}{m_1^r} = \frac{\eta_3 b_{m_3}^r}{\eta_1 b_{m_1}^r}, \quad m_{22} = \frac{m_2^r}{m_1^r} = \frac{\eta_2 b_{m_2}^r}{\eta_1 b_{m_1}^r}$$

remain [compare with Eq. (6)]. Solving the normalized version

of Eq. (16) for $[S]$ yields the error-correction formula

$$[S] = \begin{bmatrix} G_{11}m_{11} + G_{12} & H_{11}m_{12} \\ H_{11}m_{21} & G_{11}m_{22} + G_{12} \end{bmatrix} \times \begin{bmatrix} G_{21}m_{11} + G_{22} & H_{21}m_{12} \\ H_{21}m_{21} & H_{21}m_{22} + G_{22} \end{bmatrix}^{-1} \quad (17)$$

Calibration of the Five-Term Error Model. Not all of the eight error terms contained in $[G]$ and $[H]$ are needed for error correction with Eq. (17), where H_{12} and H_{22} do not appear. Furthermore, the structure of Eq. (17) reveals its invariance to multiplication of *all* error terms with a constant. The remaining error terms may therefore be computed by rearranging Eq. (17) into a linear, homogeneous system of equations in the error terms,

$$\begin{bmatrix} -m_{11}^i & -1 & S_{11}^i m_{11}^i & S_{11}^i & 0 & S_{12}^i m_{21}^i \\ 0 & 0 & S_{21}^i m_{11}^i & S_{21}^i & -m_{21}^i & S_{22}^i m_{21}^i \\ 0 & 0 & S_{12}^i m_{22}^i & S_{12}^i & -m_{12}^i & S_{11}^i m_{12}^i \\ -m_{22}^i & -1 & S_{22}^i m_{22}^i & S_{22}^i & 0 & S_{21}^i m_{12}^i \end{bmatrix} \begin{pmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \\ H_{11} \\ H_{21} \end{pmatrix} = 0 \quad (18)$$

setting one error term to unity and solving the resulting inhomogeneous system of equations. $[S^i]$ denotes the S -matrix of the i th calibration standard. Five independent equations are needed to solve for the remaining five independent error terms, hence the name five-term error model.

The problem of *how many* two-port standards are needed and how they must be built to ensure five independent equations can be tackled in the following way: A suitable combination of standards must always comprise at least one standard with transmission ($S_{21} \neq 0$ and $S_{12} \neq 0$); otherwise, H_{11} and H_{21} cannot be determined [$S_{21} = S_{12} = 0$ implies $m_{21} = m_{12} = 0$, such that the fifth and sixth column of Eq. (18) vanish]. Numerical investigations reveal that a two-port standard exhibiting transmission contributes at least two equations and may contribute four if it is not symmetric (i.e., $S_{11} \neq S_{22}$).

Because the use of reflection standards ($S_{21} = S_{12} = 0$) implies $m_{21} = m_{12} = 0$, the second and third equation of Eq. (18) degenerate, leaving only the first and fourth equation for determination of $[G]$; and with $S_{11} = \Gamma_S$ the first equation corresponds directly to Eq. (13) (three-term calibration) with $G_{11} = 1$. Transmissionless standards therefore contribute one equation per reflection coefficient connected to port 1.

Even though the general nature of Eq. (18) allows for calibration with any set of standards that yields five independent equations, traditionally the five-term model is calibrated using the three reflection standards Open, Short, and Match (contributing three equations), together with a direct connection of both reference planes, the Through standard (furnishing the remaining two equations). Figure 16 depicts this commonly used set of standards. Another name for this commonly used procedure is SOLT, which stands for Short, Open, Load, and Through.

Of course, care must be taken as to build the calibration standards such that their S -matrix or reflection coefficients match the postulated values used in the calibration process. All deviations of the standards' *real* values from the postulated *ideal* of Fig. 16 lead to inconsistencies in the equations of Eq. (18). Keeping these inconsistency errors small requires a smart choice and precisely built or modeled standards.

Because the Through standard requires a direct connection of both measurement ports, it is typically a very easy standard to realize in coaxial media. The same holds for the Short, which is easily manufactured to high precision ($<1^\circ$ deviation at 20 GHz) in coaxial media. However, care must be taken as to ensure that the Short is connected directly to the reference plane, which is a potential problem if sexed connectors (N, SMA, K) are used. Bridging same-sexed connectors with adapters introduces a phase shift $2\beta l$ into the reflection coefficient, which must be known in order to substitute the shorts reflection coefficient by $-1e^{-2\beta l}$. The same holds if offset shorts with electrical length $2\beta l$ are used. For a frequency band where $2\beta l$ is between 40° and 320° , the latter can replace the open standard, which typically suffers from fringing capacities and radiation losses. Therefore, the open standards contained in commercially available calibration kits come with detailed information on those imperfections, modeling the frequency dependence of the open's capacitance by a polynomial to enhance accuracy. For the absolute impedance reference, the Match standard, precision crafting is needed for precision measurements.

The 10-Term Error Model for Vector Network Analyzers with Three Receivers

For all practical purposes, the applicability of the five-term error model is somewhat impaired by the need for manually reversing the DUT in order to obtain error-corrected results. If a switch is incorporated into the VNA, as depicted in Fig. 11, this manual interaction can be automated, easing error-corrected measurements considerably.

In essence, the error model of the resulting bidirectional VNA with three receivers (see section entitled "The Three-Receiver Network Analyzer Architecture for Full Two-Port Measurements") consists of *two* five-term error models (Fig. 17). Because of the switch being located *between* m_1 and m_2 and m_1 and m_3 , respectively, different error two-ports are needed for either position of the switch.

With the switch in position I, the reflectometer comprising m_1 and m_2 can be represented by an error two-port $[G^I]$ and the reflectometer terminated in Z_0 and comprising m_3 by an error two-port $[H^I]$. In the second state of the switch, the four-port comprising m_1 and m_3 can be reduced to the error two-port $[G^{II}]$ and the reflectometer terminated in Z_0 and comprising m_2 by an error two-port $[H^{II}]$ with

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = [G^I] \begin{pmatrix} m_2^I \\ m_1^I \end{pmatrix}, \quad \begin{pmatrix} b_2 \\ a_2 \end{pmatrix} = [H^I] \begin{pmatrix} m_3^I \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = [H^{II}] \begin{pmatrix} m_2^{II} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} b_2 \\ a_2 \end{pmatrix} = [G^{II}] \begin{pmatrix} m_3^{II} \\ m_1^{II} \end{pmatrix}$$

$$\text{Through: } [S^T] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{Open: } S_{11}^O = 1, \quad \text{Short: } S_{11}^S = -1, \quad \text{Match: } S_{11}^M = 0$$

Figure 16. Calibration standards for the TMSO five-term procedure.

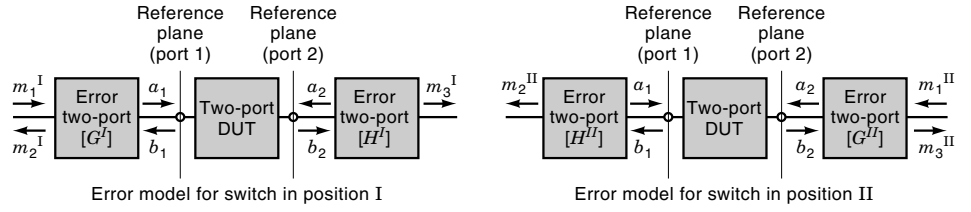


Figure 17. Error model of the bidirectional vector network analyzer with three receivers.

Using these relations together with Eq. (14) yields a vector equation for either state of the switch which can be combined into a matrix equation, very similar to Eq. (16):

$$\begin{bmatrix} G_{11}^I m_2^I + G_{12}^I m_1^I & H_{11}^{II} m_2^{II} \\ H_{11}^I m_3^I & G_{11}^{II} m_3^{II} + G_{12}^{II} m_1^{II} \end{bmatrix} = [S] \begin{bmatrix} G_{21}^I m_2^I + G_{22}^I m_1^I & H_{21}^{II} m_2^{II} \\ H_{21}^I m_3^I & G_{21}^{II} m_3^{II} + G_{22}^{II} m_1^{II} \end{bmatrix}$$

Normalization to the incident waves m_1^I and m_1^{II} now yields

$$\begin{bmatrix} G_{11}^I m_{11} + G_{12}^I & H_{11}^{II} m_{12} \\ H_{11}^I m_{21} & G_{11}^{II} m_{22} + G_{12}^{II} \end{bmatrix} = [S] \begin{bmatrix} G_{21}^I m_{11} + G_{22}^I & H_{21}^{II} m_{12} \\ H_{21}^I m_{21} & G_{21}^{II} m_{22} + G_{22}^{II} \end{bmatrix} \quad (19)$$

with

$$m_{11} = \frac{m_2^I}{m_1^I} = \frac{\eta_2 b_{m_2}}{\eta_1 b_{m_1}}, \quad m_{21} = \frac{m_3^I}{m_1^I} = \frac{\eta_3 b_{m_3}}{\eta_1 b_{m_1}}$$

$$m_{12} = \frac{m_2^{II}}{m_1^{II}} = \frac{\eta_2 b_{m_2}}{\eta_1 b_{m_1}}, \quad m_{22} = \frac{m_3^{II}}{m_1^{II}} = \frac{\eta_3 b_{m_3}}{\eta_1 b_{m_1}}$$

where comparison with the five-term error model shows that m_{12}^{II} and m_{13}^{II} have switched their position as have the error terms for the second state of the switch, a consequence of the switch now reversing the DUT and the error two-ports.

With error correction according to the 10-term error model applied, the requirements of the switch as detailed in the section entitled “The Three-Receiver Network Analyzer Architecture for Full Two-Port Measurements” become somewhat relaxed, because its mismatch errors and transmission losses are now accounted for. However, repeatability and leakage errors (finite isolation) are not contained in the error model.

Calibration of the 10-Term Error Model. As was the case with the five-term model, not all 16 error terms appear in Eq. (19). H_{12}^I , H_{22}^I , H_{12}^{II} , and H_{22}^{II} are not needed for error correction and cannot be determined by the calibration procedure.

Rearranging Eq. (19) as a linear homogeneous system of equations in the remaining error terms, as demonstrated in the section entitled “Calibration of the Five-Term Error Model,” finds the four equations to be decoupled: Two equations depend solely upon error terms of $[G^I]$ and $[H^I]$, and the other two

equations depend only upon $[G^{II}]$ and $[H^{II}]$:

$$\begin{bmatrix} -m_{11}^I & -1 & S_{11} m_{11}^I & S_{11} & 0 & S_{12} m_{21}^I \\ 0 & 0 & S_{21} m_{11}^I & S_{21} & -m_{21}^I & S_{22} m_{21}^I \end{bmatrix} \mathbf{e}^I = 0,$$

$$\mathbf{e}^I = (G_{11}^I, G_{12}^I, G_{21}^I, G_{22}^I, H_{11}^I, H_{21}^I)^T$$

$$\begin{bmatrix} 0 & 0 & S_{12} m_{22}^I & S_{12} & -m_{12}^I & S_{11} m_{12}^I \\ -m_{22}^I & -1 & S_{22} m_{22}^I & S_{22} & 0 & S_{21} m_{12}^I \end{bmatrix} \mathbf{e}^{II} = 0,$$

$$\mathbf{e}^{II} = (G_{11}^{II}, G_{12}^{II}, G_{21}^{II}, G_{22}^{II}, H_{11}^{II}, H_{21}^{II})^T$$

With one of the error terms of both $\mathbf{e}^I$ and $\mathbf{e}^{II}$ set to 1 (e.g., $G_{11}^I = G_{11}^{II} = 1$), two sets of five independent equations must be supplied by the calibration measurements in order to solve the resulting inhomogeneous system for the remaining 10 error terms.

A similar reasoning (compared with the section entitled “Calibration of the Five-Term Error Model”) as to how many and what kind of standards are needed is applicable here. At least one standard with transmission ($S_{21} \neq 0$ and $S_{12} \neq 0$) is needed in order to determine $[H^I]$ and $[H^{II}]$. Because of the different error terms, such a standard will supply a total of four equations for both positions of the switch. Reflection two-ports ($S_{21} = S_{12} = 0$), on the other hand, still contribute one equation for each of their two reflection coefficients S_{11} and S_{22} .

The same standards as with the five term error model (referred to as TMSO, Through-Match-Short-Open, or SOLT, Short-Open-Load-Through) are commonly used to calibrate the 10-term model, with the reflection standards now being reflection two-ports.

Because of the decoupled nature of Eq. (19), the reflection two-ports may be realized as one-ports by first connecting the calibration impedance to port 1 and noting the measurement values with the switch in position I. Then the appropriate calibration impedance is connected to port 2 and the measurements are taken with the switch in the second position [since these standards do not exhibit transmission, the corresponding transmission receiver (m_{13}^I , m_{12}^{II}) need not be considered]. With respect to the commonly used TMSO standards, this sequential procedure eliminates the need for two identical reflection standards but allows usage of one physical match—short and open—as reflection one-ports.

The Cross-Talk Including 12-Term Error Model for Three-Receiver VNAs

A number of commercially available VNAs provide a 12-term calibration capability (17), expanding the above-described 10-term error model by two error terms that characterize forward and reverse isolation. Their inclusion aims at extending the dynamic range for measurements where finite isolation

$$\begin{aligned} \text{Through: } [S^T] &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & \text{Open: } [S_O^I] &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \text{Short: } [S^S] &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, & \text{Match: } [S^M] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Figure 18. Calibration standards for the TMSO (SOLT) 10-term procedure.

of the switch or coupling over the DUT adversely affects measurement accuracy at high insertion loss levels of the DUT.

For that purpose an additional isolation standard (no transmission) with, for now, arbitrary reflection coefficients r_1 and r_2

$$[S^X] = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \rightarrow [M^X] = \begin{bmatrix} m_{11}^X & m_{12}^X \\ m_{21}^X & m_{22}^X \end{bmatrix}$$

must be measured (alternatively the transmission measurement of one of the reflection standards may be exploited). The transmission terms m_{12}^X and m_{21}^X of its measurement matrix are a measure of the cross-talk level and can be used for correcting the measured transmission values of the DUT:

$$[S] = \begin{bmatrix} G_{11}^I m_{11} + G_{12}^I & H_{11}^I (m_{12} - m_{12}^X) \\ H_{11}^I (m_{21} - m_{21}^X) & G_{11}^I m_{22} + G_{12}^I \end{bmatrix} \times \begin{bmatrix} G_{21}^I m_{11} + G_{22}^I & H_{21}^I (m_{12} - m_{12}^X) \\ H_{21}^I (m_{21} - m_{21}^X) & G_{21}^I m_{22} + G_{22}^I \end{bmatrix}^{-1}$$

However, this simple correction scheme can only improve results if the reflection coefficients of the isolation standard match the reflection coefficients of the DUT. For well-matched DUTs, the isolation terms can be taken from the measurement of the Match standard $[S^M]$. High transmission dynamic DUTs (e.g., filters), however, typically exhibit strongly varying reflection coefficients over frequency. If two identical DUTs are available, an isolation standard that satisfies the above condition can be generated by connecting the first DUT with port 1 to port 1 of the VNA and the second DUT with port 2 to the VNAs second port. The unused ports of the DUTs are terminated in Z_0 . Consequently the isolation measurement must be repeated with a suitable isolation standard whenever DUTs with differing reflection coefficients are to be measured.

Considering the above limitations, the cross-talk reduction with the 12-term model should be regarded as a cross-talk *normalization* rather than a *calibration*.

In the section entitled “The Cross-Talk Correcting 15-Term Error Model for Four-Receiver VNAs,” it will be shown that for an error model to correctly include cross-talk, eight error terms are needed for its characterization.

The Seven-Term Error Model for Vector Network Analyzers with Four Receivers

The addition of a fourth receiver to independently measure all four waves at the DUT, as depicted in Fig. 12, leads to very interesting instrument properties and calibration possibilities.

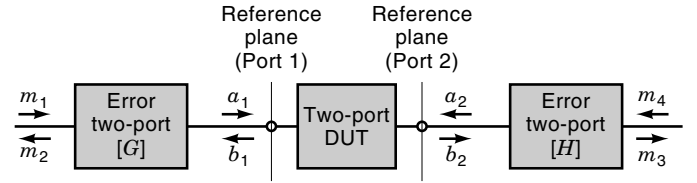


Figure 19. Seven-term error model of the bidirectional vector network analyzer with four receivers.

The DUT is now embedded in a symmetrical test set, consisting of two reflectometers, each of which can be represented by an error two-port (see section entitled “The Three-Term Error Model for Reflection Measurements,” which discusses four-port/two-port reduction).

The subtle but far-reaching advantages of this architecture and the accompanying error-model are:

1. The error model (Fig. 19) is valid for *either* position of the switch, reducing the number of error terms to eight and the number of two-port calibration standards required to three.
2. The four-receiver architecture provides redundant information in the calibration measurements that can be exploited to allow calibration with *partially unknown* standards. The section entitled “Modern Self-Calibration Techniques for Four-Receiver VNAs” details these important and powerful techniques.
3. The switch is no longer part of the error two-ports. Instead it has been moved into the generator port of the reflectometers, whose waves were eliminated during the four-port/two-port reduction. Therefore, the imperfections of the switch no longer influence the quality of an error-corrected measurement. The switch may be mismatched, may exhibit transmission losses and finite isolation, and does not even need to be repeatable.

The mathematical derivation of the seven-term error model unfolds in the usual way. With

$$\begin{aligned} \begin{pmatrix} b_1 \\ a_1 \end{pmatrix} &= [G] \begin{pmatrix} m_2 \\ m_1 \end{pmatrix}, & \begin{pmatrix} b_2 \\ a_2 \end{pmatrix} &= [H] \begin{pmatrix} m_3 \\ m_4 \end{pmatrix}, & \text{and} \\ \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &= [S] \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \end{aligned} \quad (20)$$

vector equations for both positions of the switch (I and II) are obtained and combined into a matrix equation:

$$\begin{bmatrix} G_{11}m_2^I + G_{12}m_1^I & G_{11}m_2^{II} + G_{12}m_1^{II} \\ H_{11}m_3^I + H_{12}m_4^I & H_{11}m_3^{II} + H_{12}m_4^{II} \end{bmatrix} = [S] \begin{bmatrix} G_{21}m_2^I + G_{22}m_1^I & G_{21}m_2^{II} + G_{22}m_1^{II} \\ H_{21}m_3^I + H_{22}m_4^I & H_{21}m_3^{II} + H_{22}m_4^{II} \end{bmatrix} \quad (21)$$

The eight measurement values contained in Eq. (21) can be

condensed to four by expanding Eq. (21) as

$$\begin{bmatrix} G_{12} & 0 \\ 0 & H_{12} \end{bmatrix} \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix} + \begin{bmatrix} G_{11} & 0 \\ 0 & H_{11} \end{bmatrix} \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} \\ = [S] \left\{ \begin{bmatrix} G_{22} & 0 \\ 0 & H_{22} \end{bmatrix} \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix} + \begin{bmatrix} G_{21} & 0 \\ 0 & H_{21} \end{bmatrix} \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} \right\}$$

and right-multiplying both sides with the inverse measurement matrix of the first and fourth receiver. The result is much simplified error-model representation

$$\begin{bmatrix} G_{12} & 0 \\ 0 & H_{12} \end{bmatrix} + \begin{bmatrix} G_{11} & 0 \\ 0 & H_{11} \end{bmatrix} [M] \\ = [S] \left\{ \begin{bmatrix} G_{22} & 0 \\ 0 & H_{22} \end{bmatrix} + \begin{bmatrix} G_{21} & 0 \\ 0 & H_{21} \end{bmatrix} [M] \right\} \quad (22)$$

with a 2×2 measurement matrix

$$[M] = \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix}^{-1} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

With the measurement values pertaining to the incident waves m_1^I and m_1^{II} residing on the main diagonal of the right submatrix and the off-diagonal elements being small, the matrix is guaranteed to be invertible. For $m_4^I = m_1^I = 0$ (three receiver analogy) the entries of $[M]$ are normalized with respect to the incident waves m_1^I and m_1^{II} . The measurement matrix $[M]$ is therefore attributed S -parameter character. Furthermore, the computation of $[M]$ provides a first level of error correction, because finite isolation of the switch is eliminated. This is achieved by the fourth receiver, which precisely measures the level of the switch's leakage signals through m_1^I and m_1^{II} .

Equation (22) is easily returned to the structure of Eq. (21), with the four measurement values now being direct counterparts of the corresponding error-corrected S -parameters of the DUT:

$$\begin{bmatrix} G_{11}m_{11} + G_{12} & G_{11}m_{12} \\ H_{11}m_{21} & H_{11}m_{22} + H_{12} \end{bmatrix} \\ = [S] \begin{bmatrix} G_{21}m_{11} + G_{22} & G_{21}m_{12} \\ H_{21}m_{21} & H_{21}m_{22} + H_{22} \end{bmatrix} \quad (23)$$

It may be interesting to note that $[G]$ and $[H]$ are of the cascading *transfer matrix* type. This type of matrix also appears in all the other error models introduced in this section. The fact that transfer matrices exhibit a singularity if the underlying network has zero transmission can safely be ignored as far as error two-ports of practical reflectometer or network analyzer realizations are concerned. Those networks *must* exhibit transmission in order to perform the desired function. The device under test, however, may be perfectly isolating and should therefore always be represented by its S -matrix (the section entitled "The General TAN Procedure" introduces *pseudo-transfer matrices* to work around that singularity).

Calibration of the Seven-Term Error Model. With Eq. (23) being of the same structure as the five-term model [Eq. (16)] and the 10-term model [Eq. (19)], the determination of the error terms proceeds in the same way. Equation (23) is rearranged into a linear homogeneous system of equations

$$\begin{bmatrix} -m_{11}^i & -1 & S_{11}^i m_{11}^i & S_{11}^i & 0 & 0 & S_{12}^i m_{21}^i & 0 \\ 0 & 0 & S_{21}^i m_{11}^i & S_{21}^i & -m_{21}^i & 0 & S_{22}^i m_{21}^i & 0 \\ -m_{12}^i & 0 & S_{11}^i m_{12}^i & 0 & 0 & 0 & S_{12}^i m_{22}^i & S_{12}^i \\ 0 & 0 & S_{21}^i m_{12}^i & 0 & -m_{22}^i & -1 & S_{22}^i m_{22}^i & S_{22}^i \end{bmatrix} \mathbf{e} = 0, \\ \mathbf{e} = (G_{11}, \dots, G_{22}, H_{11}, \dots, H_{22})^T$$

with every calibration measurement i contributing four equations. One error term is set to 1, and seven independent equations are required to solve for the remaining seven error terms. It turns out that the eight equations stemming from the measurement of two calibration two-ports are not sufficient to provide the necessary rank of 7. Measurement of a third calibration two-port is needed such that a total of 12 equations exist.

The resulting surplus in information can be used in one of the following ways:

1. The linear system with all 12 equations is solved in the least-squares sense (e.g., see Ref. 19), using the extra equations to minimize residual errors (contacting or repeatability errors).
2. Even though contacting errors cannot be accounted for due to their statistical nature, in Ref. 20 the extra information is used to assess these errors by deriving error bounds. As these techniques are not yet widely used, they will not be further treated here.
3. The calibration standards can be permitted to have unknown parameters (13,21), which are determined using the extra information contained in the additional equations. Because the construction of standards, which do not have to be fully known, can result in considerable savings, this "self-calibration" technique is of much practical importance. The section entitled "Modern Self-Calibration Techniques for Four-Receiver VNAs" has the details.

Performing a seven-term calibration with completely known standards is most easily done using the TMS procedure, where the three calibration two-ports

$$\text{Through: } [S^T] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{Match: } [S^M] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \text{Short: } [S^S] = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

are used. Comparison with the 10-term and the 5-term model shows that the third reflection standard is no longer needed. Using a short, instead of an open, as the third standard is in many cases the more convenient choice, because shorts can be manufactured to higher precision in coaxial media or waveguides (see section entitled "Calibration of the Five-Term Error Model"). At low frequencies, however, using an open may be the more advantageous choice, as the test port itself serves as a comparatively good open up to about 100 MHz.

The computed error terms are used for error correction by simply solving Eq. (23) for $[S]$:

$$[S] = \begin{bmatrix} G_{11}m_{11} + G_{12} & G_{11}m_{12} \\ H_{11}m_{21} & H_{11}m_{22} + H_{12} \end{bmatrix} \times \begin{bmatrix} G_{21}m_{11} + G_{22} & G_{21}m_{12} \\ H_{21}m_{21} & H_{21}m_{22} + H_{22} \end{bmatrix}^{-1}$$

Calibration of the Seven-Term Error Model using an Unknown Through. For applications where a Through standard of known properties is not available, the seven-term error model may be calibrated by performing simple reflectometer calibrations for the error two-ports $[G]$ and $[H]$, according to the section entitled “Calibration of the Three-Term Error Model,” and using a transmission standard whose only known property is its reciprocity (e.g., $S_{21} = S_{12}$).

Such a calibration method may, for instance, be desirable for on-wafer measurements where nonaligned measurement ports render a known Through difficult to build (22). Calibration without an unknown Through was first introduced in Ref. 23 and is very simple to derive if the cascade structure of the error model (Fig. 19) is written using transfer matrices. With the error two-ports $[G]$ and $[H]$ already in that notation [Eq. (20)], the difference to Eq. (22) is in the description of the DUT

$$\begin{pmatrix} b_1 \\ a_1 \end{pmatrix} = [\Sigma] \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \frac{1}{S_{21}} \begin{bmatrix} -\Delta S & S_{11} \\ -S_{22} & 1 \end{bmatrix} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} \quad (24)$$

and in the evolving measurement matrix

$$\begin{bmatrix} m_1^I & m_1^{II} \\ m_2^I & m_2^{II} \end{bmatrix} \begin{bmatrix} m_3^I & m_3^{II} \\ m_4^I & m_4^{II} \end{bmatrix}^{-1} = [P] = [G]^{-1}[\Sigma][H]$$

with the latter now also being a transfer matrix.

Performing reflectometer calibrations, as described in the section entitled “Calibration of the Three-Term Error Model,” for both error two-ports establishes three of the four error terms in $[G]$ and $[H]$, leading to the preliminary error matrices $[\tilde{G}]$

and $[\tilde{H}]$, with one error term normalized to 1. Without loss of generality, let us assume $\tilde{G}_{11} = \tilde{H}_{11} = 1$. This knowledge of $[G]$ and $[H]$, except for a common factor, can be expressed as

$$[P] = [G]^{-1}[\Sigma][H] = \frac{1}{K_G} [\tilde{G}]^{-1}[\Sigma]K_H[\tilde{H}] = \alpha [\tilde{G}]^{-1}[\Sigma][\tilde{H}]$$

$$\alpha = \frac{K_H}{K_G}$$

The still unknown term α can be regarded as the seventh error term, still needed for a full two-port calibration. As the determinant of the transfer matrix of a reciprocal two-port is unity, using such a device for calibration produces the equation

$$\det[P^{\text{reci}}] = \alpha^2 \det[\tilde{G}]^{-1} \det[\Sigma^{\text{reci}}] \det[\tilde{H}] = \alpha^2 \frac{\det[\tilde{H}]}{\det[\tilde{G}]}$$

which is solved for α :

$$\alpha = \pm \sqrt{\frac{\det[\tilde{G}]}{\det[\tilde{H}]} \det[P^{\text{reci}}]} \quad (25)$$

According to Ref. 23, the correct root in Eq. (25) is most easily found using

$$S_{21}^{\text{reci}} = \frac{\alpha}{P_{22}^{\text{reci}}}$$

if the phase of the reciprocal transmission-factor is known to $\pm 90^\circ$.

The Cross-Talk Correcting 15-Term Error Model for Four-Receiver VNAs

Whenever energy emanating from the measurement planes finds a path to leak around the DUT or a coupling path between the reflectometers of the VNA exists (Fig. 20), the error models of the sections entitled “The Five-Term Error Model for Unidirectional Vector Network Analyzers,” “The 10-Term Error Model for Vector Framework Analyzers with Three Receivers,” “The Cross-Talk Including 12-Term Error Model for

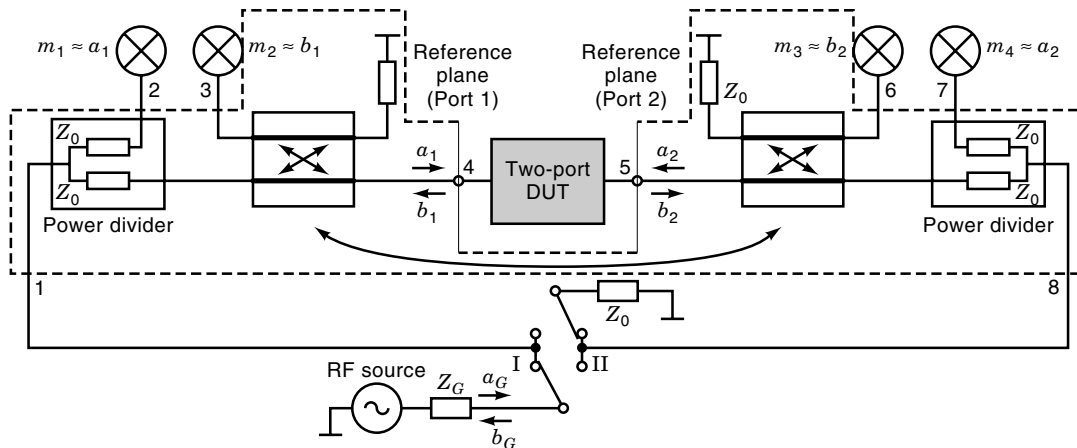


Figure 20. Eight-port hardware representation of the vector network analyzer, including coupling of the reflectometers.

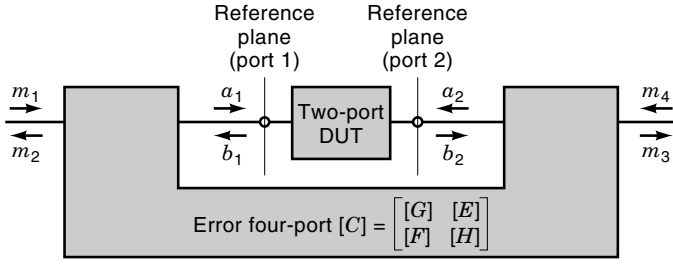


Figure 21. Fifteen-term error-model of the four-receiver vector network analyzer.

Three-Receiver VNAs,” and “The Seven-Term Error Model for Vector Network Analyzers with Four Receivers” cannot adequately handle the associated cross-talk error. On-wafer measurements typically suffer from this kind of leakage error, due to the small separation of the measurement planes and the radiation of the probe tips.

An error model, capable of accounting for those leakage errors was first detailed in Ref. 24 and is also termed *full-model* (14). It can be derived by describing the four-receiver VNA's hardware as a general eight-port. The coupling between the reflectometers may be arbitrary and nonreciprocal but must be linear and not affected by the parameters of the DUT. Comparable to the four-port/two-port reduction, introduced in the section entitled “Calibration of the Five-Term Error Model,” the reflection coefficients and conversion constants of the four receivers are used to eliminate the reflected waves at the receivers and the generator waves a_G and b_G .

The resulting error four-port $[C]$ (Fig. 21) relates the measured values m_i ($i = 1, 2, 3, 4$) to the waves a_1, a_2, b_1, b_2 at the DUT such that its S -matrix can be computed from the measurement data and $[C]$. Because the switch is not part of the error network and the waves of the RF source are eliminated in Eq. (26), the properties of the switch have no influence on the quality of error correction, and the same error network may be used for either state of the switch.

The mathematical formulation of the error model is derived by partitioning the 4×4 matrix $[C]$ into four 2×2 matrices. With

$$\begin{pmatrix} b_1 \\ b_2 \\ a_1 \\ a_2 \end{pmatrix} = [C] \begin{pmatrix} m_1 \\ m_4 \\ m_2 \\ m_3 \end{pmatrix} = \begin{bmatrix} [G] & [E] \\ [F] & [H] \end{bmatrix} \begin{pmatrix} m_1 \\ m_4 \\ m_2 \\ m_3 \end{pmatrix} \quad (26)$$

the two vector equations

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = [G] \begin{pmatrix} m_1 \\ m_4 \end{pmatrix} + [E] \begin{pmatrix} m_2 \\ m_3 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = [F] \begin{pmatrix} m_1 \\ m_4 \end{pmatrix} + [H] \begin{pmatrix} m_2 \\ m_3 \end{pmatrix}$$

can be derived to express the waves at the DUT as

$$[G] \begin{pmatrix} m_1 \\ m_4 \end{pmatrix} + [E] \begin{pmatrix} m_2 \\ m_3 \end{pmatrix} = [S] \left\{ [F] \begin{pmatrix} m_1 \\ m_4 \end{pmatrix} + [H] \begin{pmatrix} m_2 \\ m_3 \end{pmatrix} \right\}$$

An equation of this type is obtained for measurement in either position of the switch, and combining both vector equations into a matrix equation yields

$$[G] \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix} + [E] \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} = [S] \left\{ [F] \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix} + [H] \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} \right\}$$

which simplifies to

$$[G] + [E][M] = [S]\{[F] + [H][M]\} \quad (27)$$

by introduction of the measurement matrix

$$[M] = \begin{bmatrix} m_2^I & m_2^{II} \\ m_3^I & m_3^{II} \end{bmatrix} \begin{bmatrix} m_1^I & m_1^{II} \\ m_4^I & m_4^{II} \end{bmatrix}^{-1} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

Comparison with the formulation of the seven-term model [Eq. (22)] reveals the very close relationship, because their structure and even the measurement matrix are identical. In fact, the seven-term error model evolves as a special case of the more general 15-term model. Consequently, the eight off-diagonal error terms that default to zero in Eq. (21), where no leakage is assumed, are needed in the 15-term model to correctly describe the cross-talk.

Calibration of the 15-Term Error Model. Determination of the 16 error terms in the error quadrants $[G]$, $[E]$, $[F]$, and $[H]$ is achieved by expanding Eq. (27) into a system of four linear equations in the error terms per calibration measurement (25). One error term may be arbitrarily set to one, and the resulting inhomogeneous system may be solved for the remaining 15 error terms. Care must be taken to choose one of the error terms *on* the diagonal of the error-quadrants for this normalization, because the *off*-diagonal elements vanish in the case of a leakage-free error four-port.

It turns out that at least five completely known and distinct calibration two-ports must be measured to determine the 15 error terms (25,26). With five calibration measurements, 20 equations in the 15 unknowns exist, but only 15 are needed. Analogous to the section entitled “The Seven-Term Model for Vector Network Analyzers with Four Receivers” either five suitable equations may be neglected or, alternatively, the system of 20 equations in 15 unknowns is solved in the least-squares sense. Use of partially unknown standards and their determination through self-calibration is the third option and is outlined in the section entitled “Self-Calibration for the 15-Term Error Model.”

An especially well-suited combination of standards for the 15-term model, first mentioned in Ref. 25, consists of a Through two-port and the three reflection one-ports *match*, *short*, and *open*, combined to yield four reflection two-ports such that calibration is performed with the five two-ports of Fig. 22.

Contrary to the 7-term, 5-term, or 10-term error-models, the 15-term model does not allow reflection two-ports to be measured sequentially, by separating the reflection measurements at both ports. Due to the presence of cross-talk that must be precisely measured, both ports need to be terminated

$$S^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad S^{MS} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \quad S^{OM} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad S^{SO} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad S^{OS} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Figure 22. Calibration standards for the 15-term Tms0-procedure.

in the prescribed impedance while performing the calibration measurement. Cyclically interchanging the three reflection one-ports to create the calibration standard lets this Tms0 procedure get by with one match-, short-, and open standard. Lowercase letters are used to indicate that match, short, and open are one-ports and need to be combined to form the necessary calibration two-ports.

Once all error terms are known, the measured values of an arbitrary DUT are to be error-corrected. Contrary to the calibration process, error correction is performed with every measurement. Numerical efficiency is therefore in demand.

Equation (27) is easily solved for the S -matrix $[S]$ of the DUT

$$[S] = \{[G][M] + [E]\}\{[H][M] + [F]\}^{-1}$$

and devises a very simple algorithm for the error-correction process. Only 2×2 matrices need to be manipulated.

Figure 23 (taken from Ref. 27) demonstrates the capabilities of the 15-term procedure in the presence of an artificial leakage error between the measurement planes. Figure 24 depicts the setup. The DUT is a 20 dB attenuator, and the cross-talk is modeled by a second 20 dB attenuator in parallel to the DUT.

While the presence of the leakage error corrupts the result in a way that the uncorrected data bears no resemblance with the transfer characteristics of the DUT, correcting the data according to the 15-term procedure recovers the attenuation value with good precision, even though the amount of cross-talk is of the same magnitude as the DUT's transmission.

MODERN SELF-CALIBRATION TECHNIQUES FOR FOUR-RECEIVER VNAs

Calibration of four-receiver VNAs according to the 7-term or the 15-term error model yields five redundant equations that are not needed for the computation of the error terms. Using this extra information for the determination of unknown parameters of the calibration standards is a very powerful technique that allows the calibration to be performed with *partially known* standards. Besides reducing the cost of standards, self-calibration techniques also enhance the accuracy of the calibration as inconsistency errors are reduced.

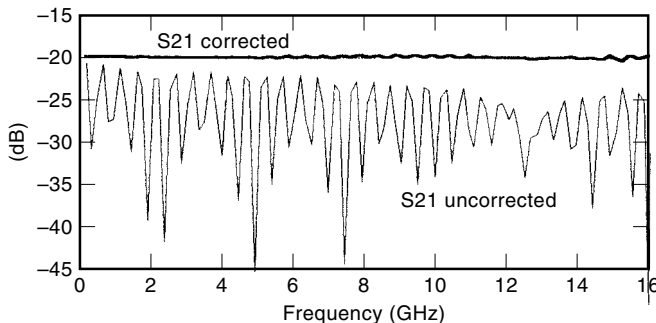


Figure 23. Transmission measurement of a 20 dB attenuator (corrected/uncorrected). (From Ref. 27, with permission.)

One of the most popular self-calibration representative, introduced in 1979 by Engen and Hoer (13) for the double six-port reflectometer (seven-term error model), is the TLR procedure. It allows calibration with the following standards:

- (T)hrough (direct connection of both measurement planes)
- (L)ine (a line of unknown electrical length, between 20° and 160° , perfectly matched)
- (R)eflect (unknown, but symmetrical reflection standard)

The L standard and the R standard are only partially known, with no requirements imposed upon the reflect, except that the same reflect has to be used on both measurement planes. Its value and the exact electrical length of the line are determined through the self-calibration process. Different methods can be employed for the determination of the unknown parameters of the standards, three of which are as follows:

1. The *linear* system of equations in the error terms [Eqs. (23) and (27)] is treated as a *nonlinear* system in the error terms *and* the unknowns of the standards. This nonlinear system is solved iteratively (28) or, where possible, analytically, providing a simultaneous solution of error terms and unknowns of the standards (17,18,29).
2. For a unique solution of the error terms to exist, five of the 12 (respectively 20) equations must be linearly dependent. If the 12×8 (20×16) coefficient matrix of the homogeneous linear system of equations [Eqs. (23) and (27)] can be *analytically* brought to tridiagonal form, the last five lines are required to vanish, resulting in five equations in which only measurement values and parameters of the standards appear. It may be argued that the analytical reduction of a 20×16 or even a 12×8 matrix to tridiagonal form is impracticable. However, Eqs. (23) and (27) are sparse enough to render the derivation of self-calibration equations according to this scheme feasible, even for the 15-term error model (30). Once the unknown parameters of the standards are determined, the standards are completely known and so are all nonzero entries of the tridiagonal matrix, which is then used to compute the error terms by a backsubstitution process (19, p. 37).

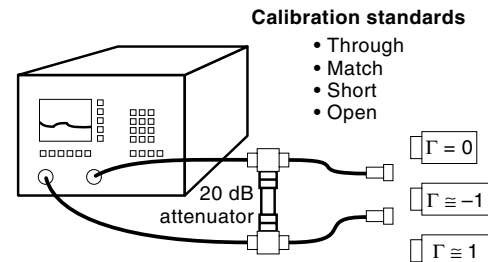


Figure 24. Measurement setup of a network analyzer with artificial cross-talk.

3. The formulation of the error model is rearranged to yield relations between the standards' S -parameters and the measured values, eliminating the error terms. These relations are exploited to solve for the unknown parameters of the calibration standards which are then completely known and can be used to solve the linear system of equations [Eqs. (23) and (27)] for the error terms. In the following sections, this method, which is based on similarity transformations (21,27), will be presented for the 7-term and the 15-term error model.

Self-Calibration for the Seven-Term Error Model

Elimination of the error parameters to arrive at a relation between the standards' parameters and the measurement values was first described by Eul and Schiek (21) and is most easily achieved by using transfer parameters to describe the cascade structure of the error model (Fig. 19). See Eq. (24) in the section entitled "Calibration of the Seven-Term Error Model Using an Unknown Through."

As transfer parameters exhibit a singularity for transmissionless two-ports ($S_{21} = 0$), this representation of the error model is less suited for the computation of the error terms, which is why the S -parameter representation [Eq. (22)] was chosen in the section entitled "The Seven-Term Error Model for Vector Network Analyzers with Four Receivers." The seven-term error model, however, may as well be written using Eq. (24) with the three calibration measurements yielding three such equations.

Solving the i th calibration measurement for $[H]$.

$$[H] = [\Sigma_i]^{-1}[G][P_i]^{-1} \quad (28)$$

and substituting Eq. (28) into the remaining calibration measurements j and k yields

$$\begin{aligned} [P_j][P_i]^{-1} &= [G]^{-1}[\Sigma_j][\Sigma_i]^{-1}[G] \quad \text{and} \\ [P_k][P_i]^{-1} &= [G]^{-1}[\Sigma_k][\Sigma_i]^{-1}[G] \end{aligned} \quad (29)$$

Recognizing Eq. (29) as similarity transformations, the equality of the eigenvalues of the similar matrices is equivalent to the equality of their determinants and traces, hence

$$\begin{aligned} [A] &= [X][B][X]^{-1} \Rightarrow \det[A] = \det[B] \\ \text{trace}[A] &= a_{11} + a_{22} = \text{trace}[B] = b_{11} + b_{22} \end{aligned}$$

which yields the four equations

$$\begin{aligned} \det\{[P_j][P_i]^{-1}\} &= \beta_1 = \det\{[\Sigma_j][\Sigma_i]^{-1}\}, \\ \text{trace}\{[P_j][P_i]^{-1}\} &= \beta_2 = \text{trace}\{[\Sigma_j][\Sigma_i]^{-1}\} \\ \det\{[P_k][P_i]^{-1}\} &= \beta_3 = \det\{[\Sigma_k][\Sigma_i]^{-1}\}, \\ \text{trace}\{[P_k][P_i]^{-1}\} &= \beta_4 = \text{trace}\{[\Sigma_k][\Sigma_i]^{-1}\} \end{aligned} \quad (30)$$

The combination of both equations in Eq. (29) leads to a fifth condition for the standards' parameters:

$$\text{trace}\{[P_j][P_i]^{-1}[P_k][P_i]^{-1}\} = \beta_5 = \text{trace}\{[\Sigma_j][\Sigma_i]^{-1}[\Sigma_k][\Sigma_i]^{-1}\} \quad (31)$$

These five equations in Eqs. (30) and (31), in which only the transfer parameters of the standards and measurement values appear, can be used to compute a maximum of five unknown parameters of the standards.

The General TAN Procedure. One possibility to distribute the five unknowns among the three calibration standards is the general TAN procedure, where a Through standard (of known transmission factor t), a well-matched Attenuator (with unknown forward and reverse attenuation k_f and k_r), and a symmetrical, but not necessarily reciprocal, Network standard (with unknown reflection r and unknown transmission coefficients u_f and u_r) are used for calibration:

$$[\Sigma_T] = \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}, \quad [\Sigma_A] = \begin{bmatrix} k_r & 0 \\ 0 & 1/k_f \end{bmatrix}$$

$$[\Sigma_N] = \frac{1}{u_f} \begin{bmatrix} u_f u_r - r^2 & r \\ -r & 1 \end{bmatrix}$$

The inclusion of the known parameter t is for convenience if connector sex dictates the use of a nonzero length Through.

Since the A -standard may also turn into a well-matched Line standard (L) with $k_f = k_r = e^{-\gamma l}$, or into a double Match standard (M) with $k_f = k_r = 0$, and since the N standard may degenerate to a Reflection standard (R) with $u_f = u_r = 0$, the general TAN procedure also covers the most popular TLR and TMR self-calibration procedures.

However, the singularity of the transfer matrices $[\Sigma_A]$ and $[\Sigma_N]$ and the accompanying measurement matrices $[P_A]$ and $[P_N]$ for the case of vanishing transmission (M , R standard) must be taken care of. For that purpose, pseudo-transfer matrices are introduced (31) by separating the matrix of the measured quantities $[P]$ from its determinant Δm , which may be zero. The residual part of the matrix is denoted $[\tilde{P}]$

$$[P] = \begin{bmatrix} m_1^I & m_1^{II} \\ m_2^I & m_2^{II} \end{bmatrix} \frac{1}{m_3^I m_4^{II} - m_3^{II} m_4^I} \begin{bmatrix} m_4^{II} & -m_3^{II} \\ -m_4^I & m_3^I \end{bmatrix} = \frac{1}{\Delta m} [\tilde{P}] \quad (32)$$

with

$$\Delta m = m_3^I m_4^{II} - m_3^{II} m_4^I \quad \text{and} \quad [\tilde{P}] = \begin{bmatrix} m_1^I & m_1^{II} \\ m_2^I & m_2^{II} \end{bmatrix} \begin{bmatrix} m_4^{II} & -m_3^{II} \\ -m_4^I & m_3^I \end{bmatrix}$$

Combination of the determinant of the measurement matrix with the transfer matrix of the standard belonging to it yields the pseudo-transfer matrices

$$\begin{aligned} \Delta m_A [\Sigma_A] &= \frac{\Delta m_A}{k_f} \begin{bmatrix} k_r k_f & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} k_r & 0 \\ 0 & 1/k_f \end{bmatrix} = [\Sigma_A^P] \\ \Delta m_N [\Sigma_N] &= \frac{\Delta m_N}{u_f} \begin{bmatrix} u_f u_r - r^2 & r \\ -r & 1 \end{bmatrix} = \frac{1}{\mu_f} \begin{bmatrix} \mu_f \mu_r - \rho^2 & \rho \\ -\rho & 1 \end{bmatrix} \\ &= [\Sigma_N^P] \end{aligned}$$

with

$$\begin{aligned} \kappa_f &= \frac{k_f}{\Delta m_A}, & \kappa_r &= k_r \Delta m_A, \\ \mu_f &= \frac{u_f}{\Delta m_N}, & \mu_r &= r_r \Delta m_N, & \rho &= r \end{aligned} \quad (33)$$

and modifies Eqs. (30) and (31) as

$$\begin{aligned} \det\{[\tilde{P}_A][P_T]^{-1}\} &= \beta_1 = \det\{\Delta m_A[\Sigma_A][\Sigma_T]^{-1}\} \\ &= \det\{[\Sigma_A^p][\Sigma_T]^{-1}\} \\ \text{trace}\{[\tilde{P}_A][P_T]^{-1}\} &= \beta_2 = \text{trace}\{\Delta m_A[\Sigma_A][\Sigma_T]^{-1}\} \\ &= \text{trace}\{[\Sigma_A^p][\Sigma_T]^{-1}\} \\ \det\{[\tilde{P}_N][P_T]^{-1}\} &= \beta_3 = \det\{\Delta m_N[\Sigma_N][\Sigma_T]^{-1}\} \\ &= \det\{[\Sigma_N^p][\Sigma_T]^{-1}\} \\ \text{trace}\{[\tilde{P}_N][P_T]^{-1}\} &= \beta_4 = \text{trace}\{\Delta m_N[\Sigma_N][\Sigma_T]^{-1}\} \\ &= \text{trace}\{[\Sigma_N^p][\Sigma_T]^{-1}\} \\ \text{trace}\{[\tilde{P}_A][P_T]^{-1}[\tilde{P}_N][P_T]^{-1}\} &= \beta_5 \\ &= \text{trace}\{[\Sigma_A^p][\Sigma_T]^{-1}[\Sigma_N^p][\Sigma_T]^{-1}\} \end{aligned} \quad (34)$$

Because at least one standard must exhibit transmission, this property is, without loss of generality, attributed to the T standard, which therefore may be represented by a regular transfer matrix.

With β_i being nonsingular complex values, computed from the measurement values of the four receivers, according to Eq. (32), the first two equations of Eq. (34) can be used to determine the transmission coefficients of the A standard,

$$\frac{\kappa_r}{\kappa_f} = \beta_1, \quad \frac{\kappa_r}{t} + \frac{t}{\kappa_f} = \beta_2$$

which compute as

$$\kappa_r = t \left(\frac{\beta_2}{2} \pm \sqrt{\frac{\beta_2^2}{4} - \beta_1} \right) \quad \text{and} \quad \kappa_f = \frac{\kappa_r}{\beta_1} = \frac{t^2}{t\beta_2 - \kappa_r} \quad (35)$$

Similarly, the other parameters may be derived from the remaining equations in Eqs. (34) as

$$\mu_f = \frac{t^2 - \kappa_f \kappa_r}{t\kappa_f(\beta_5 - \kappa_r \beta_4)}, \quad \mu_r = \mu_f \beta_3 \quad (36)$$

$$\rho = \pm \sqrt{\mu_f \mu_r - \frac{\beta_4 \mu_f}{t} + \frac{1}{t^2}} \quad (37)$$

Using Eq. (33), the parameters of the transfer matrices of the standards $[\Sigma_A]$ and $[\Sigma_N]$ are computed from the above-derived pseudo-transfer parameters.

The root ambiguity in Eq. (35) is solved with a priori knowledge of the A standard's transmission factor. If the standard exhibits attenuation ($|k_r| < 1$), the root whose magnitude is below 1 will be chosen, since the standard is passive. If a line standard with $|k_f| = |k_r| \approx 1$ is used, the phase of the line's

transmission factor must be known to $\pm 90^\circ$ in order to select the correct root.

The root choice in Eq. (37) requires the same kind of information about the phase of the N standard's reflection coefficient, which must be known to $\pm 90^\circ$.

Equations (35) to (37) allow the computation of the calibration standards for the general TAN calibration procedure, also accounting for the special cases of the TMN, TMR, TAR, TLN, and TLR procedures, where the second and/or third standard are double one-ports without transmission.

Calibrating with TMN standards ($k_f = k_r = 0$), for instance, reduces Eqs. (35) to (37) to

$$\mu_f = \frac{\beta_2}{\beta_5}, \quad \mu_r = \frac{\beta_2 \beta_3}{\beta_5}, \quad \rho = \pm \sqrt{\mu_f \mu_r - \frac{\beta_4 \mu_f}{t} + \frac{1}{t^2}}$$

TAR (TLR), on the other hand, has arbitrary k_f and k_r with $\mu_f = \mu_r = 0$, reducing the set of unknowns to

$$\begin{aligned} \kappa_r &= t \left(\frac{\beta_2}{2} \pm \sqrt{\frac{\beta_2^2}{4} - \beta_1} \right), & \kappa_f &= \frac{\kappa_r}{\beta_1} \\ \rho &= \pm \frac{1}{t} \sqrt{1 - \frac{\beta_4(t^2 - \kappa_f \kappa_r)}{\kappa_f(\beta_5 - \kappa_r \beta_4)}} \end{aligned}$$

For the TMR procedure (32) $k_f = k_r = 0$ and $\mu_f = \mu_r = 0$. Only the unknown reflection r needs to be determined, yielding

$$\rho = \pm \frac{1}{t} \sqrt{1 - t \frac{\beta_2 \beta_4}{\beta_5}}$$

As before, the reflection coefficient must be known a priori to $\pm 90^\circ$ in order to resolve the root ambiguity. Technically this is done by choosing the solution that is closest to the estimated reflection coefficient of the R standard. Its choice as a short or an open is common practice (though not required) and provides the necessary information. Furthermore, the *true* value of r is determined by the self-calibration process and can be used to check the quality of the calibration, in those cases where the R standard is completely known.

It should be noted that the perfect match of the A , L , or M standard in the above-described procedures is the only absolute impedance standard and must always be a known quantity, determining the reference impedance of the calibration. The R standard is not required to be known exactly, and the only demand is that the same reflection is presented to both ports. By connecting the same reflection standard first to port 1 and then to port 2, this requirement is easily met. Possible inconsistencies between the phase-reference planes of the different standards, which may creep in without the use of self-calibration, can therefore not occur. The postulated symmetry of the R standard uniquely sets the phase reference at equal offsets to the test ports. These offsets are zero if the transmission factor t of the through corresponds to its true electrical length.

Besides allowing the calibration of the seven-term error model with partially known standards, self-calibration also provides an interesting means of verifying the quality of the calibration if the calibration standards are fully known. In this case, all deviations of the computed parameters of the

standards from their *known* values indicate imperfections of the calibration process.

Employing Self-Calibration for ϵ_r -Measurements

The scheme introduced above for determining the transmission factor of the *A* standard can be used in a straight-forward manner to obtain error-corrected measurements of the electrical length of an unknown line: Measurements of a through and the unknown line yield the measurement matrices $[P_T]$ and $[\tilde{P}_A]$, which, using Eq. (34) make β_1 and β_2 available. With the transmission factor of a line being

$$e^{-\gamma l} = k_f = k_r = \frac{k_r}{\Delta m_A}$$

Eq. (35) can be used to solve for γl . Furthermore, if the mechanical length l of the line is known, the complex propagation constant γ containing ϵ_r is available with very high precision, even though up to this point no calibration has been performed.

Self-Calibration for the 15-Term Error Model

Contrary to the above-presented self-calibration scheme for the seven-term error model, self-calibration formulas for the 15-term model can be derived without the need for transfer parameters and their associated singularity problems.

As was shown in the section entitled “The Cross-Talk Correcting 15-Term Error Model for Four-Receiver VNAs,” the 15-term error model is calibrated using five calibration measurements

$$[G] + [E][M_n] = [S_n]\{[F] + [H][M_n]\} \quad (n = 1, \dots, 5) \quad (38)$$

Using four of those equations, named *i*, *j*, *k*, and *l*, most of the error parameters can be eliminated, arriving at

$$\begin{aligned} &\Delta M_{j,k}^{-1} \Delta M_{i,j} \Delta M_{i,l}^{-1} \Delta M_{l,k} \\ &= ([H] + [F][M_k])^{-1} \Delta S_{j,k}^{-1} \Delta S_{i,j} \Delta S_{i,l}^{-1} \Delta S_{l,k} ([H] + [F][M_k]) \end{aligned} \quad (39)$$

with

$$\Delta S_{m,n} = [S_m] - [S_n], \quad \Delta M_{m,n} = [M_m] - [M_n] \quad (m, n = 1, \dots, 5)$$

(see Ref. 27 for a comprehensive outline of the derivation).

Equation (39) again constitutes a similarity transform, resulting in the two nonlinear relations

$$\begin{aligned} &\text{Trace}\{\Delta M_{j,k}^{-1} \Delta M_{i,j} \Delta M_{i,l}^{-1} \Delta M_{l,k}\} \\ &= b_1 = \text{Trace}\{\Delta S_{j,k}^{-1} \Delta S_{i,j} \Delta S_{i,l}^{-1} \Delta S_{l,k}\} \\ &\text{Det}\{\Delta M_{j,k}^{-1} \Delta M_{i,j} \Delta M_{i,l}^{-1} \Delta M_{l,k}\} \\ &= b_2 = \text{Det}\{\Delta S_{j,k}^{-1} \Delta S_{i,j} \Delta S_{i,l}^{-1} \Delta S_{l,k}\} \end{aligned} \quad (40)$$

between the measured values $[M_n]$ and the standards' *S*-parameters (invariance of the eigenvalues of similar matrices).

Because only four calibration measurements are used by Eq. (39), the fifth standard $[S_m]$, $[M_m]$ may be substituted into Eq. (39) to yield two more similarity transformations:

$$\begin{aligned} &\Delta M_{j,k}^{-1} \Delta M_{m,j} \Delta M_{m,l}^{-1} \Delta M_{l,k} \\ &= ([H] + [F][M_k])^{-1} \Delta S_{j,k}^{-1} \Delta S_{m,j} \Delta S_{m,l}^{-1} \Delta S_{l,k} ([H] + [F][M_k]) \end{aligned} \quad (41)$$

(*i*th standard replaced by *m*th standard) and

$$\begin{aligned} &\Delta M_{m,k}^{-1} \Delta M_{i,m} \Delta M_{i,l}^{-1} \Delta M_{l,k} \\ &= ([H] + [F][M_k])^{-1} \Delta S_{m,k}^{-1} \Delta S_{i,m} \Delta S_{i,l}^{-1} \Delta S_{l,k} ([H] + [F][M_k]) \end{aligned} \quad (42)$$

(*j*th standard replaced by *m*th standard).

Together with Eq. (40), the resulting trace and determinant equalities provide six nonlinear self-calibration equations, sufficient for computation of the maximally five unknowns that the 15-term error model allows for.

It is interesting to note that the derivation of Eq. (39) holds as long as the structure of the error model [Eq. (38)] is unchanged. The calibration standards and/or the measurement matrix $[M]$ may therefore also be expressed in *T*-(transfer) parameters or even in chain parameters, using voltages and currents instead of waves. Even the extension to the error model of an *N*-port VNA, as described in Ref. 24, is straightforward. In this case, all quadrants of the error matrix, the standards' *S*-(*T*-) matrix, and the measurement matrices are $N \times N$ matrices. The invariance of the eigenvalues of the two similar matrices in Eq. (39) yields *N* relations between the measured values and the standards.

The Tmrg Procedure. The self-calibration equations [Eqs. (39), (41), and (42)] permit construction of a multitude of self-calibration procedures, allowing calibration with the set of standards, best suited for a specific application. A particularly interesting variant, the Tmrg procedure, refining the Tmso procedure of the section entitled “The Cross-Talk Correcting 15-Term Error Model for Four-Receiver VNAs,” shall now be introduced.

Whereas Tmso requires a *perfect* match, short and open, the above-derived formalism allows the reflection coefficients of the short and open to be computed through self-calibration. Furthermore, a nonzero but known reflection coefficient may be used as the *m* standard, reducing its requirement from *perfect* to *known*.

Comparable with Fig. 22, Fig. 25 shows the five standards to be constructed from a Through (with known transmission factor *t*) and three reflection one-ports with reflection coefficients *m*, *r*, and *g*. The first four standards suffice for the determination of the unknown reflection coefficients *r* and *g*:

Substituting the standards *i*, *j*, *k*, and *l* into Eq. (39) and evaluating the product of *S*-matrix differences yield

$$\begin{aligned} &\Delta S_{j,k}^{-1} \Delta S_{i,j} \Delta S_{i,l}^{-1} \Delta S_{l,k} = [B] \\ &= \frac{1}{t^2 - rm} \begin{bmatrix} t^2 C_1 - m^2 C_2 & t(mC_2 - gC_1) \\ t(rC_1 - mC_2) & t^2 C_2 - rgC_1 \end{bmatrix} \end{aligned}$$

$$[S_{(k)}^T] = \begin{bmatrix} 0 & t \\ t & 0 \end{bmatrix}, \quad [S_{(j)}^m] = \begin{bmatrix} r & 0 \\ 0 & m \end{bmatrix}, \quad [S_{(i)}^{mg}] = \begin{bmatrix} m & 0 \\ 0 & g \end{bmatrix}, \quad [S_{(i)}^{gr}] = \begin{bmatrix} g & 0 \\ 0 & r \end{bmatrix}, \quad [S_{(mg)}^{rg}] = \begin{bmatrix} r & 0 \\ 0 & g \end{bmatrix}$$

Figure 25. Calibration standards for the 15-term Tmrg self-calibration procedure.

with

$$C_1 = \frac{r-m}{r-g}, \quad C_2 = \frac{g-r}{g-m}$$

resulting in

$$\begin{aligned} & \text{trace}\{\Delta M_{j,k}^{-1} \Delta M_{i,j} \Delta M_{i,l}^{-1} \Delta M_{l,k}\} \\ &= \text{trace}[B] = b_1 = \frac{m^2(g-m) + gr(m-r) + t^2(r-g)}{(rm-t^2)(m-g)} \end{aligned} \quad (43)$$

$$\det\{\Delta M_{j,k}^{-1} \Delta M_{i,j} \Delta M_{i,l}^{-1} \Delta M_{l,k}\} = \det[B] = b_2 = \frac{(t^2-mg)(r-m)}{(t^2-rm)(m-g)} \quad (44)$$

Equation (44) can be solved to yield a linear relation for r

$$r = \frac{b_2 t^2 (g-m) + m(gm-t^2)}{b_2 m(g-m) + (gm-t^2)} \quad (45)$$

and can be combined with Eq. (43) to form a quadratic equation for g

$$g^2 - g \frac{2mt^2 a_1 a_2}{m^2 a_3 - t^2 b_2^2} + \frac{t^4 a_3 - t^2 m^2 b_2^2}{m^2 a_3 - t^2 b_2^2} = 0 \quad (46)$$

with

$$a_1 = b_1 - 1 - 2b_2, \quad a_2 = b_2 + 1$$

$$a_3 = (b_1 - 1)(b_2 + 1) - b_2(2 + b_2)$$

Choosing the proper root for g requires knowledge about the sign of that reflection standard. Using a short for g and an open for r provides the necessary sign information and makes the standards sufficiently distinct for subsequent use as fully known calibration standards.

The only parameters that must be known are the transmission coefficient t of the T standard and the reflection m . The quantity m should be small for numerical reasons and must be known, but the standard is not required to be an *ideal* match.

Since only one set of reflection one-ports is physically required, the postulated equality of the reflection coefficients that enter the different standards is guaranteed, an important advantage over calibration procedures, which require the *same* reflection coefficient to be connected to both ports simultaneously.

OUTLOOK AND RELATED TOPICS

The continuously growing need for fast (production) and precision (laboratory) network measurements has led to the development of sophisticated instrumentation equipment (33–35) and many optimized calibration procedures. This process is expected to continue with the focus shifting toward even higher frequencies (>110 GHz), on-wafer cross-talk correction, and N -port measurements (36,37). A lot of interest is spawned by

the communication industry—for instance, in the characterization of IC packages (N -port measurements) and in high production throughput. Semiautomatic calibration procedures, avoiding the necessity of reconnecting the calibration standards for every calibration, address this need but are not covered here. The interested reader is referred to Ref. 38 for an overview. References 39 and 40 describe commercially available solutions.

Finally, Refs. 20 and 41 present statistical methods to assess errors not originally contained in the error model, like repeatability or contacting errors. Because these errors meanwhile constitute the accuracy limit of error-corrected network analyzer measurements, their inclusion is expected to further boost measurement accuracy.

BIBLIOGRAPHY

1. M. Sucher and J. Fox, *Handbook of Microwave Measurements*, Vol. 1, Brooklyn, NY: Polytechnic Press of the Polytechnic Institute of Brooklyn, 1963.
2. P. I. Somlo and J. D. Hunter, *Microwave Impedance Measurements, IEEE electrical measurement series*, Vol. 2, Stevenage, UK: Peregrinus, 1985.
3. G. U. Sorger, Coaxial swept-frequency VSWR measurements using slotted lines, *IEEE Trans. Instrum. Meas.*, **IM-17**: 403–412, 1968.
4. R. Caldecott, The generalized multiprobe reflectometer and its application to automated transmission line measurements, *IEEE Trans. Antennas Propag.*, **AP-21**: 550–554, 1973.
5. C. L. J. Hu, A novel approach to the design of multiple-probe high-power microwave automatic impedance measuring schemes, *IEEE Trans. Microw. Theory Tech.*, **MTT-28**: 1422–1428, 1980.
6. C. A. Hoer, The six-port coupler: a new approach to measuring voltage, current, power, impedance and phase, *IEEE Trans. Instrum. Meas.*, **IM-21**: 466–470, 1972.
7. G. F. Engen, The six-port reflectometer: an alternative network analyzer, *IEEE Trans. Instrum. Meas.*, **IM-21**: 1075–1080, 1972.
8. P. I. Somlo and J. D. Hunter, A six-port reflectometer and its complete characterization by convenient calibration procedures, *IEEE Trans. Microw. Theory Tech.*, **MTT-30**: 186–192, 1982.
9. G. F. Engen, An improved circuit for implementing the six-port technique of microwave measurements, *IEEE Trans. Microw. Theory Tech.*, **MTT-25**: 110–112, 1977.
10. D. M. Pozar, *Microwave Engineering*, Reading, MA: Addison-Wesley, 1990.
11. S. Uysal, *Nonuniform Line Microstrip Directional Couplers and Filters*, Norwood, MA: Artech House, 1993.
12. H. M. Churchill and L. Susman, A six-port automatic network analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-25**: 1086–1091, 1977.
13. G. F. Engen and C. A. Hoer, Thru-reflect-line: An improved technique for calibrating the dual six port automatic network analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-27**: 987–993, 1979.
14. H.-G. Krekels and B. Schiek, A full model calibration algorithm for a dual six-port network analyzer, *Proc. IEEE Instrum. Meas. Conf. (IMTC)*, Hamamatsu, Japan, 1994, pp. 990–993.
15. R. J. King, *Microwave Homodyne Systems*, Stevenage, UK: Peregrinus, 1978.

16. U. Gärtner and B. Schiek, A broad-band homodyne network-analyzer with binary phase-modulation, *IEEE Trans. Microw. Theory Tech.*, **MTT-34**: 902–906, 1986.
17. D. Rytting, An analysis of vector measurement accuracy enhancement techniques, Hewlett Packard, April 1981.
18. D. Rytting, Appendix to an analysis of vector measurement accuracy enhancement techniques, Hewlett Packard, April 1981.
19. W. H. Press, et al., *Numerical Recipes in C*, Cambridge, UK: Cambridge Univ. Press, 1988.
20. H. Van Hamme and M. Vanden Bosche, Flexible vector network analyzer calibration with accuracy bounds using an 8-term or a 16-term error correction model, *IEEE Trans. Microw. Theory Tech.*, **MTT-42**: 976–987, 1994.
21. H.-J. Eul and B. Schiek, A generalized theory and new calibration procedures for network analyzer self-calibration, *IEEE Trans. Microw. Theory Tech.*, **MTT-39**: 724–731, 1991.
22. A. Ferrero, Two-port network analyzer calibration using an unknown “Thru”, *IEEE Microwave Guided Wave Lett.*, **2**: 505–507, 1992.
23. B. Saswata and L. Hayden, An SOLR calibration for accurate measurement of orthogonal on-wafer DUTs, *Proc. IEEE MTT-S Int. Microw. Symp.*, Denver, CO, 1997, pp. 1335–1338.
24. R. A. Speciale, A generalization of the TSD network-analyzer calibration procedure, covering n -port scattering-parameter measurements, affected by leakage errors, *IEEE Trans. Microw. Theory Tech.*, **MTT-25**: 1100–1115, 1977.
25. K. J. Silvonen, Calibration of 16-term error-model, *Electron. Lett.*, **29**: 1544–1545, 1993.
26. H. Heuermann and B. Schiek, Results of network analyzer measurements with leakage errors corrected with the TMS-15-term procedure, *Proc. IEEE MTT-S Int. Microw. Symp.*, San Diego, CA, 1994, pp. 1361–1364.
27. A. Gronefeld and B. Schiek, Network-analyzer self-calibration with four or five standards for the 15-term error-model, *Proc. IEEE MTT-S Int. Microw. Symp.*, Denver, CO, 1997, pp. 1655–1658.
28. G. L. Madonna, A. Ferrero, and U. Piani, Multiport network-analyzer self-calibration: A new approach and some interesting results, *49th ARFTG Conf. Dig.*, Denver, CO, 1997, p. 142.
29. K. J. Silvonen, LMR 16—a self-calibration procedure for a leaky network analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-45**: 1041–1049, 1997.
30. A. Gronefeld and B. Schiek, Eine neue Methode zur Generierung von Selbstkalibrierverfahren für Netzwerkanalysatoren, anwendbar auf alle bekannten Fehlermodelle, *Kleinheubacher Berichte 1998*, Band 41, pp. 117–126.
31. H. Heuermann and B. Schiek, Robust algorithms for Txx network analyzer self-calibration procedures, *IEEE Trans. Instrum. Meas.*, **IM-43**: 18–23, 1994.
32. H. J. Eul and B. Schiek, Thru-match-reflect: One result of a rigorous theory for de-embedding and network analyzer calibration, *Proc. 18th Eur. Microw. Conf.*, Stockholm, 1988, pp. 909–914.
33. O. Ostwald and C. Evers, Vector network analyzer family ZVR, to the heart of the chart, *News from Rohde & Schwarz*, **150**: 6–9, 1996.
34. A. Wiltron, 360/37XXXA/B Series Vector Network Analyzers, Application Note AN360B/37XXXA/B-1, Mai 1996.
35. Hewlett Packard, “HP8510”, Product information.
36. A. Ferrero, F. Sanpietro, and U. Pisani, Multiport vector network analyzer calibration: A general formulation, *IEEE Trans. Microw. Theory Tech.*, **MTT-42**: 2455–2461, 1994.
37. J. C. Tippet and R. A. Speciale, A rigorous technique for measuring the scattering matrix of a multiport device with a 2-Port network-analyzer, *IEEE Trans. Microw. Theory Tech.*, **MTT-30**: 661–666, 1982.
38. H.-G. Krekels and B. Schiek, A novel procedure for an automatic network-analyzer calibration, *IEEE Trans. Instrum. Meas.*, **IM-44**: 291–294, 1995.
39. H. G. Krekels, AutoKal: Automatic calibration of vector network analyzer ZVR, *Rohde & Schwarz Application Note 1EZ30 IE*, 1996.
40. Hewlett Packard, *Ecal, Product Information*.
41. R. B. Marks, A multiline method of network analyzer calibration, *IEEE Trans. Microw. Theory Tech.*, **MTT-39**: 1205–1215, 1991.

BURKHARD SCHIEK
Oakland University
ANDREAS GRONEFELD
Ruhr-Universität Bochum

MILLIMETER-WAVE MEASUREMENT

The millimeter-wavelength spectral band covers the frequency range 30 GHz ($\lambda = 10$ mm) to 300 GHz ($\lambda = 1$ mm). In the larger view, it can include a part of the submillimeter band: the extended range up to 1 THz ($\lambda = 0.3$ mm), which represents one of the least explored portions of the electromagnetic spectrum. The frontier between the millimeter–submillimeter region and the far-infrared region is arbitrary and variable. The distinction comes mainly from the detection techniques employed (coherent or incoherent detection). The millimeter spectrum is presented in Fig. 1.

In the microwave domain, the atmosphere is transparent to frequencies up to 40 GHz except for a weak water vapor absorption line at 22 GHz. However, in the millimeter domain, there are several strong absorption lines: (1) a large and complex set of oxygen lines around 55 GHz to 60 GHz, (2) a single oxygen line around 119 GHz, and (3) a water vapor line around 183 GHz. Above 300 GHz, several absorption lines exist, mainly due to the water vapor. The spectral region located in between these lines, currently called “windows,” is decreasingly transparent when the frequency increases.

Millimeter waves offer a solution to the increasing demand in frequency allocation due to the low-frequency-band saturation and the requirement for higher data rates. Moreover, a high directivity can be obtained with small antennas associated with small sized circuits which become more easily integrable. Applications are numerous, ranging from mobile communications, local area networks, and collision avoidance radars to satellite communications, radio-astronomy, radio-altimetry, and robotics.

In the millimeter-wave range up to 100 GHz, the equipment and methods of measurement have been extended from the microwave domain. The major problems in the millimeter field are due to the small size of the devices and the transmission line losses. Above 100 GHz, as an alternative, other equipment and methods of measurement, using a quasi-optic setup, have been developed or adapted from far-infrared techniques (dielectric waveguide cavity resonator, free-space methods).

MILLIMETER-WAVE AUTOMATIC NETWORK ANALYZER

The reader is invited to consult Electric noise measurement.

As all other types of automatic network analyzer (ANA), the millimeter-wave automatic network analyzer (MWANA) measures magnitudes and phases of scattering parameters (S parameters) of the device under test (DUT).

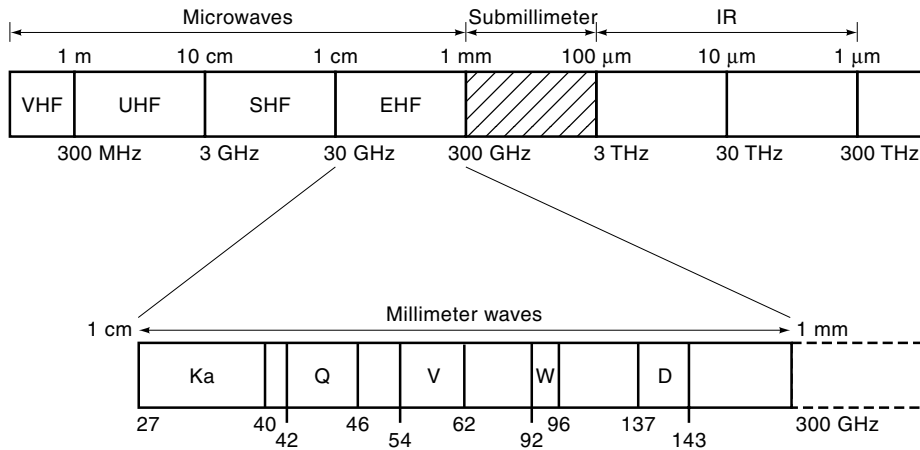


Figure 1. Atmospheric transmission in the millimeter domain.

Main Types of MWANA

Broadband Coaxial Systems. In this group, single or multiple synthesized sweeper network analyzers are commercially available (1). The single synthesized source systems perform *S*-parameter measurements up to 50 GHz using 2.4 mm coaxial accessories, and up to 67 GHz using 1.85 mm coaxial elements (V connectors). The multiple synthesized sources system may cover the 40 MHz to 110 GHz frequency range using 1 mm coaxial elements (W connectors).

Rectangular Waveguide Systems. These network analyzers perform *S*-parameter measurements in Q (33 GHz to 50 GHz), U (40 GHz to 60 GHz), V (50 GHz to 75 GHz), and W (75 GHz to 110 GHz) frequency ranges; the rectangular

waveguide standards are WR-22, WR-19, WR-15, and WR-10, respectively.

In the multiple-source network analyzer, one synthesized source provides the radiofrequency (RF) (stimulus) signal and the second provides the local oscillator (LO) signal. Figure 2 shows a simplified block diagram of this system common to all waveguide bands. This system consists of a conventional network analyzer, two microwave sources (RF and LO), and a pair of band-dependent millimeter-wave test set modules covering the frequency bands given above. The RF signal after amplification is routed to the Port 1 test set module for forward measurements (S_{11} and S_{21}) or to the Port 2 test set module for reverse measurements (S_{22} and S_{12}). Components in the millimeter-wave test set module provide frequency multiplication, signal separation to sample incident, reflected and

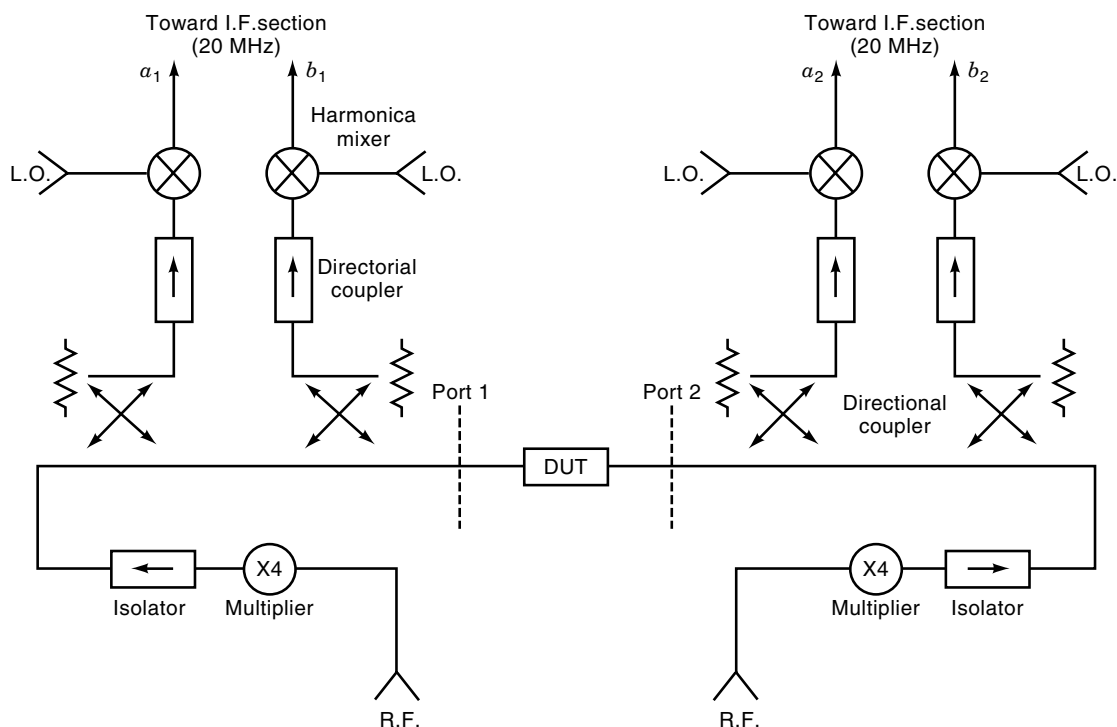


Figure 2. Simplified synopsis of the MWANA test set, for an HP85106D V-band system (50 GHz to 75 GHz).

Table 1. Dynamic Range of HP8510 C for Transmission Measurements as a Function of the Frequency Band

Frequency range (GHz):	33–50	40–60	50–75	75–110
Maximum power measured at port 2, nominal value:	+12 dBm	+10 dBm	+10 dBm	0 dBm
Reference power at port 1, nominal value:	0 dBm	0 dBm	0 dBm	–3 dBm
Minimum power measured at port 2:	–87 dBm	–87 dBm	–75 dBm	–79 dBm
Receiver dynamic range ^a :	99 dB	97 dB	85 dB	79 dB
System dynamic range ^b :	87 dB	87 dB	75 dB	75 dB

^a “Receiver dynamic range” is defined as the ratio of the maximum signal level at port 2 for 0.1 dB compression to the system noise floor.

^b “System dynamic range” is defined as the ratio of the maximum signal at port 1 and the system noise floor.

transmitted signals, and the harmonic mixers to accomplish the intermediate frequency (IF) conversion (generally first IF) to some MHz (e.g., 20 MHz). The second source provides the LO for the four harmonic mixers. This LO source is set such that the millimeter-wave RF test signal frequency and the appropriate LO harmonic are offset by exactly the IF (e.g., 20 MHz). For instance, in the case of the HP8510C MWANA (1) with V-band millimeter-wave test set modules, the frequency of the two microwave sources (RF and LO) can be expressed as follows:

$$\text{RF} = \frac{F_{\text{op}}}{4} \quad \text{and} \quad \text{LO} = \frac{(F_{\text{op}} + 20 \text{ MHz})}{14}$$

where F_{op} is the operating frequency.

As compared with a single-source network analyzer (coaxial), the rectangular waveguide system has inherent drawbacks. Indeed the power of the RF signal injected to the DUT cannot be controlled due to the frequency multiplication. This power may be close to 0 dBm (1 mW on 50 W system) and may be more dependent on the frequency band. This feature may induce nonlinear phenomena (compression, distortion) when the DUT is an active device (transistor, amplifier, etc.). Moreover, reactive impedance of a rectangular waveguide below its cutoff frequency may allow instability of the active DUT.

Dynamic Range of Millimeter-Wave Automatic Network Analyzer

Dynamic range, which is the key consideration in most measurement systems, relates to the ability of a receiver to accurately detect a signal over a large amplitude range. The largest input signal is usually limited by compression in the input receiver, while the smallest ones that can be detected are limited by the noise floor and other undesirable signals. Dynamic range can be improved by increasing the number of measurement averages and changing video IF bandwidth. Table 1 summarizes the dynamic range of HP8510C for transmission measurements as function of the frequency band.

On-Wafer Probing System

Commercially available coplanar probes cover the full millimeter-wave band (2). On-wafer probing in millimeter-wave measurement is by far the most precise technique, due to (1) better positioning and (2) better contact repeatability. For millimeter-wave measurement, only ground-signal-ground (G-S-G) topology is useful, since fundamental modes must only be excited at the probe tip. There are mainly two

types of coplanar probes: coaxial-to-coplanar probe tips and waveguide-to-coplanar probe tips. The former one uses the internal MWANA test set bias tee, while the latter one may include a direct-current (dc) bias tee inside the probe (Fig. 3). Typical values for return loss and insertion loss are, respectively, between 10 dB to 15 dB and 1 dB to 2 dB (2). The connection between the probe and the test set port must be kept as short as possible, and the millimeter-wave coaxial cable may easily add several decibels to the insertion loss.

Specific On-Wafer Calibration Technique

High-precision measurement relies on careful reference plane definition and on-chip parasitic access determination (4). Reference plane definition strongly correlates with the calibration used. For SOLT (short, open, load and through standards) calibration, the reference plane is defined by the coherent values declared for “short” and “open.” Through losses must be kept low, and delay declaration must be coherent with reference plane positioning. Any inconsistency will lead to poor measurement. For TRL (through, reflect, line standards) or LRM (line, reflects, match standards) calibration, the reference plane is always located in the center of the through, but may be moved to some other convenient place after calibration. When using on-chip standards, some or all of the on-chip access parasitics may be included in the calibration, while use of specific

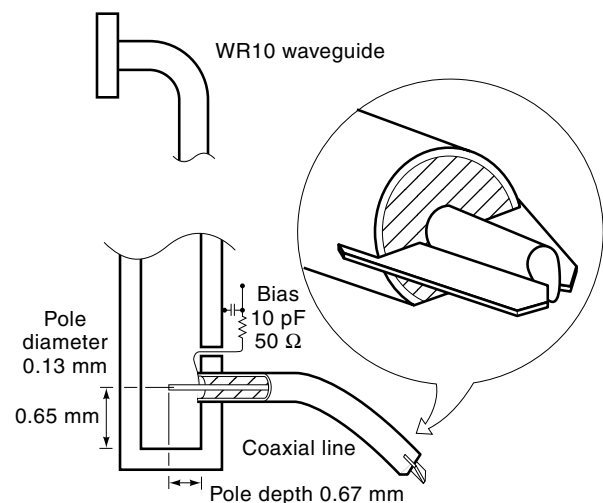


Figure 3. Mechanical structure of the waveguide-coaxial transition and the coaxial probe. (From Ref. 3.)

standards (on alumina substrate) implies subsequent determination (deembedding) of the access parasitics. In the latter case, the reference plane is usually located under the probe tip. The main error sources are (1) bad calibration and (2) bad access parasitic determination. Advanced calibration techniques have been devised for improving calibration while use of specific on-wafer elements may improve deembedding of the access parasitic.

Advanced Calibration Technique. This technique uses more standards in order to (1) obtain better standards definition (SOLT calibration) and (2) perform vector errors correction (LRM calibration and TRL calibration). None of these techniques are implemented in ANA hardware, so specific computer programs are needed.

SOLT Enhancement. The reference plane is solely determined by the short declaration (usually 0 pH). An open-ended long line measurement is performed with an incorrect open declaration and an error model to allow finding for each frequency the open declaration error, thus leading to a better frequency dependent open declaration (5). This allows precise measurement with SOLT up to 110 GHz.

LRRM Calibration. A standard LRM calibration is performed, a new reflect is measured (a short if the calibration reflect was an open), and a new set of error vectors is calculated (6). This allows us to correct for a small probe misplacement in addition to true load deviation.

NIST Multiline Calibration. The TRL calibration technique is based only on the accurate knowledge of the characteristic impedance of transmission line standards. One of the main drawbacks of TRL is its relatively narrow operating frequency range. To perform a very broadband (up to 110 GHz) TRL calibration, a multiline calibration technique has been proposed by the National Institute of Standards and Technology (NIST) (7).

Deembedding of the Access Parasitic

The use of an on-chip-specific design may allow precise determination of all access parasitics between a reference plane and a DUT port. This is an alternative to techniques based on frequency dependence of Y and Z parameters which allow parasitic determination for transistor measurement (8). The deembedding uses direct S -parameter correction or correction through precise parasitic modeling using a specially designed on-chip test device.

Direct S -Parameter Correction. The measurement of open and short placed at the DUT port position allows direct S -parameter correction, using S to Y transformation:

$$Y_{\text{device}} = ((Y_{\text{meas}} - Y_{\text{open}})^{-1} - (Y_{\text{short}} - Y_{\text{open}})^{-1})^{-1}$$

This technique is frequently used for microwave measurement on silicon devices but is also interesting in millimeter-wave measurement. However, extreme care must be taken to compensate for open capacitance (fringing field) and short inductance (ground access) when designing the specific open and short device.

Precise Parasitic Modeling. This usually uses several short, open, and through devices. A careful modeling of all these elements allows us to find out the true access parasitic and the intrinsic device parasitic. Once the access parasitic models

are known, correction of the DUT measurement are obtained through the use of a linear simulator.

Specific Characterizations of Transistors in Millimeter Wave

In view of the increasing number of applications in the centimeter-wave range, the millimeter-wave range is now largely used. MVDS (40.5 GHz to 42.5 GHz), wireless local area networks (60 GHz WLAN), and automotive radar (77 GHz) are among the most focused millimeter-wave applications today. In addition, advanced technologies are now available for manufacturing integrated circuits used in this range. The main challenge is to design this integrated circuit accurately. To this end, reliable broadband transistor models are needed for designing a millimeter-wave integrated circuit. Linear models (or equivalent circuit) including high-frequency noise sources are usually deduced from S parameter and noise-parameter on-wafer measurements. The accuracy of each element of such models depends on the measurement accuracy. The determination of equivalent circuit elements may be difficult and inaccurate in the millimeter-wave range. The key considerations in designing a reliable equivalent circuit of transistors in the millimeter-wave range are as follows:

1. The choice of calibration technique as a function of the topology of the transistor and the nature of the substrate,
2. The choice of the equivalent circuit topology including parasitic elements.

Another solution consists in establishing an equivalent circuit of transistors from S parameters and noise parameters performed in a relatively lower frequency range (for instance, up to 50 GHz). The main advantage is that the accuracy of measurement in this frequency range is better controlled than that in the millimeter-wave range. To validate the reliability of such an equivalent circuit, we calculate the S parameters and noise parameters from the elements of the equivalent circuit and we compare these calculated data with measured ones in the millimeter-wave range.

Millimeter-Wave Cryogenic On-Wafer Measurement

There are basically two different solutions depending on the temperature range. For measurements down to 200 K, the setup is similar to that of the system used for high temperature measurement. The system works at ambient pressure, only the chuck is cold, and a local overpressure of drier air or nitrogen is used to prevent icing of wafer or probe tips. In this case, the temperature gradient is mainly located on the probe itself, so cable length at low temperatures is kept minimal. The calibration substrate may be kept at room temperature.

For measurements down to a few kelvin, the device and probes are kept under vacuum in a nitrogen or helium flow cryostat. Probe displacement under vacuum is obtained through the use of a bellow, cable length is significant, and calibration and measurement must be made at the same temperature.

VOLTAGE AND POWER RATIO TECHNIQUES: SIX-PORT NETWORK ANALYZER

The voltage and power ratio techniques and the six-port network analyzer (SPNA) are based on direct detection of the

millimeter-wave. The hardware configuration of these measurement systems is simple because it is composed of diode or thermal detectors and of directional couplers or probes. In contrast, heterodyne detection systems involve multiple frequency conversions requiring local oscillators. The complexity of the measurement system makes random and systematic errors more difficult to estimate. That is why direct detection techniques provide much of the basis for precision microwave metrology. This article deals with the measurement of the scattering parameters S_{ij} of n -port millimeter devices using a slotted line, a (tuned) reflectometer, and an SPNA.

Slotted Line

This is the oldest method for measuring the reflection coefficient S_{11} of an impedance. In the millimeter frequency range, the slotted line is realized using a piece of metallic rectangular waveguide with a slot located at the center of the broad wall of the guide. The electric field inside the guide is sampled with a wire antenna connected to a Schottky diode detector. The magnitude of S_{11} is given by the voltage standing wave ratio (VSWR). The phase of S_{11} is given by the position of the antenna for which the detected voltage is minimum. This technique has been largely replaced by an automated method.

The Tuned Reflectometer

A simple reflectometer requires one or two directional couplers and power detectors in order to measure the magnitude of S_{11} . These techniques suffer from low directivity of the couplers and from the mismatches of the source and the measurement port Γ_0 . A tuned reflectometer includes tuners in order to overcome these difficulties. The measurement system is composed of a millimeter-wave source, one coupler of directivity D , one power detector, and two tuners. The detected power P may be written as follows:

$$P = K \left| \frac{S_{11} + D}{1 - S_{11}\Gamma_0} \right|^2$$

where K is a constant characterizing the measurement system.

The measurement procedure consists of successively connecting a sliding load and a sliding short in order to null D and Γ_0 using the tuners. Thereafter, the magnitude of S_{11} is given by the power ratio

$$|S_{11}| = \frac{P}{P_{cc}}$$

where P_{cc} is the power measurement when a short circuit takes place at the DUT. For a frequency equal to 110 GHz, the uncertainty measurement (defined at 2σ) of $|S_{11}|$ using the tuned reflectometer is in the range of 0.005 to 0.06 when $|S_{11}|$ varies from 0.01 to 0.5. In metrological labs, transmission measurements (S_{21}) are performed using an IF attenuator (IF substitution method).

Six-Port Network Analyzer

The name “six-port” is due to the six-port millimeter-wave junction (Fig. 4). At its four output ports it provides power readings P_3 to P_6 which are a weighted addition of the incident a_2 and reflected b_2 waves. The complex value of S_{11} (b_2/a_2) derives from

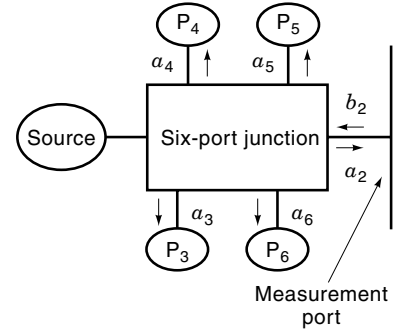


Figure 4. Six-port measurement system. It provides the complex value of the reflection coefficient S_{11} of the load connected at the measurement port. The power detector, connected at each output port, measures the power of b_i , where $i = 3$ to 6.

the six-port equations:

$$\frac{P_i}{P_3} = K_i \left| \frac{\gamma_i a_2 + \beta_i b_2}{\gamma_3 a_2} \right|^2, \quad i = 4, 5, 6$$

where α_i and β_i are the weighted factor of the waves a_2 and b_2 at the i th port and K_i is a constant of the power detector. The four scattering parameters may be obtained by the connection of two SPNAs at the two ports of the DUT or one SPNA in the reflection or transmission mode.

Practical SPNA Junctions

Six-port theory is, in principle, applicable to arbitrary design. However, for better accuracy assessment, design objectives should be obtained:

- At one output port, the wave is proportional (α_3) to the incident wave a_2 .
- At the three remaining ports we have $|q_i| = 1.5$ and $\arg(q_i - q_j) = 120^\circ$, where $q_i = -\alpha_i/\beta_i$ for $i, j = 4, 5, 6$.

A simple six-port junction consists of one directional coupler and three voltage probes (as used in the slotted line) separated by about $\lambda/6$. A similar junction replaces the probes by a waveguide coupling structure (9). This structure contains two E-plane T junctions at the upper broad wall of the main R320 (26.5 GHz to 40 GHz) waveguide and one E-plane T junction at the lower broad wall. The distances between the T junctions are about $\lambda/6$.

Figure 5 shows a six-port junction using techniques (10) at submillimeter wave (300 GHz). Similar quasi-optical techniques have been applied in optic domains for a wavelength of $0.633 \mu\text{m}$ (11). The beam splitter may be replaced by directional couplers using a metallic waveguide or dielectric waveguide structure (94 GHz) (12).

A more wideband system (13) (75 GHz to 110 GHz) has been realized by means of connecting five 3 dB 90° hybrid couplers. It can be shown that the q_i points are frequency independent and are equal to $(-j, 1 + j, -1 + j)$ assuming identical and symmetrical couplers with a coupling factor of 3 dB. This feature is interesting in the millimeter-wave range because the phase property of commercial couplers are usually unknown.

Another technique is the multistate reflectometer. It consists of two directional couplers. The internal matched

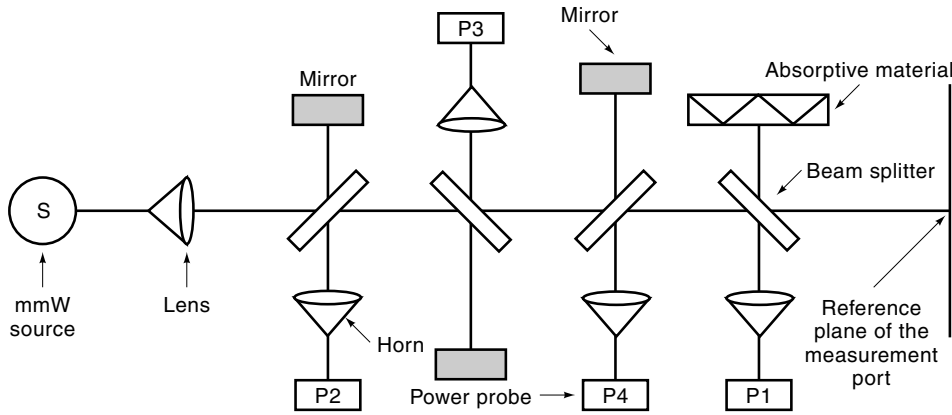


Figure 5. Six-port millimeter-wave junction using quasioptic techniques. It comprises five horns and four dielectric sheets. Each of the dielectric sheets is a beam splitter. A metallic mirror is placed on the fourth branch of each beam splitter except the one which involves the measurement of the source signal. The distance between the mirror and the dielectric sheet gives the weight of the added signals.

termination on the fourth arm of one coupler has been replaced by a phase shifter. Three states of the phase shifter provide the three equivalent power ratios of the six-port technique. Currently, this system permits on-wafer measurement at a frequency of 140 GHz (14).

Experimental Results

Table 2 shows S_{11} measurement results obtained with different systems. The measurement results labeled “SPNA,” “HP8510,” or “AB millimeter” can be compared with the values labeled “LCIE” given by the calibration center of LCIE (Laboratoire Central des Industries Electriques en France), which are considered arbitrarily to be the references. In this case, the magnitude of S_{11} as determined with a tuned reflectometer while the phase is obtained with a slotted line. The mean standard deviation is equal to 0.01 for the magnitude and 4° for the phase. The small differences may be due to temperature effects or the nonrepeatability of the connections.

Future Trends

The six-port junction may be realized using a microwave monolithic integrated circuit (MMIC). The MMIC chips can be used as a sensor in an antenna array or integrated inside the tips of a probe station. In the latter case, the series of losses of the probe tips and the line connection do not decrease the measurement accuracy of the wafer probe station.

SOURCE-PULL AND LOAD-PULL TECHNIQUE

Large-signal millimeter-wave measurements of representative samples of semiconductor devices are of prime importance for

two main reasons: (1) accuracy and consistency check of nonlinear transistor models for CAD and (2) experimental optimization of transistor optimum operating conditions without the use of any model.

Nonlinear devices demonstrate different aspects of their behavior depending on the source and load match. Therefore, large-signal measurement systems use either computer-controlled tuners or active loads to change source and load impedances of the DUT to reach the optimum matching conditions under large-signal operation (load-pull system).

Tuner systems operating up to the W band are commercially available. They are widely used for the design of low-noise amplifiers (15), power amplifiers and oscillators (16), and mixers (17). However, such systems do not allow synthesis of impedances close to the edge of the Smith chart. This main drawback becomes more and more crucial if the operating frequency increases (millimeter wave) or if on-wafer measurements are performed. For these reasons the active source and load-pull technique has emerged. Going further in the large-signal characterization, novel measurement systems allowing the extraction of voltage/current waveforms at the DUT's ports have been developed.

Basic Considerations on the Source and Load-Pull Techniques

The principle of the large-signal characterization of any nonlinear two-port is sketched in Fig. 6. If a single-tone power source is used, the four power waves are expressed as follows:

$$a_1(t) = \sum_n A_{1n} \cos(n\omega t + \varphi_{1n}), \quad b_1(t) = \sum_n B_{1n} \cos(n\omega t + \theta_{1n})$$

$$a_2(t) = \sum_n A_{2n} \cos(n\omega t + \varphi_{2n}), \quad b_2(t) = \sum_n B_{2n} \cos(n\omega t + \theta_{2n})$$

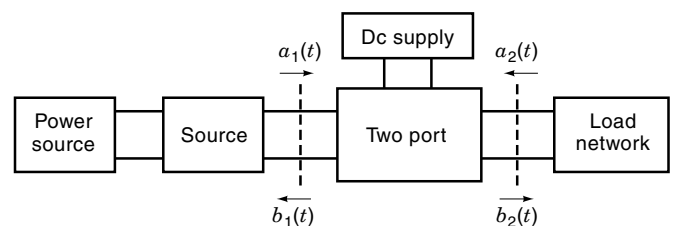


Figure 6. Source and load-pull techniques: principle.

Table 2. Measurement Comparison Among Network Analyzers

Network Analyzer	93 GHz	94 GHz	95 GHz	96 GHz
LCIE	0.415 −100.8°	0.447 174.8°	0.479 91.4°	0.515 13°
SPNA	0.01 −0.1°	−0.013 +0.2°	0.012 −0.8°	−0.013 7°
HP8510	0.005 5°	0.002 3°	0.01 5°	0.002 3°
AB millimeter	0.006 0.1°	−0.007 −4°	0.02 −3°	0.01 −1°

A vector network analyzer (VNA) or six-port reflectometer provides the measurements of the magnitudes $|A_{in}|$, $|B_{jn}|$ ($i, j = 1, 2$) and the power wave ratios at the same frequency. From this information, powers, impedances and gains can be calculated. Unfortunately, classical VNAs do not allow the measurements of absolute phases ϕ_{in} and θ_{in} . As a consequence, time-domain waveforms cannot be extracted. A novel system allowing the measurements of time-domain waveforms will be described later.

First a conventional source and load-pull system (measurements of impedances and powers) is considered. Referring to Fig. 6, a systematic approach for performing large-signal characterization of a DUT is as follows:

1. Impose desired dc voltages or currents.
2. Tune the source and load networks.
3. Sweep the power level of the input source and measure powers, efficiency, and gain.

Then the same procedure can be repeated for different operating conditions. This implies the use of a fully automated measurement system.

Multiharmonic Active Source and Load-Pull System

Multiharmonic source and load-pull systems are very useful in designing optimized nonlinear microwave circuits. Both source and load impedances have a great influence on the performances of DUTs in terms of efficiency and linearity. The load-pull characterization has also become a key step in the whole modeling process of semiconductor devices.

Load-Pull (18). The block diagram of a multiharmonic load-pull system is shown in Fig. 7. The measurements of absolute powers and power wave ratios are performed by using a VNA

(receiver operation mode) calibrated with a TRL procedure. The synthesis of load impedances at the first three harmonics coming out of the DUT is performed by using active loops and monitoring the complex gain of each loop with attenuators and phase shifters. Once the gains are fixed, a power sweep at the input of the DUT is performed and input/output power characteristics of the DUT are measured.

Source-Pull (19). Figure 8 shows a measurement system based on the use of six-port reflectometers. This system integrates both input and output active loops to perform source and load-pull measurements. Depending on the position of switch 1, the input six-port measures either the input reflection coefficient of the DUT or the reflection coefficient of the source load presented to the DUT. For both cases the error terms found by a classical calibration procedure are valid.

Time-Domain Waveform Measurement System (20)

As previously mentioned, conventional VNAs do not allow the measurements of absolute phases of harmonically related signals. As a consequence, time-domain waveforms cannot be extracted. Therefore, different institutes have developed measurement systems to extract time waveforms in one or another way usually based on the HP microwave transition analyzer. The potential of the combination of the nonlinear network measurement system (NNMS) with active source and load-pull techniques is under study.

This NNMS is mainly composed of a four-channel broadband downconverter followed by digitizers. It uses the harmonic mixing principle to convert RF fundamental and harmonics into IF fundamental and harmonics. This instrument takes the place of the VNA in the system previously presented.

The calibration of the system is performed in three main steps:

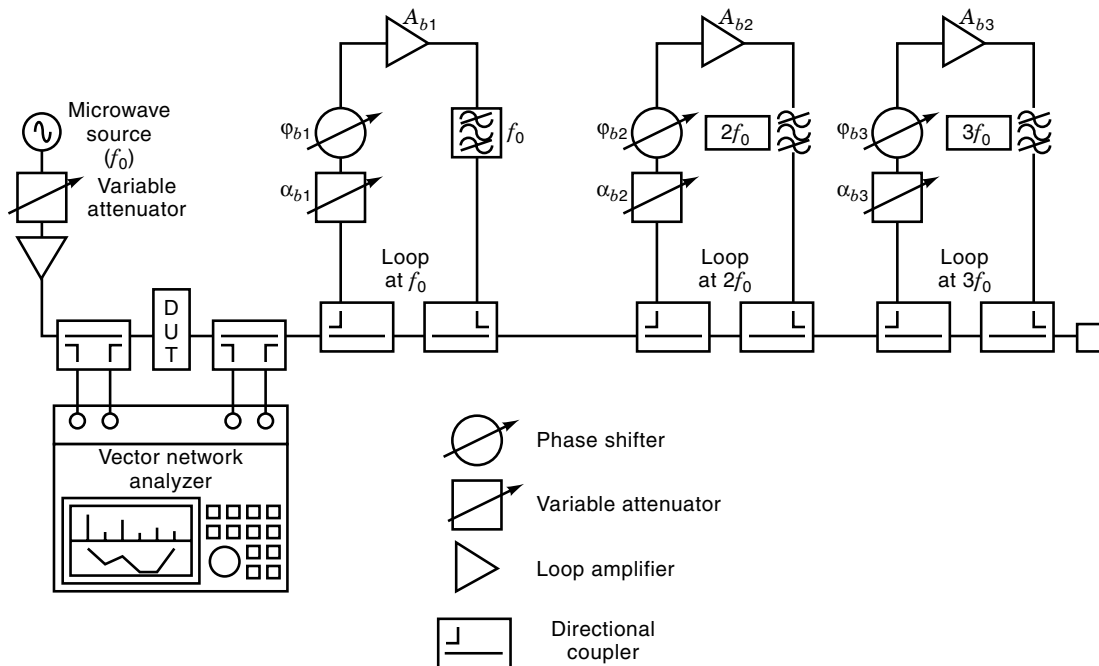


Figure 7. Multiharmonic load-pull system block diagram.

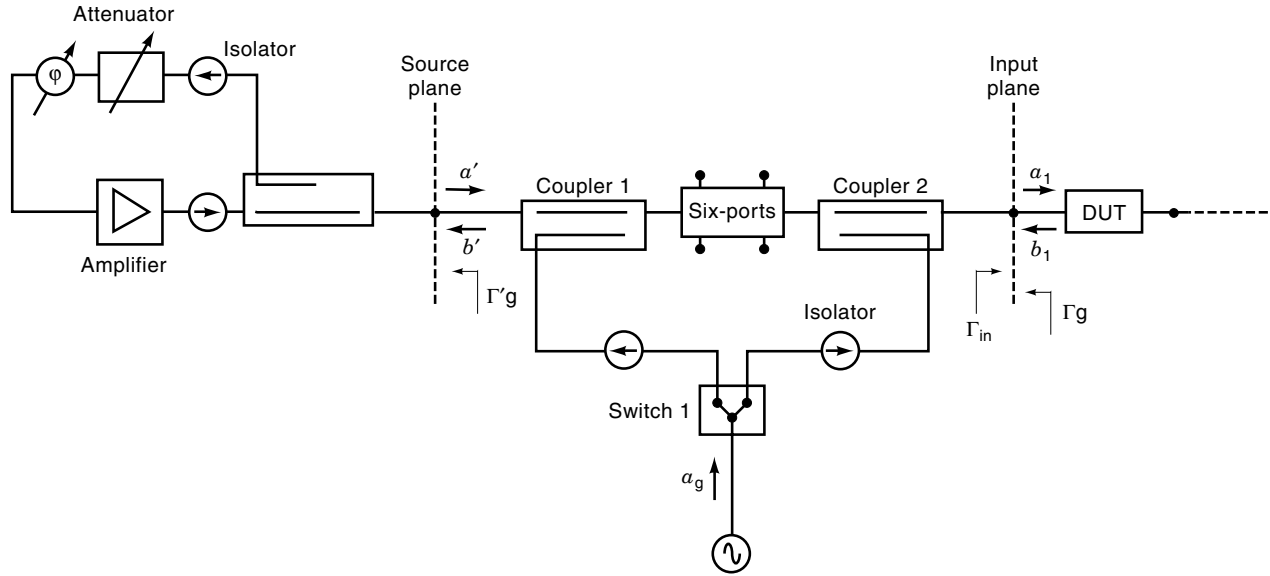


Figure 8. Source-pull implementation with a millimeter-wave six-port junction.

1. Classical TRL calibration
2. Power calibration
3. Phase calibration

During the last step, a reference generator (step recovery diode) is connected instead of the DUT (21). The reference generator is calibrated using the nose-to-nose calibration procedure (22).

DIELECTRIC WAVEGUIDE CAVITY RESONATOR

At millimeter-waves, resonators are useful for a large number of applications in communication systems and measurements of dielectric properties. In the millimeter-wave and submillimeter-wave ranges, difficulties arise from wavelengths which are very short, and devices are difficult to machine with a large degree of accuracy. So the problem is to achieve high circuit Q for volumic or hybrid millimeter-wave integrated circuits.

Different resonator structures are used. Some of them are derived from low-frequency application, like a cylindrical metallic cavity, but other devices have been developed specially for millimeter-wave measurement. In the following subsections we present mainly the devices given in Fig. 9 which are often used.

Cylindrical Metallic Cavity

This structure presented in Fig. 9(a) is composed of a cylindrical metallic waveguide closed at the top and the bottom by a metallic plane. The resonant frequency depends on the dimensions of the cavity (diameter and height) and the mode which is excited in the structures. These modes are chosen to be TE_{01n} or TM_{01n} modes and depend on excitation line position. The unloaded Q factor of these resonators increases with the axial number n . But it is difficult to use axial numbers greater than five, because a lot of modes are excited in a frequency band and it is difficult to obtain good frequency isolation. Typically, at room temperature and with copper to realize the cavity, values of unloaded Q factor are equal to 12,000 at 30 GHz and 7000 at 100 GHz on the TE_{013} modes.

Open Resonators

The most popular of this type of resonator is the Fabry-Pérot, which is presented in Fig. 9(b) (23,24). These resonators are used from the short-microwave to the optical domains (25). The basic device is composed of two reflectors of arbitrary radius of curvature separated by a length d . At low frequencies, the dimensions of the mirror will be very large, so for this reason these devices are used essentially at very high frequencies. TEM_{plq} mode is excited in these structures, where p , l , and q are, respectively, radial, azimuthal, and axial variations of

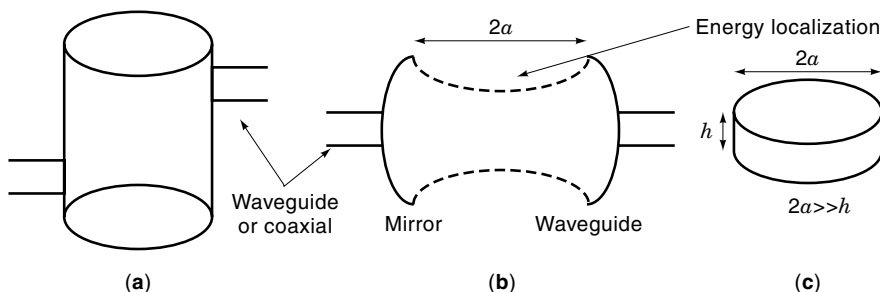


Figure 9. Example of millimeter-wave resonators. (a) Cylindrical metallic cavity. (b) Open resonators. (c) Whispering gallery dielectric resonators.

the energy which is localized in the center between the two mirrors. In a great number of applications, the TEM_{00q} mode is used and resonant frequency of these modes are periodics along the q parameter. As in metallic cavities, the unloaded Q factor increases with the number of axial variations, and values of the Q factor greater than 10^6 are possible in millimeter-wave measurement.

Dielectric Resonators

For high frequencies the dimension of resonators excited on conventional modes becomes impractically small. A solution consists of using dielectric resonators excited on whispering gallery modes (WGMs), which are higher-order modes. The first advantage of this solution is the dimension of the resonators, which is approximatively 10 times bigger than resonator excited on conventional modes. The geometry of the resonators is a disk with a diameter greater than thickness, as shown in Fig. 9(c) (26,27). So these resonators are easy to integrate in planar circuits.

Moreover, acting on these modes, energy is confined at the periphery of the dielectric resonator, and radiation losses are negligible. Thus, unloaded Q factors are very large and only limited by dielectric losses of the material used to realize the resonators.

At room temperature and using quartz material, a measured Q factor of 30,000 has been obtained at 100 GHz. Placed in a metallic cavity and at 77 K, a Q factor of 30,000,000 has been measured at 7 GHz with sapphire.

Applications to Millimeter Devices

In millimeter-wave devices, a large number of applications use resonators circuits. These elements are used in devices such as filters or oscillators, or for material measurements to determine complex permittivity and permeability. In both cases, it is very interesting to have a high Q factor of the resonances modes.

Filtering. Insertion losses and rejection depend on Q factor of the resonators. To realize these circuits, cylindrical metallic cavity or WGM dielectric resonators are suitable because association of several resonators is possible. At high frequencies, topologies of these structures are the same as for low-frequency devices.

Oscillator. Frequency stabilization and phase noise need a high Q value of the resonant device. For millimeter-waves, dielectric resonators excited on WGM give good results and are easy to integrate in the devices. With these modes, the original topology of oscillators can be realized by using the wave propagation effect at the periphery of the resonators, which is another property of these modes.

Dielectric Material Measurement. These resonator devices are currently used because they permit good accuracy with regard to the complex permittivity of the material. For metallic cavities or open resonators, the method consists of comparing the resonant frequency and the unloaded Q factor of the empty and loaded resonators. This method is convenient if the thickness of the material under test is smaller than the wavelength. For material with a large thickness, methods using WGM are suitable. In this case, measurements of resonance frequency and Q factor are compared with results obtained by electro-

magnetic simulator. These methods can be used for anisotropic dielectric or magnetic material (28).

Future Trends

The performance of millimeter-wave resonator devices is limited by the difficulty of integration of resonators in devices (in particular, for cavity or open resonators) or by losses of metallic or dielectric materials. Over the past 10 years, with the development of new dielectric materials like sapphire in the microwave domain, performances have been improved with regard to the unloaded Q factors. Unfortunately, characteristics of these materials change with temperature, and frequency stabilization is difficult to obtain without using regulating temperature devices. In the future, with technology development, we can hope to obtain material with optimum characteristics.

FREE-SPACE METHODS: INTERFEROMETRY

Waveguide loss becomes important for millimeter waves; free-space transmission has lower loss and is good for low-noise applications as well as for high-power applications (in addition, larger area of beam spread produces a lower power density). Free-space measurement is required when contact is not possible. Such is the case in radiometry for measurement of temperature and chemical composition, as well as in interferometry and in radar detection for measurement of distance, velocity, and position.

Very often for millimeter waves, the beam diameter is a relatively small number of wavelengths; thus, diffraction must be considered. A wide variety of components and systems have been developed using quasi-optical techniques, either similar to waveguide devices or derived from infrared and optical techniques (29,30).

Quasi-Optical Techniques

Gaussian Beams. Paraxial propagation of a beam in free space is relatively simple to analyze if the transverse electric field amplitude variation has a Gaussian form:

$$E(r)/E(0) = \exp[-(r/w)^2]$$

where r is the distance from the axis of propagation, and w is called a "beam radius." A Gaussian beam is produced with, or focused to, a minimum size; this minimum beam radius w_0 called a "beam waist."

The feedhorn is the best coupling device between the Gaussian beam and the guided wave (Fig. 10). The best coupling (98%) is obtained with a scalar feedhorn pattern. Several types of planar antennas (patch, bow tie, traveling-wave slot) can also be used. An associated lens is used to obtain reasonable coupling efficiency. Several types of planar antennas (patch, bow tie, traveling-wave slot) can be used. An associated lens allows us to reduce the beam size and increase the coupling efficiency.

Quasi-Optical Components Used in Millimeter-Wave Measurement. Quasi-optical components provide a wide variety of functions used for millimeter measurements:

- Beam transformations require focusing elements like parabolic or ellipsoidal mirrors and lenses. To minimize the absorptive loss of lenses, low-loss dielectrics must be

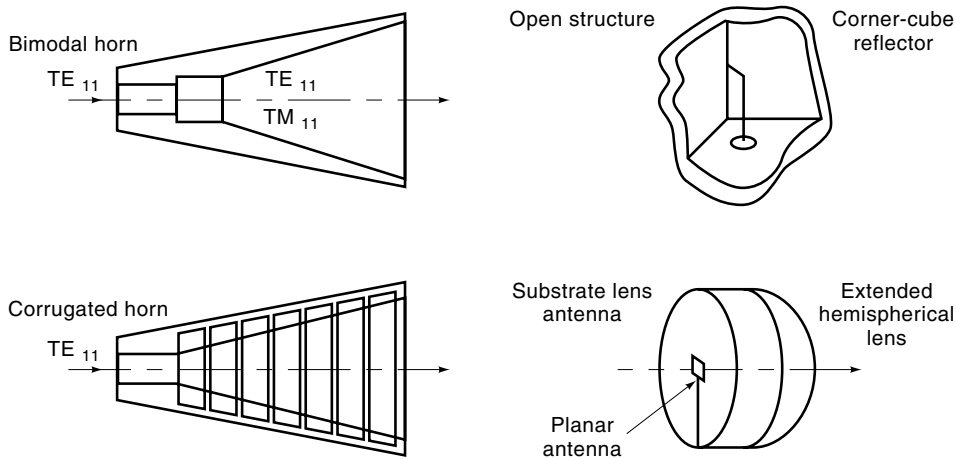


Figure 10. Devices for millimeter-wave and submillimeter-wave beam production (horn and open-structure examples).

selected (PTFE, alumina, fused silica, etc.); and to obtain low reflection loss, a matching layer or grooves are essential, except for low-index materials.

- High Q -factor ($\gg 10,000$) resonant cavities can be formed with two spherical mirrors or one spherical and one plane mirror.
- Signal filtering can be achieved by interferometers (see below) and by plate filters: Perforated conductive plates or arrays of resonant patterns are printed on a dielectric substrate.
- Polarizing grids are usual in quasi-optical systems, often used as beam splitters for a polarized signal. These grids can be formed with free-standing wires or with dielectric-supported conducting strips. Dielectric plates are also used as beam splitters and can function as hybrids (90° phase shift between reflected and transmitted beams).
- Different types of interferometers are developed from beam splitters and reflective devices: dual-beam interferometers or Fabry–Perot interferometers.

Quasi-Optical Bench. The purpose of the quasioptical bench is to create a “beam waveguide” including a sufficient measurement area. Figure 11 shows a basic bench: The measurement area is located between a signal generator and a detector equipped with free-space coupling devices (horns and lenses). The relative positions must be finely adjustable (in three rectangular directions and two or three rotating angles) while staying extremely stable. As in coaxial or waveguide measurements, generators and detectors can use frequency multipliers and heterodyne and phase-locked systems to increase the sensitivity and stability.

Free-space measurements may use a device for analog-to-coaxial calibration set parts. The methods are identical, but special care is required to (1) decrease the multiple reflections by using absorbing shields and anechoic rooms, (2) take VSWR into account, and (3) manage external (or internal) electromagnetic interferences. Another source of error and instability is atmospheric absorption, when the measuring frequency band comes over the absorption bands of an atmospheric molecule.

Free-Space Antenna Measurement

The antenna characteristics which have to be measured in the millimetric range are mainly radiation patterns in co-

and cross-polarization. Phase center measurements of primary feeds, as well as beam efficiency, also have a great importance for reflector antenna design. Measurement techniques are much the same as at lower frequencies, but with specific difficulties and requirements (31). (See also Radiometry, Electromagnetic field measurement.)

Radiation Pattern Measurements. Far-field measurements must be performed outdoors if antenna dimensions are large compared with the wavelength, which generally is the case for reflector antennas at millimetric frequencies. However, atmospheric attenuation and geographic implementation become prohibitive when the far field exceeds 1 km. Compact antenna test ranges (CATRs) remedy this problem for medium-to-large reflector antennas in the millimetric range.

For antennas of moderate (cm) dimensions, far-field measurements can be performed indoors. Horns and printed antennas are tested in anechoic chambers. In CATRs, a local plane wave is created in a zone called the “quiet zone,” by way of one or several reflectors used to collimate the beam of a smaller source. Various designs exist, ranging from the basic one with a single offset reflector to triple reflector systems, according to the required cross-polarization and spillover levels and the size of the antennas under test. Diffraction at the edges of the reflectors is less critical than in the microwave range, but reflector surfaces requirements are more stringent because the root-mean-square (rms) surface error should be better than $1/100$ wavelength to obtain good precision on the plane wave phase. Corrugated or special multimode horns are used as sources.

Hologram CATRs are being developed. Reflectors are replaced by a hologram, with a surface accuracy requirement divided by 10. This technique is thus less expensive, but it is still very new and faces problems concerning the size of the required holograms as well as frequency bandwidth limitations (20% to 30%) and polarization difficulties.

In the near-field scanning technique, fields are measured close to the antenna under test, on either a planar, cylindrical, or spherical surface. This technique requires both amplitude and phase measurements, because the sampled fields are used to calculate the radiated far-field through a near-field to far-field transformation. In the millimeter-wave domain, this technique encounters problems of time consumption and precision phase measurement.

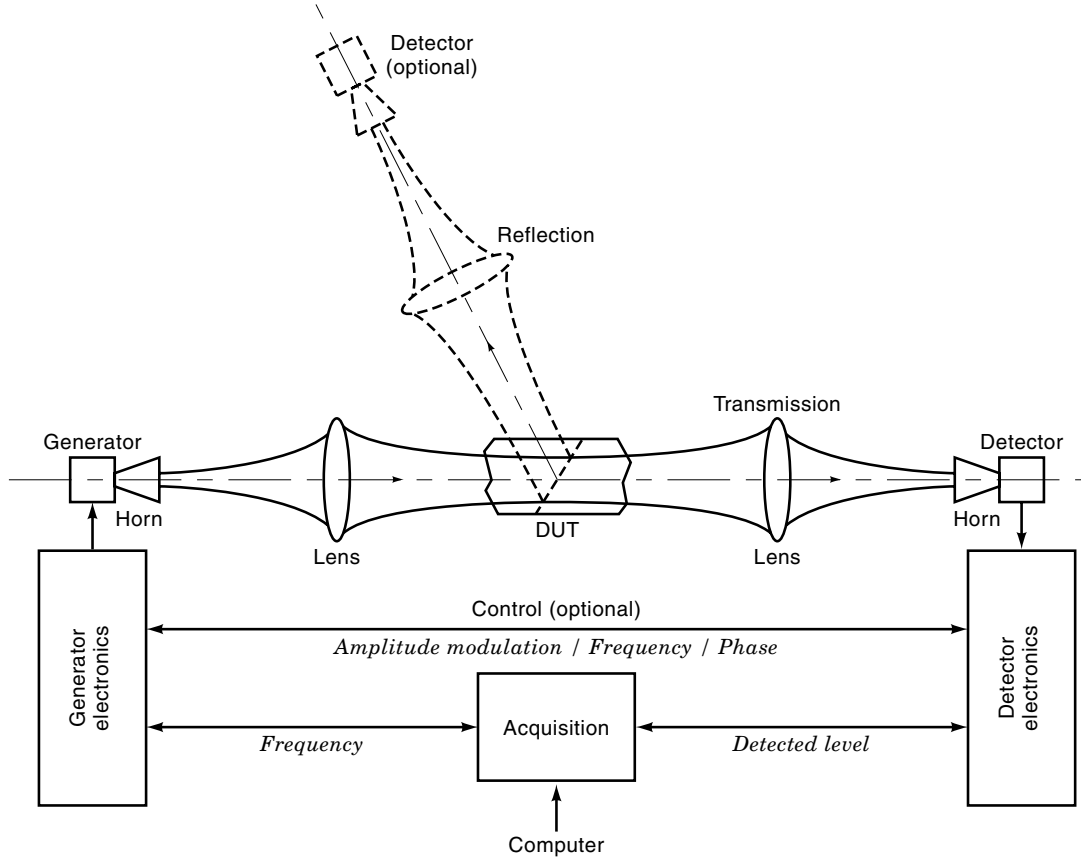


Figure 11. Quasioptical bench.

Other Antenna Performance Measurement. The “beam efficiency measurement” is performed by measuring the power radiated within the main beam of the antenna. It is especially important for radiometer antennas, which must have very low sidelobes. It requires both radiation pattern measurements, although not with wide-angle scanning, and absolute power measurements. The “phase center position measurement” is useful only for horns used as primary sources in reflector antennas. It is performed by positioning the center of phase patterns in different planes along the axis of the horn. It requires precise phase measurements and mechanical positioners.

Quasi-Optical Measurement

Power Measurement. Most of the power detectors used in the microwave measurement (Schottky diodes for instance) still work in the millimeter-wave frequency range. Moreover, bolometers and the calorimeters also operate in this range. These devices mounted in a waveguide structure can be associated with a horn to make up a beam detector. To increase the sensitivity, synchronous detection and heterodyne conversion may be used. Absolute calibration must be performed with photoacoustic detectors (used at Brewster angle and through amplitude modulation).

Quasi-Optical Device Characteristics Measurement. A basic quasi-optical bench allows us to measure the main of millimeter-wave characteristics of a DUT inserted in the optical

path: transmittance, loss and scattering by insertion, and reflection by comparison with a good reflector. Much attention must be paid to (1) the radial size of the DUT compared with the usable beam size, (2) the compensation for phase differences, and (3) the VSWR (which can be reduced by choosing an incidence angle other than zero).

For low losses, noise measurements provide a better accuracy. The equivalent noise temperature T_1 is increased to T_2 by the insertion of a DUT at the front of a low noise receiver. Giving the DUT physical temperature T_d , the loss factor L is obtained from

$$T_2 = T_d(L - 1) + LT_1(1 + \Gamma^2)$$

with $\Gamma^2 = [(VSWR - 1)/(VSWR + 1)]^2$

Noise Measurement. In addition to classic noise measurement using a noise source (diode, gas tube) associated in this case with a horn, the common noise measurement uses two absorbing targets with two different radiant temperatures. The target with the higher temperature T_h , the “hot load,” takes the place of the target with the lower temperature T_c , the “cold load,” in front of the DUT. The respective output powers are P_h and P_c . With ideal targets the equivalent noise temperature of the DUT is

$$T_x = (T_h \cdot P_c - T_c \cdot P_h)/(P_h - P_c)$$

Radiometer systems must be extremely stable over long periods and linear over the whole level range. The target size must be enough to cover the whole beam. Temperature and absorption coefficients must be homogeneous over the target surface. At millimeter wavelengths, $h\nu < kT$, the brightness temperature is very near the physical temperature. Measurement uncertainties come from variations of the effective emissivity of the target and from the mismatch with the receiver. In addition, as a result of the standing wave effect, the total noise entering the receiver becomes frequency-dependent.

The standard ferrite-loaded foam absorbers may be used as calibration targets in the lower frequency range, but for higher frequencies the reflected power unfortunately reaches -20 dB, depending on the polarization angle. For a single polarization and when the configuration is fixed, a specially developed ridged absorber or dielectric surface at the Brewster angle acts as a quasi-perfect absorber.

To achieve good precision ($<1\%$), a lot of specific targets have been developed from the principle of a conical hole with an angle of $<10^\circ$ to increase the number of reflections.

Other Quasi-Optical Measurements. The frequency measurement may use a down conversion by means of a millimeter mixer coupled with a local oscillator by a quasi-optical coupler (an interferometer or a simple dielectric plate). The local oscillator and the low frequency counter are phase/frequency locked on a reference ultra stable oscillator. On the other hand, wave-meters may use very-high- Q -factor cavities in quasi-optical techniques.

The measurement of the polarization of a signal usually takes advantage of the sensitivity of the detector for one electromagnetic field direction (detection diodes and rectangular waveguide mounts). To increase the accuracy or to use a nonpolarized detector, a polarized plate (grid) may be inserted (with an incident angle to decrease a possible VSWR effect). This method requires us to rotate the whole receiver or to insert a waveguide twist behind the horn (causing a calibration problem). The other solution is to use a quasi-optical polarization rotator on the optical path. This device uses three grids in transmission (32) or one grid and one reflector (33). In this way the mechanical rotation is limited to one light device, and the transition time may be very short (<10 ms). Circular polarization may be measured with a particular disposition of Martin-Puplett interferometer.

The knowledge of the insertion effects (loss and phase variations) allows us to compute the complex dielectric constant of a material. Other methods for material characteristics measurement use modifications of a cavity resonator's Q -factor.

Interferometry

The resolution of an antenna has a diffraction limit of $\approx \lambda/D$, and an interferometer increases the resolution according to the area covered by two or several connected antennas. More generally, an interferometer can be used to measure the Fourier components of a brightness distribution. Since Ryle and Hewish (34) have formulated the principle of aperture synthesis, many interferometers have been built or are in progress (mainly for radioastronomy). The aperture may be synthesized by multiplication, physically moving of elements, or by using the rotation of the earth. As baselines are increased, the major problem in millimeter-wave measurement is the maintenance of phase

stability of local oscillators; for very long baseline interferometry (VLBI), the local oscillators are independent and need very accurate frequency standards.

BIBLIOGRAPHY

1. Technical information available from Hewlett-Packard [Online]. Available www: <http://www.hp.com> Anritzu®.
2. Technical information available from Cascade Microtech Inc. [Online]. Available www: <http://www.cmicro.com> and Picoprobe, a GGB Industries Inc. [Online]. Available www: <http://www.picoprobe.com>
3. S. M. J. Liu and G. G. Boll, A new probe for W-band on-wafer measurements, *IEEE MTT-S Dig.*, 1993.
4. See application note from Cascade Microtech Inc., HP, and Picoprobe.
5. P. Crozat, J. C. Henaux, and G. Vernet, Precise determination of the open circuit capacitance of coplanar probes for on-wafer automatic network analyzer measurements, *Electron. Lett.*, **27**: 1476–1478, 1991.
6. A. Davidson, K. Jones, and E. Strid, LRM and LRRM calibrations with automatic determination of load inductance, *36th ARFTG Conf. Dig.*, 1990, pp. 57–63.
7. R. B. Marks, A multiline method of network analyzer calibration, *IEEE Trans. Microw. Theory Tech.*, **39**: 1205–1215, 1991.
8. G. Dambrine et al., A new method for determining the F.E.T. small signal equivalent circuit, *IEEE Trans. Microw. Theory Tech.*, **36**: 1151–1159, 1988.
9. U. Stumper, Experimental investigation of mmW Six port incorporating simple waveguide structure, *IEEE Trans. Instrum. Meas.*, **40**: 469–472, 1991.
10. U. Stumper, A six port reflectometer operating at submillimeter wavelengths, *Proc. 15th Eur. Conf.*, 1985.
11. N. C. Wolker and J. E. Carrol, Simultaneous phase and amplitude measurements on optical signals using a multiport junction, *Electron. Lett.*, **20**: 981–983, 1984.
12. G. Hjiipieris, R. J. Collier, and J. Griffin, A mmW six port using dielectric waveguide, *IEEE Trans. Microw. Theory Tech.*, **38**: 54–61, 1990.
13. S. A. Chahine et al., A six-port reflectometer calibration using Schottky diodes operating in AC detection mode, *IEEE Trans. Instrum. Meas.*, **42**: 505–510, 1993.
14. R. J. Collier and I. M. Boese, Impedance measurements using a multistate reflectometer from 110–170 GHz, *BEMC*, 3-1, 3-4, 1996.
15. Focus Microwave Inc., An ultra wideband tuner system for load pull and noise characterization, *Microw. J.*, **38** (6): 90–94, 1995.
16. F. M. Ghannouchi and R. G. Bosisio, Source pull/load pull oscillators measurements at microwave/MM wave frequencies, *IEEE Trans. Instrum. Meas.*, **41**: 32–35, 1992.
17. D. L. Le and F. M. Ghannouchi, Source-pull measurements using reverse six-port reflectometers with application to MESFET mixer design, *IEEE Trans. Microw. Theory Tech.*, **42**: 1589–1595, 1994.
18. F. Blache et al., A novel computerized multiharmonic active load pull system for the optimization of high efficiency operating classes in power transistors, *IEEE MTT Symp.*, Orlando, FL, 1995, pp. 1037–1040.
19. G. Berghoff et al., Automated characterization of HF power transistors by source pull and multiharmonic load pull, *IEEE Trans. Microw. Theory Tech.*, to be published.
20. D. Barataud et al., A novel time domain characterization technique of intermodulation in microwave transistors. Application to the visualization of the distortion of high efficiency power amplifiers, *IEEE MTT Symp.*, Denver, 1997, pp. 1687–1690.

21. J. Verspecht et al., Accurate on wafer measurement of phase and amplitude of the spectral components on incident and scattered voltage waves at the signal ports of a nonlinear microwave device, *IEEE MTT-Symp.*, Orlando, 1995, pp. 1029–1032.
22. J. Verspecht and K. Rush, Individual characterization of broadband sampling scopes with a nose to nose calibration procedure, *IEEE Trans. Instrum. Meas.*, **43**: 347–354, 1994.
23. J. C. McCleavy and K. Chang, Low-loss quasi-optical open resonators filters, *IEEE MTT Symp. Dig.*, 1991, pp. 313–316.
24. D. Steup, Quasioptical SMMW resonator with extremely high Q factor, *Microw. Opt. Technol. Lett.*, **8** (6): 275–279, 1995.
25. D. Cros and P. Guillon, Whispering gallery dielectric resonator modes for W-band devices, *IEEE Trans. Microw. Theory Tech.*, **38**: 1667–1674, 1990.
26. O. Di Monaco et al., Mode selection for a whispering gallery mode resonator, *Electron. Lett.*, **32** (7): 669–670, 1996.
27. J. Krupka et al., Study of whispering gallery modes in anisotropic single-crystal dielectric resonators, *IEEE Trans. Microw. Theory Tech.*, **42**: 56–61, 1994.
28. A. Parash, J. K. Vaid, and A. Mansinch, Measurement of dielectric parameters at microwave frequencies by cavity perturbation technique, *IEEE Trans. Microw. Theory Tech.*, **27**: 791–795, 1979.
29. P. F. Goldsmith, Quasi-optical techniques at millimeter and submillimeter wave-lengths, in K. J. Button (ed.), *Infrared and Millimeter Waves*, Vol. 6, New York: Academic Press, 1982, pp. 277–343.
30. J. C. G. Lesurf, *Millimeter-wave Optics, Devices & Systems*, Bristol: Adam Hilger, 1990.
31. A. D. Olver and C. G. Parini, Millimetre wave compact antenna test ranges, *Proc. JINA 92*, Nice, 1992, pp. 121–128.
32. R. K. Garg and M. M. Pradhan, Far-infrared characteristics of multielement interference filters using different grids, *Infrared Phys.*, **18**: 292–298, 1978.
33. C. Prigent, P. Abba, and M. Cheudin, A quasi-optical polarization rotator, *Int. J. Infrared Millimeter Waves*, **9** (5): 447–490, 1988.
34. M. Ryle and A. Hewish, *Monthly Notices R. Astron. Soc.*, 120–220, 1960.

GUY VERNET
 Université Paris—Sud
 GERARD BEAUDIN
 Observatoire de Paris—Meudon
 DOMINIQUE CROS
 Université de Limoges
 PAUL CROZAT
 Université Paris-Sud
 GILLES DAMBRINE
 Institut d'Electronique et
 Microélectronique du Nord (IEMN)
 BERNARD HUYART
 Ecole Nationale Supérieure des
 Télécommunications (ENST)
 JEAN-MICHEL NEBUS
 Université de Limoges

MECHANICAL VARIABLES

DISPLACEMENT MEASUREMENT

The measurement of displacement is an important aspect of many industrial, scientific, and engineering systems. The displacement is associated with the motion or the change in the position of an object. It is basically the vector representing a change in position of a body or from one stable point to another with respect to a reference point. In measurement systems, when the word *displacement* is used, the change in the relative position is considered to be very small—within few millimeters or less. Because displacement is a fundamental component of many instrumentation and measurements systems, many different devices have been developed to measure it ranging from a few nanometers to a few meters. A list of manufacturers is given in Table 1.

Displacement is closely associated with acceleration and velocity. Double integration of acceleration and single integration of velocity yield the displacement of an object. Therefore, some devices resembling acceleration and velocity sensors are modified for displacement measurements. Displacement sensors are manufactured in different dimensions and sizes, and they are based on many different physical principles (e.g., capacitance, magnetism, optics, piezoelectricity, resistivity). Here, the most frequently used displacement sensors will be explained.

CAPACITIVE DISPLACEMENT SENSORS

Capacitive displacement sensors enjoy wide application in industry. They satisfy the requirements of applications where high linearity and wide ranges from a few centimeters to a couple of nanometers are needed. Precision signal processing allows sensing of capacitance changes down to a few femtofarads.

The basic sensing element of a typical displacement sensor consists of two simple electrodes with capacitance C . The capacitance is a function of the distance d (meters) between the electrodes of a structure, the surface area A (square meters) of the electrodes, and the permittivity ϵ (8.85×10^{-12} F/m for air) of the dielectric between the electrodes; therefore,

$$C = f(d, A, \epsilon) \quad (1)$$

There are three basic methods for realizing a capacitive displacement sensor (i.e., by varying d , A , or ϵ , as discussed next).

Variable Plate-Distance Sensors

A variable-distance capacitive displacement sensor, made from two flat coplanar plates with a variable distance x apart, is illustrated in Fig. 1. Ignoring fringe effects, the capacitance of this arrangement may be expressed by

$$C(x) = \epsilon A/x = \epsilon_r \epsilon_0 A/x \quad (2)$$

where ϵ is the dielectric constant or permittivity, ϵ_r is the relative dielectric constant (in air and vacuum $\epsilon_r \approx 1$),

$\epsilon_0 = 8.854188 \times 10^{-12}$ F/m is the dielectric constant of vacuum, x is the distance of the plates in meters, and A is effective area of the plates in square meters.

The capacitance of this transducer is nonlinear with respect to distance x , which has hyperbolic transfer function characteristics. The sensitivity of capacitance to changes in plate separation may be found as

$$dC/dx = -\epsilon_r \epsilon_0 A/x^2 \quad (3)$$

Equation (3) indicates that the sensitivity increases as x decreases. Nevertheless, from these equations, it follows that the percent change in C is proportional to the percent change in x , which can be expressed as

$$dC/C = -dx/x \quad (4)$$

This type of sensor is often used for measuring small incremental displacements without making contact with the object.

Variable-Dielectric Displacement Sensors

The displacement may also be sensed by the relative movement of the dielectric material, attached to the object under investigation, between the plates. The corresponding equations will be

$$C = \epsilon_0 w [\epsilon_2 l - (\epsilon_2 - \epsilon_1)x] \quad (5)$$

where ϵ_1 is the relative permittivity of the dielectric material and ϵ_2 is the permittivity of the displacing material (e.g., liquid).

In this case, the output of the transducer is linear. This type of transducer is mostly used in the form of two concentric cylinders for measuring the level of nonconducting fluids in tanks.

Variable Plate-Area Displacement Sensors

The displacements may be sensed by varying the surface area of the electrodes of a flat plate capacitor, as illustrated in Fig. 2. In this case, the capacitance can be written as

$$C = \epsilon_r \epsilon_0 \frac{A - wx}{d} \quad (6)$$

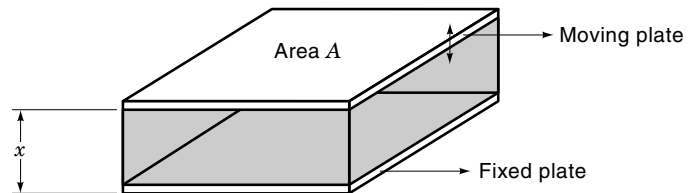


Figure 1. A variable plate-distance capacitive displacement sensor. In response to a physical stimuli, one of the plates of the capacitor moves; this varies the distance between plates. The outputs of these sensors are nonlinear with respect to distance x , having an hyperbolic transfer function characteristics. Appropriate signal processing must be employed for linearization.

Table 1. List of Manufacturers

Adsen Tech, Inc. 18310 Bedford Circle La Puente, CA 91744 Fax: 818-854-2776	Locon Sensor Systems, Inc. 1750 S. Eber Road P.O. Box 789 Holland, OH 43526 Tel: 419-865-7651 Fax: 419-865-7756
ANALITE, Inc. 24-T Newtown Plaza Plainview, NY 11803 Tel: 1-800-229-3357	Lucas 1000 Lucas Way Hampton, VA 23666 Tel: 800-745-8008 Fax: 800-745-8004
Analog Devices, Inc. 1 Technology Way P.O. Box 9106 Norwood, MA 02062-9106 Tel: 800-262-5643 Fax: 781-326-8703	Motion Sensors, Inc. 786 Pitts Chapel Road Elizabeth City, NC 27909 Tel: 919-331-2080 Fax: 919-331-1666
Dynalco Controls 3690 N. W. 53rd Street Ft. Lauderdale, FL 33309 Tel: 305-739-4300 & 800-368-6666 Fax: 305-484-3376	Rechner Electronics Industries, Inc. 8651 Buffalo Avenue Niagara Falls, NY 14304 Tel: 800-544-4106 Fax: 716-283-2127
Electro Corporation 1845 57th Street Sarasota, FL 34243 Tel: 813-355-8411 & 800-446-5762 Fax: 813-355-3120	RDP Electrosense, Inc. 2216-Dept. B Pottstown, PA Tel: 1-800-334-5838
FSI/FORK Standards, Inc. 668 Western Avenue Lombard, IL 60148-2097 Tel: 708-932-9380	Reed Switch Developments Company, Inc. P.O. Drawer 085297 Racine, WI 53408 Tel: 414-637-8848 Fax: 414-637-8861
Gordon Engineering Corp. 67 Del Mar Drive Brookfield, CT 06804 Tel: 203-775-4501	Smith Research & Technology, Inc. 205 Sutton Lane, Dept. TR-95 Colorado Springs, CO 80907 Tel: 719-634-2259 Fax: 719-634-2601
Honeywell Dept. 722 11 West Spring Street Freeport, IL 61032 Tel: 800-537-6945 Fax: 815-235-5988	Smith Systems, Inc. 6 Mill Creek Drive Box 667 Brevard, NC 28712 Tel: 704-884-3490 Fax: 704-877-3100
Hecon Corporation 15-T Meridian Road Eatontown, NJ 07724 Tel: 1-800-524-1669	Standex Electronics 4538 Camberwell Road Dept. 301L Cincinnati, OH 45209 Tel: 513-871-3777 Fax: 513-871-3779
Kaman Instrument Company 1500 Garden of the Gods Road Colorado Springs, CO 80907 Tel: 719-599-1132 & 800-552-6267 Fax: 719-599-1823	Turck, Inc. 3000 Campus Drive Minneapolis, MN 55441 Tel: 612-553-7300 & 800-544-7769 Fax: 612-553-0708
Kavlico Corporation 14501 Los Angeles Avenue Moorpark, CA 93021 Tel: 805-523-2000 Fax: 805-523-7125	Xolox Sensor Products 6932 Gettysburg Pike Ft. Wayne, IN 46804 Tel: 800-348-0744 Fax: 219-432-0828
Kistler Instrumentation Corporation Amherst, NY 14228-2171 Tel: 716-691-5100 Fax: 716-691-5226	

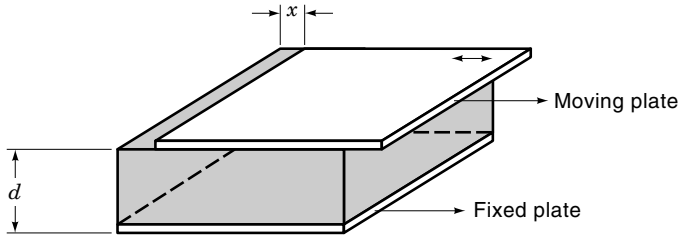


Figure 2. A variable-area capacitive displacement sensor. The sensor operates on the variation of the effective area between plates. The output is linear with displacement x . This type of sensor is normally implemented as a rotating capacitor for measuring angular displacement.

where w is the width and wx is the reduction in the area caused by the movement of the plate.

The transducer output is linear with displacement x . This type of sensor is normally implemented as a rotating capacitor for measuring angular displacement. The rotating capacitor structures are also used as an output transducer for measuring electric voltages.

Differential Capacitive Sensors

Some of the nonlinearity can be eliminated by using differential capacitive sensors. These sensors are basically three terminal capacitors, as shown in Fig. 3. Slight variations of these sensors find many applications including differential pressure measurements. In some versions, the center plate responds to physical variables while the other two plates are fixed. In the others, the center plate is fixed, and the outer plates are allowed to move. The output from the center plate is zero at the center position, and it increases as it moves left or right. The range is equal to twice the separation d . For a small displacement d , the change in capacitance will be

$$2\delta C = C_1 - C_2 = \frac{\epsilon_r \epsilon_0 l w}{(d - \delta d)} - \frac{\epsilon_r \epsilon_0 l w}{(d + \delta d)} = \frac{2\epsilon_r \epsilon_0 l w \delta d}{d^2 + \delta d^2} \quad (7)$$

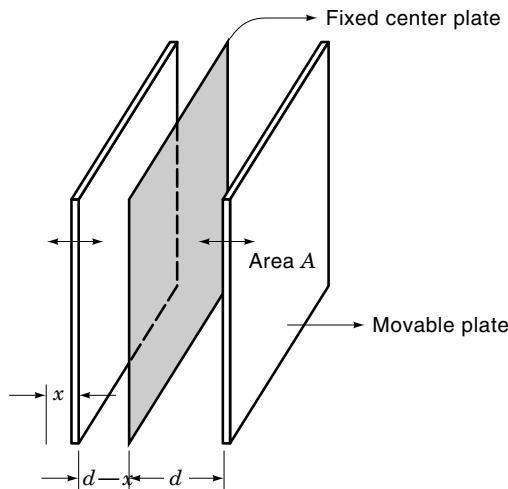


Figure 3. A differential capacitive sensor is three terminal capacitors with one fixed center plate and two outer plates. The response to motion is linear. In some versions, the center plate moves in response to a physical variable with respect to the two outer plates; in others, the center plate is fixed, and the outer plates are allowed to move.

and

$$C_1 + C_2 = 2C = \frac{\epsilon_r \epsilon_0 l w}{d - \delta d} + \frac{\epsilon_r \epsilon_0 l w}{d + \delta d} = \frac{2\epsilon_r \epsilon_0 l w d}{d^2 + \delta d^2} \quad (8)$$

Giving approximately

$$\delta C / C = \delta d / d \quad (9)$$

This indicates that the response of the device is more linear than the response of the two parallel plate types. However, some nonlinearity is still observed in practice as a result of defects in the structure. Therefore, the outputs of these types of sensors may still need to be processed carefully by appropriate signal-processing techniques.

MAGNETIC DISPLACEMENT SENSORS

Magnetic displacement sensors are commonly used in industry, particularly in manufacturing. The majority of the magnetic displacement sensors are inductive types based on Faraday's Law of Induction. They can be classified as reluctance, transformative, tachogenerator, induction potentiometer, eddy current, and magnetometer types. They are designed to measure linear as well as rotary displacements.

Linear Variable-Reluctance Types

The variable-reluctance transducers are based on change in the reluctance of a magnetic flux path. These types of transducers find applications particularly in acceleration measurements. However, they can be configured for sensing displacements and velocities. They are constructed in many different forms as described later.

Single-Coil Linear Variable-Reluctance Sensors. A typical single-coil variable-reluctance displacement sensor is illustrated in Fig. 4. The sensor consists of three elements: a

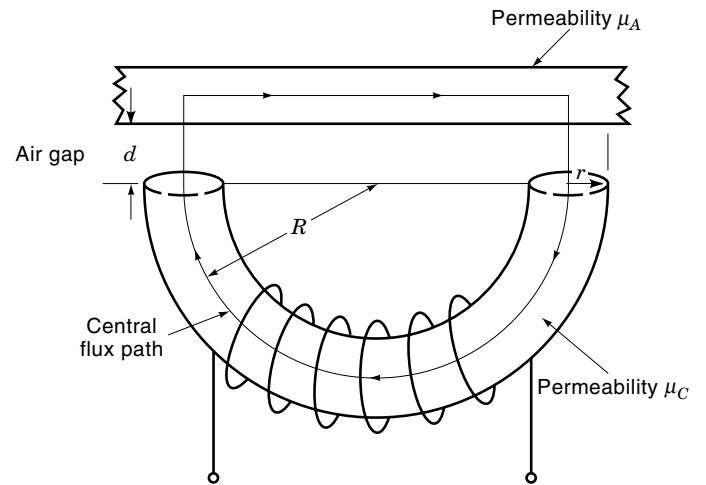


Figure 4. A typical single-coil variable-reluctance displacement sensor. The sensor consists of three elements: a ferromagnetic core in the shape of a semicircular ring, a variable air gap, and a ferromagnetic plate. The reluctance of the coil is dependent on the air gap. The air gap is the single variable, and the reluctance increases nonlinearly with the increasing gap.

ferromagnetic core in the shape of a semicircular ring, a variable air gap, and a ferromagnetic plate. The total reluctance of the magnetic circuit is the sum of the individual reluctances:

$$\mathcal{R}_T = \mathcal{R}_C + \mathcal{R}_G + \mathcal{R}_A \quad (10)$$

where $\mathcal{R}_C$, $\mathcal{R}_G$, and $\mathcal{R}_A$ are the reluctances of the core, air gap, and armature, respectively.

Each one of these reluctances can be determined from its material properties and dimensions. In this particular case, $\mathcal{R}_T$ can be approximated as

$$\mathcal{R}_T = \frac{R}{\mu_C \mu_0 r^2} + \frac{2d}{\mu_0 \pi r^2} + \frac{R}{\mu_A \mu_0 r t} \quad (11)$$

In Eq. (11), the length of the flux path in the core is taken as πR , and the cross-sectional area is assumed to be uniform with a value of πr^2 . The total length of the flux path in air is $2d$, and it is assumed that there is no fringing or bending of the flux through the air gap, so that the cross-sectional area of the flux path in air will be close to the cross section of the core. The length of an average central flux path in the armature is $2R$. The calculation of an appropriate cross section of the armature is difficult, but it may be approximated to $2rt$, where t is the thickness of the armature.

Equation (11) may be rewritten by fixing all the parameters except the air gap, which is the independent variable. Hence,

$$\mathcal{R}_T = \mathcal{R}_0 + kd \quad (12)$$

where $\mathcal{R}_0 = R/(\mu_0 r^2)[1/(\mu_C r) + 1/(\mu_A t)]$ and $k = 2/(\mu_0 \pi r^2)$. Now, the inductance can be written as

$$L = \frac{n^2}{\mathcal{R}_0 + kd} = \frac{L_0}{1 + \alpha d} \quad (13)$$

where L_0 represents the inductance at zero air gap and

$$\alpha = k/\mathcal{R}_0$$

The values of L_0 and α can be determined mathematically; they depend on the core geometry, permeability, and the like, as explained previously. From Eq. (13), it can be seen that the relationship between L and α is nonlinear. Despite this nonlinearity, this type of single-coil sensor finds application in such areas as force measurements and telemetry. In force measurements, the resultant change in inductance can be made to be a measure of the magnitude of the applied force (say on a spring). A coil usually forms one of the components of an LC oscillator whose output frequency varies with the applied force. Hence, the coil modulates the frequency of the local oscillator.

Variable-Differential Reluctance Sensors. The problem of the nonlinearity may be overcome by modifying a single-coil system into a variable-differential reluctance sensor (also known as a push-pull sensor), as shown in Fig. 5. This sensor consists of an armature moving between two identical cores separated by a fixed distance $2d$. If the core is displaced by x , Eq. (13) can now be modified to consider the effect of the two coils, as

$$L_1 = \frac{L_{01}}{1 + \alpha(d - x)} \quad L_2 = \frac{L_{02}}{1 + \alpha(d + x)} \quad (14)$$

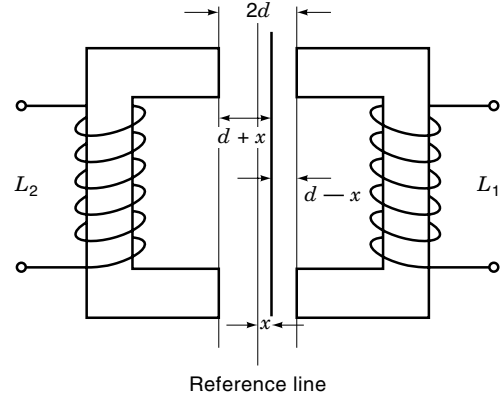


Figure 5. A variable-differential reluctance sensor consists of an armature moving between two identical cores separated by a fixed distance. The armature moves in the air gap in response to a mechanical input. This movement alters the reluctance of coils 1 and 2, thus altering their inductive properties. This arrangement overcomes the problem of nonlinearity inherent in single-coil sensors.

Although the relationship between L_1 and L_2 is still nonlinear, the sensor can be incorporated into an ac deflection bridge to give a linear output for small movements. The hysteresis errors of these transducers is almost entirely limited to the mechanical components. These sensors respond to static and dynamic displacements. They have continuous resolution and high outputs, but they may give an erratic performance in response to external magnetic fields. A typical sensor of this type has an input span of 1 cm, a coil inductance of 25 mH, and a coil resistance of 75 Ω . The resistance of the coil must be carefully considered when designing oscillator circuits. The maximum nonlinearity is 0.5%.

A typical commercially available variable differential sensor is shown in Fig. 6. The iron core is located half way between the two E-shaped frames. The flux generated by primary coils depends on the reluctance of the magnetic path, the main reluctance being the air gap. Any motion of the core increases the air gap on one side and decreases it on the other side. Consequently, reluctance changes in accordance with the principles already explained, thus inducing more voltage on one of the coils than the other. Motion in the other direction reverses the action with a 180° phase shift occurring at null. Depending on the requirements in signal processing, the output voltage can be modified by means of rectification, demodulation, and filtering. In these instruments, full-scale motion may be extremely small, on the order of few thousandths of a centimeter.

In general, variable-reluctance transducers have small ranges and are used in specialized applications such as pressure transducers. Magnetic forces imposed on the armature are quite large, and this can limit the application severely.

Linear Variable-Inductive Sensors

A typical linear-variable inductor consists of a movable iron core, which provides the mechanical input, and the two coils forming two legs of a bridge network. A typical example of such a transducer is the variable coupling transducer.

Variable-Coupling Sensors. Variable-coupling sensors consist of a former holding a center-tapped coil and a ferromagnetic

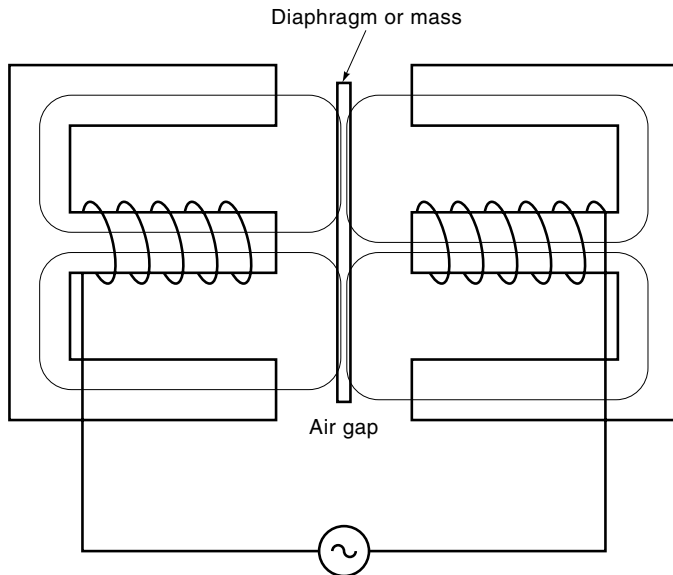


Figure 6. A typical commercial variable-differential sensor. The iron core is located half way between the two E frames. Motion of the core increases the air gap for one of the frames while decreasing for the other. This causes reluctances to change, thus inducing more voltage on one side than the other. Motion in the other direction reverses the action, with a 180° phase shift occurring at null.

plunger. The plunger and the two coils have the same length l . As the plunger moves, the inductances of the coils change. The two inductances are usually placed to form two arms of a bridge circuit with two equal balancing resistors. The bridge is excited with 5 V to 25 V ac at a frequency of 50 Hz to 5 kHz. At the selected excitation frequency, the total transducer impedance at null is set typically to the $100\ \Omega$ to $1000\ \Omega$ range. The resistors are chosen to have about the same value as the transducer impedances. The load for the bridge output must be at least ten times the resistance R value. When the plunger is in the reference position, each coil will have equal inductances of value L . As the plunger moves by δl , changes in inductances $+\delta L$ and $-\delta L$ create a voltage output from the bridge. By constructing the bridge carefully, the output voltage may be made a linear function of the displacement of the moving plunger within a rated range.

It is easy to construct transducers of this type, by simply winding a center-tapped coil on a suitable former. Variable inductance-transducers are commercially available in strokes from about 2 mm to 500 cm. The sensitivity ranges between 1% full scale to 0.02% in long-stroke special constructions. These devices are also known as linear displacement transducers (LDTs) and are available in various shape and sizes.

In addition to linear variable inductors, rotary types are available. Their cores are specially shaped for rotational motion applications. Their nonlinearity can vary between 0.5% and 1% full scale over a range of 90° rotation. Their sensitivity can be up to $100\text{ mV}/1^\circ$ of rotation.

Linear Variable-Differential Transformers

The linear variable-differential transformer (LVDT) is a passive inductive transducer finding extensive applications;

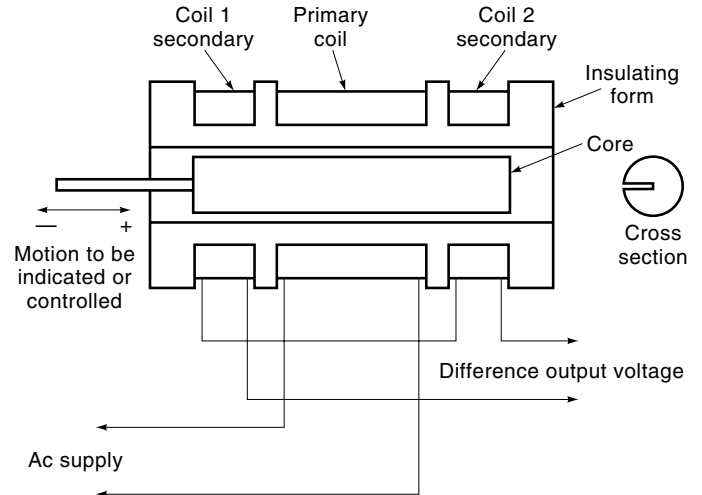


Figure 7. An LVDT is a passive inductive transducer consisting of a single primary winding positioned between two identical secondary windings wound on a tubular ferromagnetic former. As the core inside the former moves, the magnetic paths between primary and secondaries alter, thus producing secondary outputs proportional to the movement. The two secondaries are made to be as identical as possible by having equal sizes, shapes, and number of turns.

therefore, they deserve lengthy discussion here. An LVDT consists of a single primary winding positioned between two identical secondary windings wound on a tubular ferromagnetic former, as shown in Fig. 7. The primary winding is energized by a frequency of 50 Hz to 20 kHz ac. The two secondaries are made identical by having an equal number of turns. They are connected in series opposition so that the induced output voltages oppose each other.

The ferromagnetic core or plunger moves freely inside the former, thus altering the mutual inductance between the primary and secondaries. With the core in the center, or at the reference position, the induced electromagnetic fields (emfs) in the secondaries are equal; because they oppose each other, the output voltage is zero. When the core moves, say to the left, from the center, more magnetic flux links with the left-hand coil than the right-hand coil. The voltage induced in the left-hand coil is, therefore, larger than the induced emf on the right-hand coil. The magnitude of the output voltage is then larger than at the null position and is equal to the difference between the two secondary voltages. The net output voltage is in phase with the voltage of the left-hand coil. The output of the device is then an indication of displacement of the core. Similarly, movement in the opposite direction to the right from the center reverses this effect, and the output voltage is now in phase with the emf of the right-hand coil.

In many applications, the outputs are connected in opposing form, as shown in Fig. 8(a). The output voltages of the individual secondaries v_1 and v_2 at the null position are illustrated in Fig. 8(b). However, in opposing connection, any displacement in the core position x from the null point causes the amplitude of the voltage output v_o and the phase difference α to change. The output waveform v_o in relation to the core position is shown in Fig. 8(c). When the core is positioned in the middle, there is an equal coupling between the primary and the secondaries, thus giving a null point or reference point

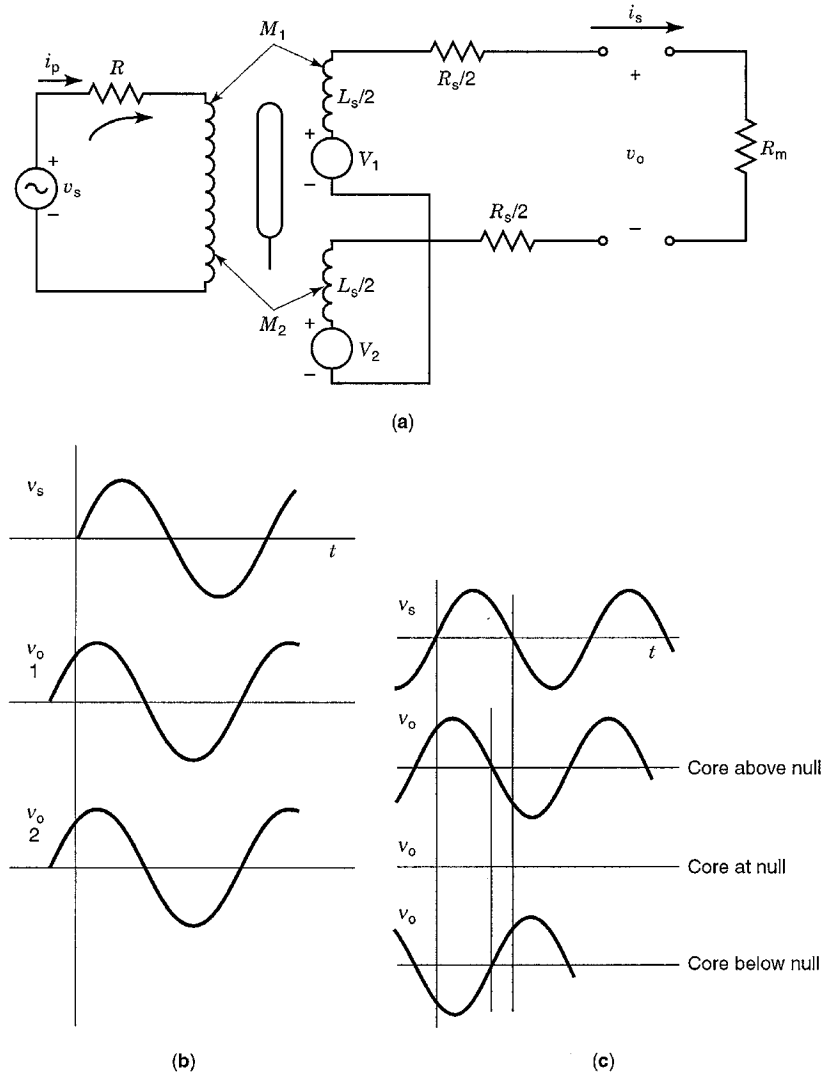


Figure 8. The voltages induced in the secondaries of a linear variable-differential transformer (a) may be processed in a number of ways. The output voltages of individual secondaries v_1 and v_2 at the null position are illustrated in (b). In this case, the voltages of individual coils are equal and in phase with each other. Sometimes the outputs are connected opposing each other, and the output waveform v_o becomes a function of core position x and phase angle α as in (c). Note the phase shift of 180° as the core position changes above and below the null position.

of the sensor. As long as the core remains near the center of the coil arrangement, the output is very linear. The linear ranges of commercial differential transformers are clearly specified, and the devices are seldom used outside this linear range.

For mathematical analysis of the operation of LVDTs, Fig. 8(a) may be used. The voltages induced in the secondary coils are dependent on the mutual inductance between the primary and individual secondary coils. Assuming that there is no cross coupling between the secondaries, the induced voltages may be written as

$$v_1 = M_1 s i_p \quad \text{and} \quad v_2 = M_2 s i_p \quad (15)$$

where M_1 and M_2 are the mutual inductances between the primary and secondary coils for a fixed core position, s is the Laplace operator, and i_p is the primary current.

In the case of opposing connection, no load output voltage v_o without any secondary current may be written as

$$v_o = v_1 - v_2 = (M_1 - M_2) s i_p \quad (16)$$

writing

$$v_s = i_p (R + s L_p) \quad (17)$$

Substituting i_p in Eq. (16) gives the transfer function of the transducer as

$$\frac{v_o}{v_s} = \frac{(M_1 - M_2) s}{R + s L_p} \quad (18)$$

However, if there is a current resulting from output signal processing, then the equations may be modified as

$$v_o = R_m i_s \quad (19)$$

where

$$i_s = \frac{(M_1 - M_2) s i_p}{R_s + R_m + s L_s}$$

and

$$v_s = i_p (R + s L_p) - (M_1 - M_2) s i_s \quad (20)$$

Eliminating i_p and i_s from Eqs. (19) and (20) results in a transfer function

$$\frac{v_o}{v_s} = \frac{R_m(M_1 - M_2)s}{[(M_1 - M_2)^2 + L_s L_p]s^2 + [L_p(R + R_m) + R L_s]s + (R_s + R_m) + R} \quad (21)$$

This is a second-order system, which indicates that with the effect of the numerator, the phase of the system changes from $+90^\circ$ at low frequencies to -90° at high frequencies. In practical applications, the supply frequency is selected such that at the null position of the core the phase angle of the system is 0° .

The amplitudes of the output voltages of the secondary coils are dependent on the position of the core. These outputs may directly be processed from each individual secondary coil for slow movements of the core, if the direction of the movement of the core does not bear any importance. However, for fast movements of the core, the signals may be converted to dc, and the direction of the movement from the null position may be detected. There are many options to do this; however, a *phase-sensitive demodulator* and filter are the most popular methods.

Phase-sensitive demodulators are extensively used in differential-type inductive sensors. They basically convert the ac outputs to dc values and also indicate the direction of movement of the core from the null position. A typical phase-sensitive demodulation circuit may be constructed, based on the diodes as shown in Fig. 9(a). This arrangement is useful for very slow displacements, usually less than 1 or 2 Hz. In Fig. 9(a), bridge 1 acts as a rectification circuit for secondary 1, and bridge 2 acts as a rectifier for secondary 2. The net output voltage is the difference between the outputs of two bridges as in Fig. 9(b). The position of the core is given by the amplitude of the dc output, and the direction of the movement of the core can be determined from the polarity of the dc voltage. For rapid movements of the core, the output of the diode bridges is filtered. This output then passes only the frequencies of the movement of the core but filters all the other frequencies produced by the modulation process. For this purpose, a suitably designed simple *RC* filter may be sufficient.

There are phase-sensitive demodulator chips available, such as AD598 offered by Analog Devices, Inc., in the marketplace. These chips are highly versatile and flexible in order to satisfy particular application requirements. These chips offer many advantages over conventional phase-sensitive demodulation devices. For example, frequency of excitation may be adjusted to any value between 20 Hz and 20 kHz by connecting an external capacitor between two pins. The amplitude of the excitation voltage can be set up to 24 V. The internal filters may be set to the required values by external capacitors. Connections to analog-to-digital converters are made easy by converting the bipolar output to a unipolar scale.

The frequency response of LVDTs is primarily limited by the mechanical inertia characteristics of the device. In general, the frequency of the applied voltage should be at least ten times the desired frequency to be measured. Commercial LVDTs are available in a broad range of sizes, and they are extensively used for displacement measurements in a variety of applications. The displacement sensors are available to cover ranges from ± 0.25 mm to ± 7.5 cm. They are sensitive enough to

respond to displacements well below $0.5 \mu\text{m}$. They can have an operational temperature range from -265° to 600°C , also are available in radiation-resistant designs for operation in nuclear reactors. For a typical sensor in the range of ± 25 mm, the recommended supply voltage is 4 V to 6 V, with a nominal frequency of 5 kHz and a maximum nonlinearity of 1% full scale. Several commercial models that can produce a voltage output of 300 mV for a 1 mm displacement of the core are available.

One important advantage of the LVDT is that there is no physical contact between the core and the coil form; hence, there is no friction or wear. Nevertheless, there are radial and longitudinal magnetic forces on the core at all times. These magnetic forces may be regarded as magnetic springs that try to displace the core from its null position. This may be a critical factor in some applications.

One problem with LVDTs is that it may not be easy to make the two halves of the secondary identical; their inductance, resistance, and capacitance may be different, causing a large unwanted quadrature output in the balance position. Precision coil-winding equipment may be required to reduce this problem to an acceptable level.

Another problem is associated with null-position adjustments. The harmonics in the supply voltage and stray capacitances result in small null voltages. The null voltage may be reduced by center-tapped voltage source arrangements and proper grounding, which reduces the capacitive effects. In center-tapped supplies, a potentiometer may be used to obtain a minimum null reading.

The LVDTs find a variety of applications including controlling jet engines and measuring roll positions in the thickness of materials in hot-slab steel mills. Force and pressure measurements may also be made from LVDTs after some mechanical modifications.

Eddy Current Displacement Sensors

Inductive transducers based on eddy currents are mainly probe types, containing two coils. One of the coils, known as the active coil, is influenced by the presence of the conducting target. The second coil, known as the balance coil, completes the bridge circuit and provides temperature compensation. The magnetic flux from the active coil passes into the conductive target by means of the probe. When the probe is brought close the target, the flux from the probe links with the target, producing eddy currents within the target.

The eddy current density is greatest at the target surface and becomes negligibly small about three skin depths below the surface. The skin depth depends on the type of material used and the excitation frequency. Even though thinner targets can be used, a minimum of three skin depths may often be necessary to minimize temperature effects. As the target comes closer to the probe, the eddy currents become stronger, causing the impedance of the active coil to change, altering the balance of the bridge in relation to target position. This unbalance voltage of the bridge may be demodulated, filtered, and linearized to produce a dc output proportional to the target displacement. The bridge oscillation frequency may be as high as 1 MHz. High frequencies allow the use of thin targets and provide a good system frequency response.

Probes are commercially available with full-scale ranges from 0.25 mm to 30 mm with a nonlinearity of 0.5% and a maximum resolution of $0.1 \mu\text{m}$. Targets are usually supplied by the

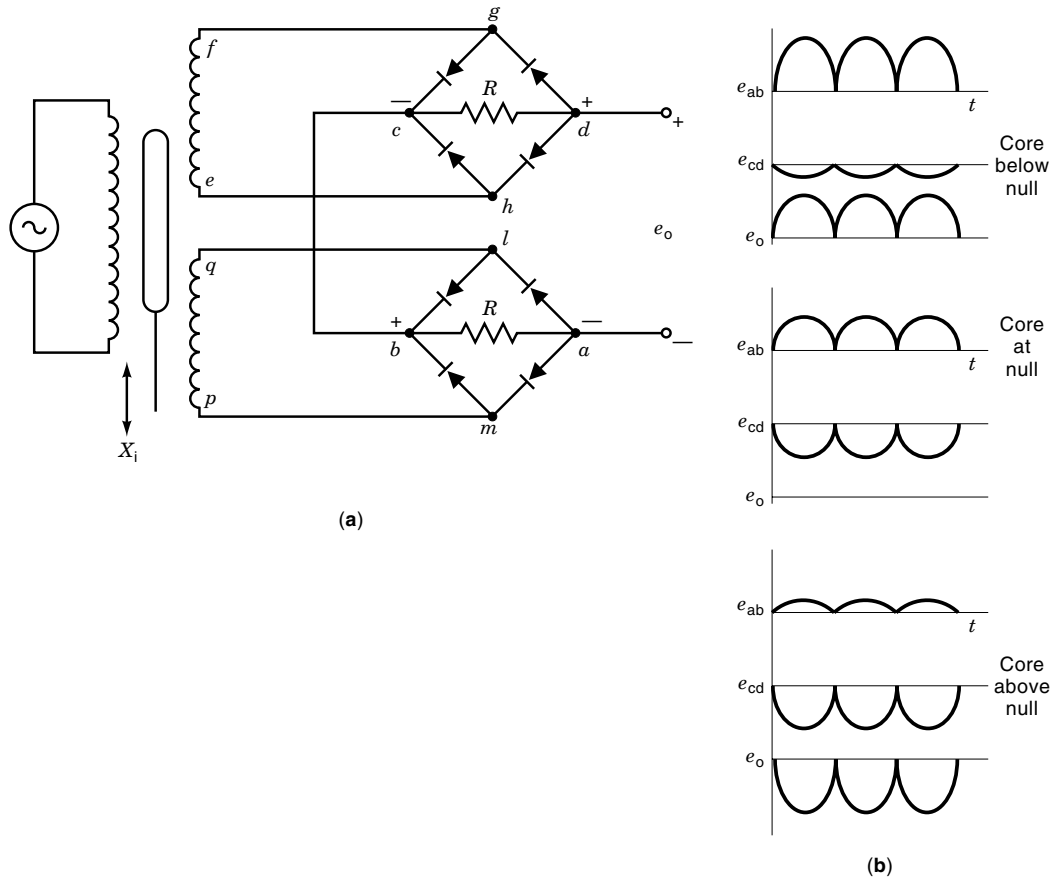


Figure 9. A typical phase-sensitive demodulation circuit based on diode bridges as in (a). Bridge 1 acts as a rectification circuit for secondary 1, bridge 2 acts as a rectifier for secondary 2, and the net output voltage is the difference between the two bridges as in (b). The position of the core can be found from the amplitude of the dc output, and the direction of the movement of the core can be determined from the polarity of the voltage. For rapid movements of the core, the outputs of the diode bridges need to be filtered. For filters, a suitably designed simple RC filter may be sufficient.

user, and often for noncontact measurements of machine parts. For nonconductive targets, conductive materials of sufficient thickness must be attached to the surface by means of commercially available adhesives. Because the target material, shape, and the like influence the output, it is necessary to calibrate the system statistically for a specific target. The recommended measuring range of a given probe begins at a standoff distance equal to about 20% of the stated range of the probe. In some cases, a standoff distance of 10% is recommended and so is calibrated as the standard for the system. A distance greater than 10% of measuring range can be used as long as the calibrated measuring range is reduced by the same amount.

Flat targets must be the same diameter as the probe or larger. If the target diameter is smaller than the probe diameter, the output drops considerably, thus becoming unreliable. Curved-surface targets may behave similarly to flat surfaces if the diameter exceeds about three or four probe diameters. In this case, the target essentially becomes an infinite plane. This also allows some cross-axis movement, which is the movement parallel to the axis of rotation, without affecting the system output. Target diameter comparable in size to the sensor can result in detrimental effects from cross-axis movements.

For curved or irregularly shaped targets, the system needs to be calibrated using the exact target that may be seen in the operation. This tends to eliminate the errors caused by the curved surfaces. However, special multiprobe systems are available for orbital motions of rotating shafts. If the curved (shaft) target is about ten times greater than the sensor diameter, it acts as an infinite plane; hence, it does not need special calibration. Special care must be exercised to deal with electrical runout caused by such factors as inhomogeneities in hardness, and are particularly valid for ferrous materials.

ANGULAR DISPLACEMENT SENSORS

Angular displacement sensors generally involve rotational capacitive or magnetic devices. The most commonly used devices involve tachogenerators, microsyns, synchros, and induction potentiometers.

Tachogenerators

An example of an angular displacement sensor is the variable-reluctance tachogenerator shown in Fig. 10. These sensors are based on Faraday's Law of Electromagnetic Induction.

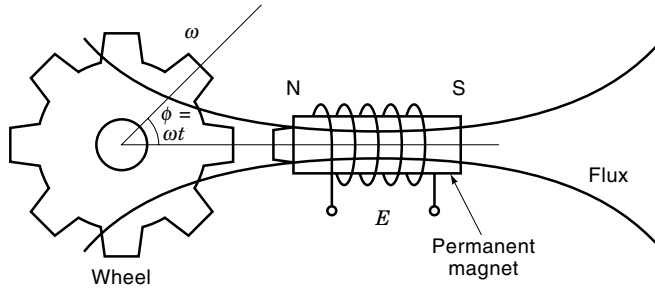


Figure 10. A variable-reluctance tachogenerator is a sensor based on Faraday's Law of Electromagnetic Induction. It consists of a ferromagnetic toothed wheel attached to the rotating shaft and a coil wound onto a permanent magnet extended by a soft iron pole piece. The wheel rotates in close proximity to the pole piece, thus causing the flux linked by the coil to change. The change in flux causes an output in the coil similar to a square waveform whose frequency depends on the speed of the rotation of the wheel and the number of teeth.

Basically, the induced emf in the sensor depends on the linear or angular velocity of the motion.

The variable-reluctance tachogenerator consists of a ferromagnetic toothed wheel attached to the rotating shaft and a coil wound onto a permanent magnet, extended by a soft iron pole piece. The wheel moves in close proximity to the pole piece, causing the flux linked by the coil to change, thus inducing an emf in the coil. The reluctance of the circuit depends on the width of the air gap between the rotating wheel and the pole piece. When the tooth is close to the pole piece, the reluctance is minimal, and it increases as the tooth moves away from the pole. If the wheel rotates with a velocity ω , the flux may mathematically be expressed as

$$\Psi(\theta) = A + B \cos m\theta \quad (22)$$

where A is the mean flux, B is the amplitude of the flux variation, and m is the number of teeth.

The induced emf is given by

$$E = \frac{-d\Psi(\theta)}{dt} = -\frac{d\Psi(\theta)}{d\theta} \times \frac{d\theta}{dt} \quad (23)$$

or

$$E = bm\omega \sin n\omega t \quad (24)$$

Both amplitude and frequency of the generated voltage at the coil are proportional to the angular velocity of the wheel. In principle, the angular velocity ω can be found from either the amplitude or the frequency of the signal. In practice, the amplitude measured may be influenced by loading effects and electrical interference. In signal processing, the frequency is the preferred option because it can be converted into digital signals easily.

Microsyn

Another commonly used example of a variable-reluctance transducer is the microsyn illustrated in Fig. 11. In this arrangement, the coils are connected in such a manner that, at the null position of the rotary element, the voltages induced in coils 1 and 3 are balanced by voltages induced in coils 2 and 4.

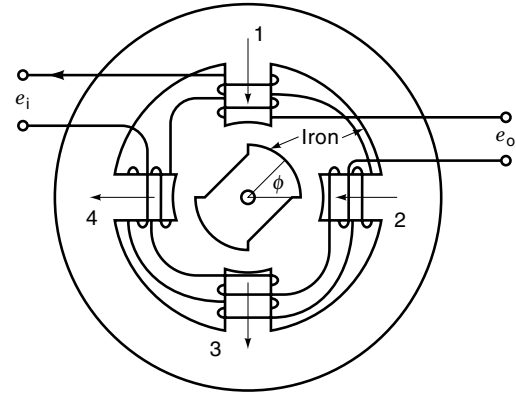


Figure 11. A microsyn is a variable-reluctance transducer, which consists of a ferromagnetic rotor and a stator carrying four coils. The stator coils are connected such that at the null position, the voltages induced in coils A and C are balanced by voltages induced in coils B and D. The motion of the rotor in one direction increases the reluctance of two opposite coils while decreasing the reluctance in others resulting in a net output voltage e_o . The movement in the opposite direction reverses this effect with a 180° phase shift.

The motion of the rotor in the clockwise direction increases the reluctance of coils 1 and 3 while decreasing the reluctance of coils 2 and 4, thus giving a net output voltage e_o . The movement in the counterclockwise direction causes a similar effect in coils 2 and 4 with a 180° phase shift. A direction-sensitive output can be obtained by using phase-sensitive demodulators, as explained in the LVDT section of this article.

Microsyn transducers are extensively used in applications involving gyroscopes. By using microsins, very small motions can be detected, giving output sensitivity as low as 0.01° of changes in angles. The output of the device can be made as high as $5 \text{ V/}1^\circ$ of rotation. The nonlinearity may vary from 0.5% to 1.0% of full scale. The main advantage of these transducers is that the rotor does not have windings and sliprings. The magnetic reaction torque is also negligible.

Synchros

The term *synchro* is associated with a family of electromechanical devices, which could be discussed under different headings. They are primarily used in angle measurements and are commonly applied in control engineering as parts of servomechanisms, machine tools, antennas, and the like.

The construction of synchros is similar to that of wound-rotor induction motors. The rotation of the motor changes the mutual inductance between the rotor coil and the three stator coils. The three voltage signals from these coils define the angular position of the rotor. Synchros are used in connection with a variety of devices, including control transformers, Scott T transformers, resolvers, phase-sensitive demodulators, and analog-to-digital (AD) converters.

In some cases, a control transformer is attached to the outputs of the stator coils such that the output of the transformer produces a resultant magnetomotive force (mmf) aligned in the same direction as that of the rotor of the synchro. In other words, the synchro rotor acts as a search coil in detecting the direction of the stator field of the control transformer. When the axis of this coil is aligned with the field, the maximum voltage is supplied to the transformer.

In other cases, ac signals from the synchros are first applied to a Scott T transformer, which produces ac voltages with amplitudes proportional to the sine and cosine of the synchro shaft angle. It is also possible to use phase-sensitive demodulation to convert the output signals to make them suitable for digital signal processing.

Induction Potentiometers

A version of the rotary-type linear inductor is the induction potentiometer. Two concentrated windings are wound on stator and rotor. The rotor winding is excited with alternating current, thus inducing voltage in the stator windings. The amplitude of the output voltage is dependent on the mutual inductance between the two coils, where the mutual inductance itself is dependent on the angle of rotation. For concentrated coil-type induction potentiometers, the variation of the amplitude is sinusoidal, so linearity is restricted to the region of the null position. A linear variation over an angle of $\pm 180^\circ$ may be obtained by carefully designed distributed coils.

Standard commercial induction pots operate in the 50 Hz to 400 Hz frequency range. They are small in size, from 1 cm to 6 cm, and their sensitivity can be in the order of $1 \text{ V}/1^\circ$ of rotation. Although the range of induction pots is limited to less than 60° of rotation, it is possible to measure displacement in angles from 0° to full rotation by suitable arrangements of a number of induction pots. As in the case of most inductive sensors, the output of the induction pots may need phase-sensitive demodulators and suitable filters. In many inductive pots, additional dummy coils are used to improve linearity and accuracy.

Rotary Variable-Differential Transformers

A variation of the linear variable-differential transformer is the rotary core differential transformer shown in Fig. 12. Here the primary winding is wound on the center leg of an E core, and the secondary windings are wound on the outer legs of the E core. The armature is rotated by an externally applied force about a pivot point above the center leg of the core. When the

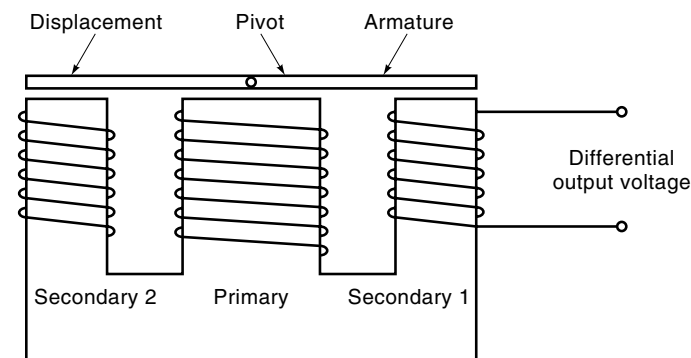


Figure 12. A rotary-core differential transformer has an E-shaped core, carrying the primary winding on the center leg and the two secondaries on the outer legs. The armature is rotated by an externally applied force about a pivot point above the center leg of the core. When the armature is displaced from its reference or balance position, the reluctance of the magnetic circuit through one secondary coil is decreased; simultaneously the reluctance through the other coil is increased. The induced emfs in the secondary windings are different in magnitude and phase as a result of the applied displacement.

armature is displaced from its reference or balance position, the reluctance of the magnetic circuit through one secondary coil is decreased; simultaneously the reluctance through the other coil is increased. The induced emfs in the secondary windings, which are equal in the reference position of the armature, are now different in magnitude and phase as a result of the displacement. The induced emfs in the secondary coils are made to oppose each other, and the transformer operates in the same manner as an LVDT. The rotating variable transformers may be sensitive to vibrations. If a dc output is required, a demodulator network may be used as in the case of LVDTs.

In most rotary linear-variable differential transformers, the rotor mass is very small, usually less than 5 g. The nonlinearity in the output ranges between $\pm 1\%$ and $\pm 3\%$, depending on the angle of rotation. The motion in the radial direction produces small output signals that can effect the overall sensitivity. But this transverse sensitivity is kept less than 1% of the longitudinal sensitivity.

LINEAR SENSORS

In this section, a number of displacement sensors, based on different principles such as resistance or piezoelectric principles, are discussed.

Resistive Displacement Sensors

In these displacement sensors, the sliding or rotational arm of a potentiometer is connected to the moving object while the main body of the potentiometer is kept in a fixed position, as illustrated in Fig. 13(a,b). Essentially, these are passive sensors requiring a stable external power source. By careful design of the potentiometers, a resolution of less than $20 \mu\text{m}$ may be obtained. They are also manufactured from conductive plastic materials, which covers a displacement span from few millimetres to 250 mm. Typical resistances are in the range of 500Ω to $100 \text{ k}\Omega$. In ideal operations, the relation between position and output voltage is linear. If the output devices draw excessive currents, the output and input relation becomes nonlinear as shown in Fig. 13(c).

A commonly used resistive displacement sensor is manufactured as a hybrid track potentiometer by depositing a conductive plastic coating on a precision wire-wound resistance track. This enables the best use of wire-wound and film technology. Although a linearity exist between the voltage and displacement, this linearity may be disturbed by the presence of load currents in the output. In many cases, the resistance of the load used with the potentiometer must be several times greater than the potentiometer resistance to avoid severe nonlinearity effects.

Strain Gage Sensors

Strain gage displacement sensors are based on resistance properties of electrical conductors. If a conductor is stretched or compressed, its resistance alters because of dimensional changes and changes in the fundamental property of material called *piezoresistance*. This indicates that the resistivity ρ of the conductor depends on the mechanical strain applied onto it. The dependence is expressed as the gage factor

$$\frac{dR/R}{dL/L} = 1 + 2\nu + \frac{d\rho/\rho}{dL/L} \quad (25)$$

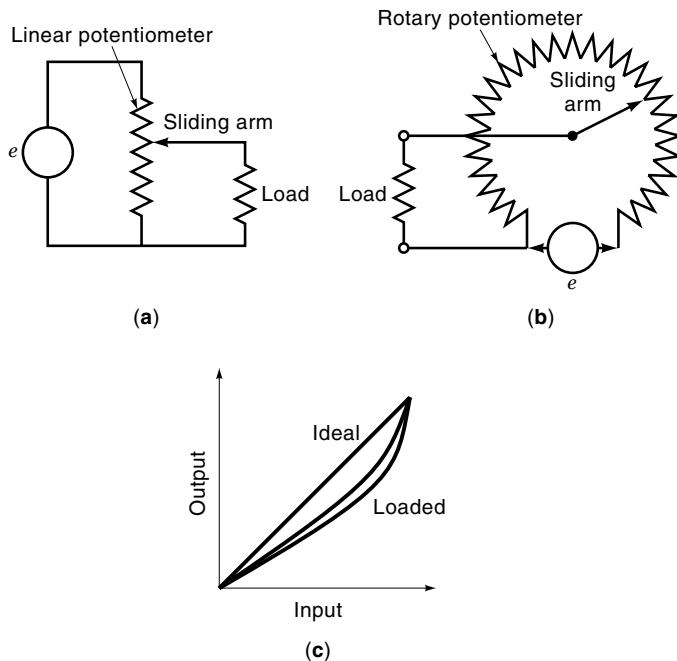


Figure 13. Resistive displacement sensors. The object under investigation is connected to the sliding arm of the potentiometer. They can measure linear motion or rotary motion as shown in (a) and (b), respectively. A loading effect occurs if the output devices have low impedance (c).

where l indicates the resistance change due to length, 2ν indicates resistance change due to area, and $(d\rho/\rho)/(dL/L)$ indicates the resistance change due to piezoresistivity.

In displacement measurements, the resistance strain gages can be selected from different types, such as unbonded metal-wire gages, bonded metal-wire gages, bonded metal-foil gages, vacuum deposited thin-metal-film gages, bonded semiconductor gages, and diffused semiconductor gages. But, generally, bonded and unbonded metal-wire gages find wider applications. Occasionally bonded semiconductor gages, known as piezoresistive transducers, are used, but they suffer from high-temperature sensitivities, nonlinearity, and some mounting difficulties.

Unbonded strain-gage displacement sensors use the strain wires as spring elements and as the motion transducer, using arrangements similar to those in Fig. 14. They are useful for general-purpose motion and vibration measurements from low- to medium-frequency displacements. They are often manufactured in the form of piezoelectric ceramic cartridges giving inexpensive but highly reliable displacement measurements.

Piezoresistive Displacement Sensors

Piezoresistive displacement sensors are essentially semiconductor strain gages with large gage factors. High gage factors are obtained because the material resistivity depends primarily on the stress, not on dimensions only. Most piezoresistive sensors use two or four active gages arranged in a Wheatstone bridge. Extra precision resistors are used (as part of the circuit) in series with the input to control the sensitivity, balancing and offsetting temperature effects. The mechanical construction of a piezoresistive displacement sensors is similar to the installation in Fig. 14.

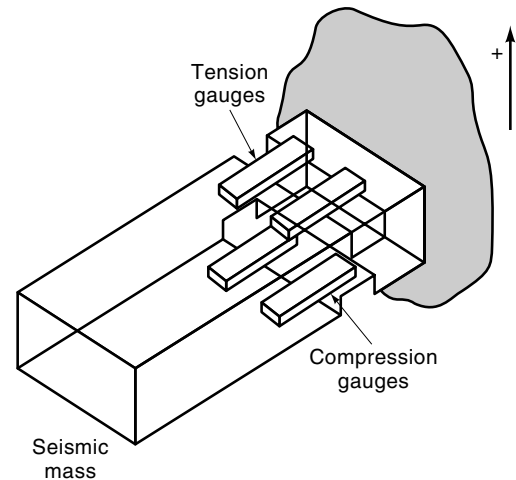


Figure 14. Bonding of piezoelectric and piezoresistive elements onto an inertial system. As the inertial member is displaced, deformation of the tension and compression gages causes the resistance to change. The change in resistance is picked up and processed further.

In some applications, overload stops are necessary to protect the gauges from high-amplitude inputs. These instrument is also useful for acquiring vibration information at low frequencies (e.g., below 1 Hz).

OPTICAL AND LASER DISPLACEMENT SENSORS

Optical methods are commonly used in displacement sensing particularly where high precision and small displacements are involved. There are many different techniques available, some of which are optical fibers, laser methods, encoders, and interferometric types. Here, some common methods will be described.

Optical Fibers

Optical fibers are suitable to measure linear and angular displacements in a noncontacting manner. Figure 15(a,b) illustrates two typical arrangements in which the intensity of the radiation incident on the detector is dependent on the relative displacement of the optical fibers. In Fig. 15(a) one of the fibers is fixed and the other is allowed to move with respect to displacement. The maximum intensity is detected when the fibers are at closest proximity and axially aligned. As the distance x between fibers increases, the light entering the second fiber decreases. Hence, the output of the detector can be related to displacement. In this arrangement, the displacement must be strictly in one direction, in a straight line of x or in y directions only. Some of the problems of using a two-fiber arrangement can be eliminated by using multiple fibers as exemplified in Fig. 16(a). In this case, the emitting fiber is fixed, and the distance between the emitting fiber and the receiving fibers is kept constant. As the fibers move in the y and x directions, the intensity of light received by each fiber varies. These variations are related to the displacement of the object. A slight modification of this arrangement can lead to reflection-type arrangements in which case the emitting and receiving fibers are bunched together, as shown in Fig. 16(b). As the reflecting surface is displaced, the detected intensity of light is varied.

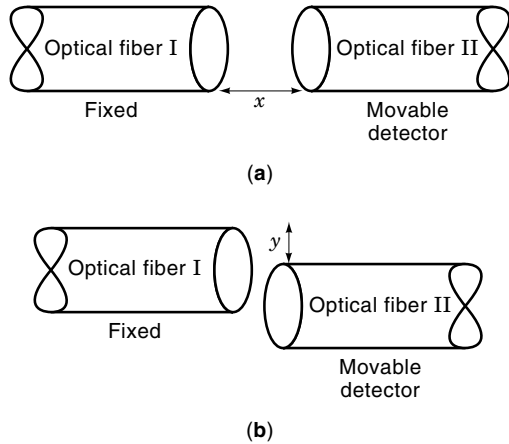


Figure 15. Examples of optical fiber displacement sensor. The intensity of the radiation incident on the detector is dependent on the relative displacement of the optical fibers. The maximum intensity of light is detected when the fibers are at closest proximity and axially aligned. Arrangement in (a) detects displacement in the x direction and that in (b) detects in the y direction.

Another common optical fiber application uses microbending principles, as illustrated in Fig. 17. As the displacement between the fixed and the displaced plates varies, the changes in the fiber dimensions between the two plates alters such fiber

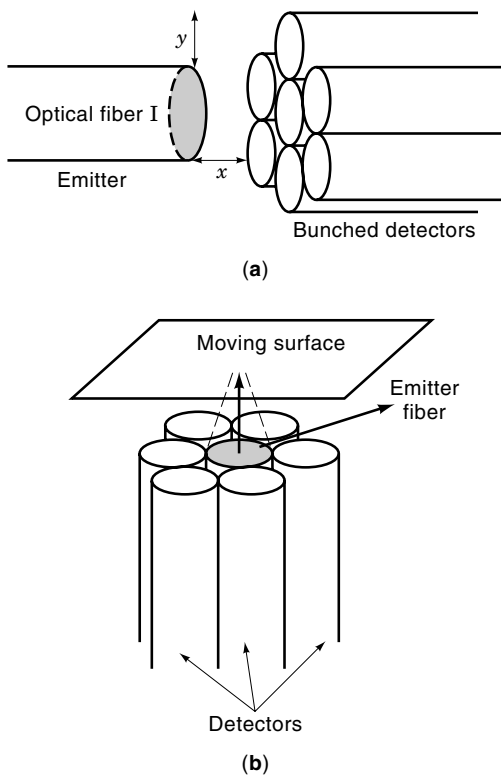


Figure 16. Multiple fiber displacement sensors. In some types, emitting fibers are fixed, and the distance between the emitting fiber and the receiving fibers is kept constant allowing vertical movements of the detectors. As the fibers move in the y and z directions, the intensity of light received by each fiber varies. In other types, the emitting and receiving fibers are bunched together, and the reflections from the moving surface are detected.

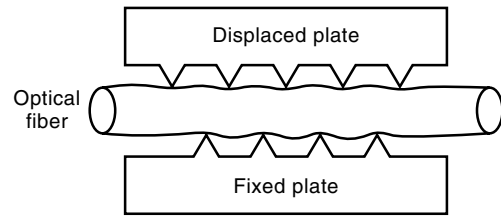


Figure 17. Microbend optical fiber displacement sensor. The displacement between the fixed plate and the displaced plate varies the fiber characteristics between the two plates, altering the attenuation and absorption. These changes are detected suitably as a measure of the displacement.

characteristics as attenuation and absorption. These changes are detected suitably as a measure of the displacement.

In all fiber optic displacement sensors, the variations in intensity and amplitudes are detected. For accurate operations, the source intensity must be stable.

Interferometric Sensors

There are many different types of interferometric sensors including optical fiber, laser, holographic, and monochromatic light types. In these sensors, Michelson, Sagnac, or Fizeau effects of interference of two or more light beams is used by suitable optical arrangements. The most commonly used optical arrangements are beam splitters, polarizing prisms, and diffraction gratings.

As two light beams interfere, interactions of light cause a series of bright and dark lines. If, for example, two interfering lights with wavelengths $1.5 \mu\text{m}$ are interfered. By suitable arrangements as in Fig. 18, the spacing of bright and dark lines will be $0.75 \mu\text{m}$. In many cases, the dark and bright points are counted in relation to the displacement of an object located within the interference. The laser interferometric sensors operate on precise wavelength monochromatic light and are the most accurate in these family of sensors. An accuracy better than 1 nm can be obtained. Many different designs based on different displacement mechanisms are available. Some of these include fringe counting interferometers, frequency modulation interferometers, heterodyne interferometers, phase-locked

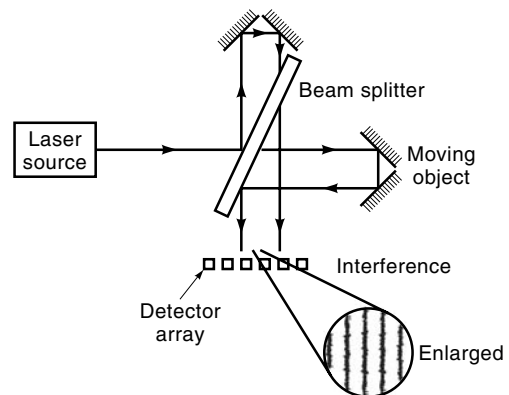


Figure 18. Interferometric displacement sensor. Interference of two light beams causes a series of bright and dark lines. In many cases, the dark and bright points are sensed by photodetectors in relation to the displacement of an object located within the interference area.

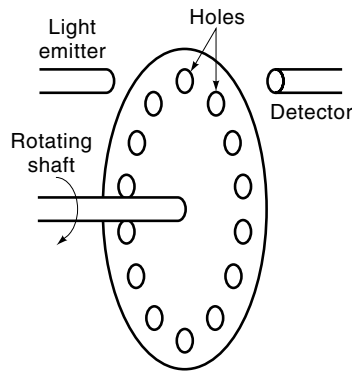


Figure 19. Encoder displacement sensors. These are digital sensors used for both angular and linear measurements. As the wheel rotates, light from the source is transmitted and stopped, giving a digital signal in the photodetectors. The output of the detector is counted and processed to measure the angular displacement of the wheel.

interferometers, laser Doppler and laser feedback interferometers, and fiber interferometers.

Encoders and Digital Displacement Sensors

Digital optical displacement sensors are used for both angular and linear measurements. An example of these sensors is shown in Fig. 19. As the wheel rotates, light from the source is transmitted and blocked, producing digital signal in the photodetectors. The output of the detector is counted and processed to give the angular displacement of the wheel. In another version, it is possible to use regularly positioned reflectors on the surface of the wheel rather than punching holes. In this case, light from the source is alternatively reflected or absorbed. Reflective-type encoders can be arranged to give natural binary-coded or Gray-coded digital signals. Some of these encoders are manufactured to give resolution better than $1\ \mu\text{m}$. In many cases, semiconductor laser light sources, which can give over 1 million samples per revolution, are used.

BIBLIOGRAPHY

- J. P. Bentley, *Principles of Measurement Systems*, 2nd ed., Burnt Mill, UK: Longman Scientific and Technical, 1988.
- E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
- W. Gopel, J. Hesse, and J. N. Zemel, *Sensors—A Comprehensive Survey*, Weinheim, Germany: WCH, 1989.
- J. P. Holman, *Experimental Methods for Engineers*, 5th ed., New York: McGraw-Hill, 1989.
- W. J. Tompkins and J. G. Webster, *Interfacing Sensors to the IBM PC*, Englewood Cliffs, NJ: Prentice-Hall, 1988.

HALIT EREN
Curtin University of Technology

VELOCIMETERS

Velocimeters are instruments used to measure the speed of a moving object or medium. Velocity is a vector quantity; that is, it has both magnitude (commonly referred to as speed) and direction. Many devices classified as velocimeters only measure

the magnitude of the motion or its speed; the direction of the motion is known a priori or must be assumed. Speed is an important quantity in many situations. We may want to know the speed of an automobile, the rate of expansion of the universe, or the flow profile in a blood vessel in the heart. Because velocity and volume flow rates are directly related, some of the techniques described in this article might also be classified as flow meters. One specific category of flow meters is anemometers—devices for measuring the speed of gases, usually air.

A transducer is a device used to convert a physical measurand or variable, such as speed, into a signal in an alternate energy form. For our purposes, this is usually an electrical signal, but it could be a mechanical deflection in a direct reading gauge, for example. The electrical parameter that varies with the measurand can be voltage, current, frequency, or phase. The electrical signal may be directly generated, as in electromagnetic induction, or it may arise from modulation, as in the change of resistance in hot-wire anemometry.

For a linear or translational frame of reference, the following definitions and relationships exist. Displacement, x , is the change in position. Velocity, v , is the rate of change of position. Acceleration, a , is the rate of change of velocity. These can be expressed mathematically by the following differential equations:

$$v(t) = \frac{d}{dt} x(t) \quad \text{and}$$

$$a(t) = \frac{d}{dt} v(t)$$

A similar set of relations exists in an angular or rotational frame of reference.

$$\omega(t) = \frac{d}{dt} \theta(t) \quad \text{and}$$

$$\alpha(t) = \frac{d}{dt} \omega(t)$$

where θ is the angular displacement, ω is the angular velocity, and α is the angular acceleration. From this, we can see that if we can measure displacement or acceleration then we can calculate velocity. However, the signal processing may introduce additional sources of error. In particular, differentiation will tend to enhance any high-frequency noise present in the displacement signal. On the other hand, an offset error or low-frequency drift in the zero level of the acceleration signal will cause errors in the integration process. The use of digital computers has greatly facilitated signal processing and has reduced the sources of errors that can corrupt the velocity signal. However, both high-frequency noise and offset or drift can occur in the analog electronics required for most transducers prior to the signal being digitized.

PHYSICAL PRINCIPLES OF DIRECT VELOCITY MEASUREMENT

As mentioned previously, several physical phenomena can be made to depend directly on the speed of a moving object or fluid. These include electromagnetic induction, the Hall effect, and several variations of the Doppler effect.

Electromagnetic Induction

Electromagnetic induction and the Hall effect (see the next subsection) have the same physical basis—namely, that an electric charge moving through a magnetic field undergoes a lateral force known as the Lorentz force. Michael Faraday first described the phenomenon of electromagnetic induction that has become familiar to electrical engineers through the mathematical relationship referred to as Faraday's law of induction. That is, a voltage will be induced across the windings of a coil that is proportional to the rate of change of flux linkages, λ , in the coil. The polarity of the voltage will be such that if a current can flow it will tend to oppose the change in flux; hence, the minus sign in Eq. (1):

$$e = -\frac{d}{dt}(\lambda) \quad (1)$$

and

$$\lambda = n\phi$$

where n is the number of turns linked with the flux and ϕ is the magnetic flux. If the number of turns linked with the flux is fixed, then a more familiar form of Faraday's law results:

$$e = N \frac{d}{dt} \phi(t) \quad (2)$$

The flux, ϕ , can be represented more usefully as

$$\phi = BA$$

where B is the flux density and A is the cross-sectional area of the coil exposed to the magnetic field. Velocity transducers based on electromagnetic induction have a coil that moves with respect to a magnetic field such that the rate of change of flux linkages is proportional to the velocity. Two possibilities, a moving coil or a moving core, are shown in Fig. 1.

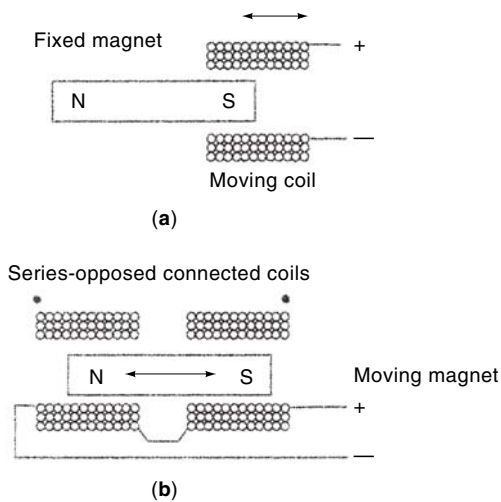


Figure 1. Electromagnetic induction of an electrical potential difference by changing the number of turns linked with the flux. (a) Moving coil and a stationary magnet, (b) two coils connected in series opposition with a moving magnet in the lumen of the coil.

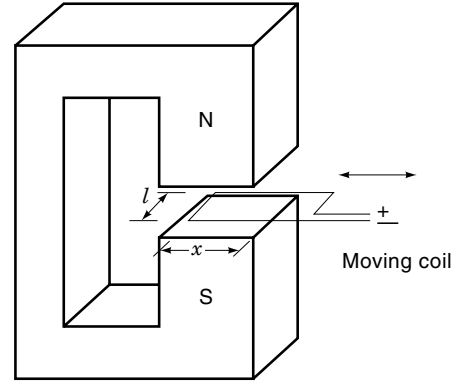


Figure 2. Electromagnetic induction of an electrical potential difference by changing the area of a coil exposed to a magnetic field; l is the width of the coil, and x is the length of the coil in the magnetic field.

If a single rectangular loop under motion is only partially within the magnetic field (Fig. 2), then Eq. (2) can be restated as

$$e = \frac{d}{dt} BA(t)$$

since $N = 1$.

The area exposed to the magnetic field is the lx , where l is fixed and x changes due to being linked to the system being measured. Then

$$\begin{aligned} e &= \frac{d}{dt} Blx(t) \\ \Rightarrow e &= Bl \frac{dx(t)}{dt} \\ \Rightarrow e &= Blv \end{aligned}$$

where the velocity, v , is the time derivative of position, x . Strictly speaking, the voltage induced is due to the vector product of the magnetic field and the velocity of the conductor:

$$e = \mathbf{v} \times \mathbf{Bl}$$

The operator $\times$ indicates the vector or cross product.

Hall Effect

Hall effect transducers are based on the same separation of charge due to the Lorentz force, as in electromagnetic induction. However, in Hall effect transducers, a current is made to flow through a stationary conductor that is exposed to a dc magnetic field (see Fig. 3):

$$\mathbf{F} = q_0 \mathbf{v} \times \mathbf{B}$$

where $\mathbf{F}$ is the force vector, q_0 is the charge, $\mathbf{v}$ is the velocity vector of the charge, and $\mathbf{B}$ is the magnetic flux density vector. The magnitude of the force is therefore

$$|\mathbf{F}| = F = q_0 v B \sin \theta$$

where θ is the angle between the velocity and the magnetic field. Since the electric charges are free to move laterally, the

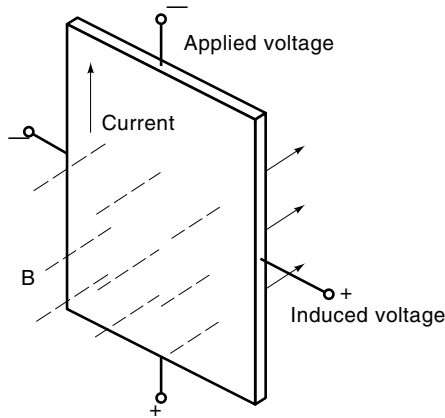


Figure 3. Hall effect principle: An external voltage source is connected across the top and bottom terminals, which causes an electric current to flow. In the presence of a magnetic field (B), the moving charges are displaced laterally due to the Lorentz force. This sets up an induced electrical potential difference (the transverse Hall potential) orthogonal to the current direction and the direction of the magnetic field.

Lorentz force will act to deflect the charges. Then an electrical potential difference will develop perpendicular to the flow path. This potential is referred to as the transverse Hall potential, V_H .

$$V_H \propto iB \sin \theta$$

This principle can also be used in flow meters for ionized solutions in which the flow of ions becomes the current. In this case, the ion concentration and temperature of the solution will also affect the Hall potential as well as the velocity of the fluid.

Doppler Effect

The Doppler effect is most commonly observed when a sound source is in motion relative to a receiver. Thus, the engine of a racecar is perceived to have a higher pitch as it approaches a spectator than after it passes the spectator. This can be extended to any source of propagating waves—ultrasound, microwave, and light:

$$f_0 + f_d = f_0 \frac{c}{c + v_s \cos \theta}$$

where f_0 is the frequency of the transmitter, f_d is the frequency shift, v_s is the velocity of the source with respect to the receiver (positive away from the receiver), θ is the angle between the trajectory of the source and the path of propagation between the source and the receiver, and c is the speed of the propagating wave. This formulation assumes a frame of reference relative to the receiver. This is the basis for the “red shift” observed in the photoemissions from stars, which has allowed astronomers to estimate the rate of expansion of the visible universe. In many applications involving Doppler frequency shift, the transmitter and receiver are both stationary, but an intermediary object, referred to as the target, is moving with respect to both. Then two shifts occur: the target with respect to the source and the receiver with respect to the target. If the transmitter and receiver are located in close proximity to each other, then the Doppler

shift becomes

$$f_0 + f_d = f_0 \frac{(c + v \cos \theta)}{(c - v \cos \theta)}$$

If $v \ll c$, then this becomes

$$f_d \cong \frac{2 f_0 v \cos \theta}{c}$$

Sonic and microwave Doppler velocimeters are usually based on this principle. The object to be measured must have a different impedance than the surrounding medium so that a reflection occurs. For microwave Doppler velocimeters (such as “radar guns”), this is not usually a problem. However, for Doppler flow meters, particles or bubbles must be suspended in the flowing medium in order for a reflected signal to occur. This formulation is applicable if the source is a continuous wave. Pulsed ultrasound flow meters are often referred to as Doppler flow meters. Strictly speaking, however, their principle of operation is time of flight for two successive pulses (1).

Transit Time

The motion of the medium through which the wave propagates can also be used to measure velocity. This does not require any reflective objects in the medium as is required in the Doppler flow method. The length of time taken for a signal to propagate from a transmitter to a receiver depends on the velocity of the medium through which the signal flows. For a more complete development, the reader is referred to Lynnworth (2):

$$T = \frac{D}{c + v_c \cos \theta}$$

where T is the transit time, D is the distance between the source and the receiver, c is the speed of wave propagation, v_c is the average velocity of the medium along the propagation path, and θ is the angle between the flow vector and the propagation path (see Fig. 4). Note that v_c is taken as positive when the transmitted wave is in the downstream direction. Note also that v_c is not the average velocity across the cross section of the flow. The propagation speed also depends on the density and adiabatic compressibility of the medium. These parameters

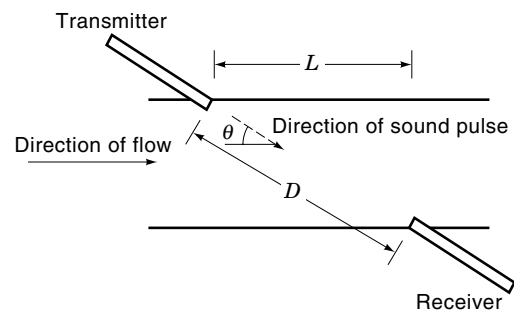


Figure 4. Schematic representation of the measurement of velocity by the transit time of a pulse of sound energy. The fluid is shown flowing in a pipe. The transmitter and receiver are on opposite sides of the pipe at an angle θ to the flow. The transmitter and receiver are separated by a distance L along the pipe and a distance D along the path of the propagated sound.

vary with temperature. Usually the transit time is measured in both the upstream and downstream directions, and the time difference is taken. The time difference varies directly with velocity (3).

$$\Delta T = \frac{2Dv_c \cos \theta}{c^2 - v_c^2 \cos^2 \theta}$$

For $c \gg v_c$,

$$\Delta T \cong \frac{2Dv_c \cos \theta}{c^2} \quad (3)$$

The distance, L , is the separation of the transmitter and receiver along the flow path (Fig. 4). This parameter may be easier to measure than the actual path length. Equation (3) becomes

$$\Delta T \cong \frac{2Lv_c}{c^2}$$

since

$$D = \frac{L}{\cos \theta}$$

For ultrasonic velocity measurement, the velocity of the fluid is typically less than 1% of the speed of sound. Therefore, the speed of sound must be known to within 0.01% for the velocity to be calculated to within 1%. The speed of sound is dependent on the bulk modulus and the temperature of the conducting medium. Therefore, where either of these variables is not carefully controlled, large errors in the calculated velocities will result from measuring the difference in transit times.

However, further calculation can be used to remove the speed of wave propagation. The sum of the upstream and downstream transit times is

$$\sum T \cong \frac{2D}{c}$$

Then dividing the time difference by the square of the sum of the transit times gives

$$\frac{\Delta T}{(\sum T)^2} \cong \frac{v_c \cos \theta}{2D}$$

Now the angle between the flow and wave propagation must be known. Subnanosecond time resolution is required for 1% accuracy for typical measurement situations (2).

An alternative approach is to have the pulse repetition rate inversely proportional to the transit time. Then the difference in frequency between the upstream and downstream pulse trains is

$$\Delta f = \frac{2v_c \cos \theta}{D}$$

or

$$\Delta f = \frac{2v_c}{L}$$

In this formulation, the propagation speed of the wave does not affect the pulse repetition frequency, and only the axial

separation of the transmitter and receivers needs to be known. However, for ultrasonic velocity measurement, Δf is expected to be small, which can result in considerable error in the calculated velocity.

The time shift can also be considered as a phase shift with respect to the source oscillator:

$$\begin{aligned} \Delta \phi &= 2\pi f_o \Delta T \\ \Rightarrow \Delta \phi &= \frac{4\pi f_o D v_c \cos \theta}{c^2} \end{aligned}$$

where $\Delta \phi$ is the phase shift, f_o is the oscillator frequency, and all other variables are as defined previously. Phase shifts can be measured optically through interference effects. With the proper choice of materials in the optical path, changes in ambient conditions have a negligible effect on the speed of light, and laser light sources provide a stable frequency (or wavelength). For acoustic velocimeters, variations in the speed of sound introduce errors in calculated velocities, as was the case for transit time. However, phase difference is easier to measure than small time differences. The disadvantage of phase measurement is the ambiguity that results from the periodicity of the phase shift. This restricts the flow range that can be measured. The maximum carrier frequency, f_o , is calculated by setting the preceding equation to π and solving for frequency given the maximum value of velocity to be measured, the separation distance and angle of the two transducers, and the minimum expected speed of sound over the operating temperature range. A standard crystal frequency below this value may then be selected.

Errors due to changes in the speed of sound with temperature can be eliminated by using two frequencies (4). The phase difference is then obtained between the two upstream signals, and a second phase difference is obtained between the two downstream signals:

$$\Delta \phi_1 = \frac{2\pi f_1 D}{c + v_c \cos \theta}$$

and

$$\Delta \phi_2 = \frac{2\pi f_2 D}{c + v_c \cos \theta}$$

where f_1 and f_2 are the two signal frequencies and all other variables are as previously defined. Then

$$\Delta \phi_u = \frac{2\pi(f_2 - f_1)D}{c - |v_c| \cos \theta} \quad (4)$$

and

$$\Delta \phi_d = \frac{2\pi(f_2 - f_1)D}{c + |v_c| \cos \theta} \quad (5)$$

where u and d indicate the upstream and downstream phase shifts, respectively. One of the frequencies, say f_1 , is still used to calculate the upstream-downstream phase shift.

$$\Delta \phi_1 = \frac{4\pi f_1 D v_c \cos \theta}{(c - |v_c| \cos \theta)(c + |v_c| \cos \theta)} \quad (6)$$

Equations (4) and (5) are rearranged and substituted into Eq. (6), and the resulting equation is rearranged to give

$$v_c \cos \theta = \frac{\pi(f_2 - f_1)^2 D \Delta \phi_1}{f_1 \Delta \phi_u \Delta \phi_d}$$

In this way, the calculation of flow velocity does not depend on the speed of wave propagation. As mentioned previously, stable oscillators are readily available, and phase differences can be measured relatively easily. A second calculation of the flow velocity can also be made using the second frequency upstream-downstream phase shift. Then the average of these two values can be taken, which provides some reduction in the error. The frequency, f_1 , is calculated as for the single-frequency method. The maximum difference between the two frequencies is found by taking the difference of the upstream and downstream phase shifts [Eqs. (4) and (5)] and setting the difference equal to π . In this case, the minimum expected speed of sound is used for the upstream phase shift, and the maximum expected speed of sound is used for the downstream phase shift. The second frequency, f_2 , is then chosen as a standard crystal frequency above that of f_1 but less than the calculated difference.

PHYSICAL PRINCIPLES OF INDIRECT MEASUREMENTS

As mentioned previously, velocity can be determined from successive measurements of position. Many displacement transducers are based on variable resistance, capacitance, inductance, or reluctance. Differentiation of an electrical signal that is modulated by the change of one of these parameters is a straightforward process. Only those displacement techniques developed specifically for use with velocimeters and anemometers will be discussed here. Time is a parameter that can be measured with good accuracy and resolution. Therefore, it is not surprising that measurement of time is the basis for two of these indirect approaches—namely, the time to travel a known distance and the time of flight of two successive pulses of a signal. A related technique is to measure the displacement before and after a known time interval. A third principle is that thermal convection is dependent on mass flow rate, from which velocity can be derived. This is the basis of hot-wire and hot-film anemometry.

Time-Based Techniques

One of the most obvious ways of measuring the average speed of an object is to measure the time it takes to travel a known distance. The transducer for detecting the presence of the object as it passes the start and end points is the major consideration in developing a velocimeter based on this technique. Usually, it is preferable to have a nonintrusive method for detecting the object. Photoelectric timers use a light source (visible or infrared) aligned with a photodetector across the path of the object. These transducers can be readily adapted to a number of applications, including industrial processes and sports activities. Contact switches are another rugged transducer that can be made in a number of form factors and for different environments, such as measuring the speed of automobiles. Because timers and counters can be made with high accuracy and resolution relative to the speed of most large objects, the largest sources of error in such systems are the transduction pro-

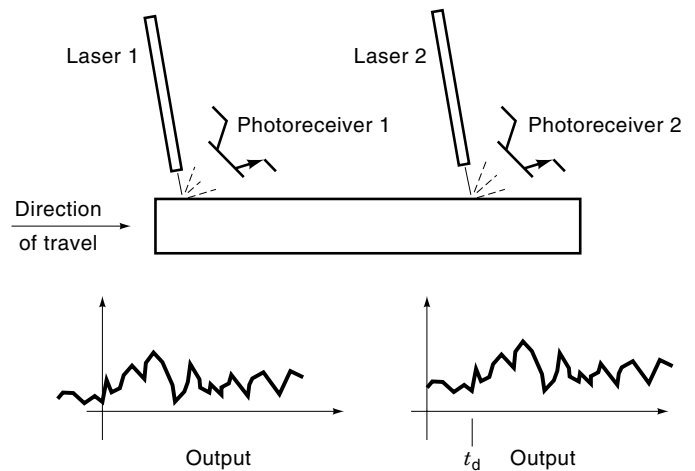


Figure 5. Surface texture correlation. The output of the second photoreceiver is ideally identical to that of the first photoreceiver, but delayed by an amount t_d that is proportional to the separation distance of the two photoreceivers and the speed of the material.

cess and measuring the separation distance of the two position detectors.

Surface texture correlation is a more sophisticated method of determining the time required to travel a known distance. In this technique, two light sources, usually laser beams, illuminate a portion of the surface of a moving object, such as steel or aluminum in a rolling mill or paper, wire, or cable as they are being metered onto rolls. The two beams are a fixed and known distance apart; they are aligned so that the same part of the material passes underneath both beams (Fig. 5). Two photoreceptors provide an analog signal of the light reflected from the two illuminated spots. Differences in surface texture or reflectance will cause variations in these signals. After the signals are digitized and buffered, crosscorrelation is performed on the two signal sequences to find the time lag required to create the largest match between the two signals. Theoretically, if the spots illuminated the same strip of material that passes under both, then the second signal will be a delayed version of the first. In practice, of course, the match is not identical due to differences in photoreceptor response and the presence of noise. High-speed, digital signal processing microprocessors have made this technique feasible for not only monitoring velocity but also, through digital integration, determining the total length of material. Accuracy of better than 0.1% is obtainable. This technique has the advantage of being a noncontact process in which the standoff distance may be more than 0.5 m. Unlike Doppler and transit time techniques, the method is not sensitive to the angle of the beams relative to the material motion, as long as the distance between the two beams on the material surface remains constant.

For time-of-flight measurements, a transmitter emits a pulsed wave that propagates at a known speed. The pulse is reflected off the object of interest (target), and the reflected pulse is received by a detector mounted immediately adjacent to the transmitter. (In the case of ultrasonic time of flight, the transducer can be both the transmitter and the receiver.) The time taken to travel from the transmitter to the detector is proportional to twice the distance from the transmitter to the target. Then successive time measurements allow the calculation of

the change in position, and the velocity can be calculated given the interval between pulses:

$$v \cos \theta = \frac{c \Delta t}{2PRI}$$

where v is the velocity, θ is the angle between the path of the target and the path of the propagating pulse, c is the speed of propagation of the pulse, Δt is the difference in the time of flight of two successive pulses, and PRI is the pulse repetition interval. This is the technique used in optical radar guns, or lidar.

In pulsed Doppler ultrasound, the reflected pulse can be demodulated by being mixed with a phase-shifted version of the original pulse (1). The phase shift of the original pulse controls the sample depth at which the velocity will be measured. After demodulation and filtering, the phase of the resulting low-frequency signal is proportional to the velocity of a single target:

$$\phi = \frac{4\pi K v \cos \theta}{c}$$

where ϕ is the phase and K is a multiple of PRI . Rather than using demodulation, direct sampling of the high-frequency return signal is also possible at a fixed time with respect to the time of pulse emission. If the pulse repetition rate is sufficiently high with respect to the velocity of the target and if a large number of pulse echoes are sampled, then the frequency of the sampled return signal will be proportional to the velocity of the target:

$$f_p = \frac{2f_0 v \cos \theta}{c}$$

where f_p is the frequency of the sampled return signal and f_0 is the frequency of the ultrasonic wave in the pulse. If the pulse is reflected off multiple objects, as is the case for red blood cells, then the signal processing becomes more complicated, and a frequency spectrum results.

Displacement-Based Techniques

A wide variety of displacement transducers have been manufactured based on changes in resistance, capacitance, reluctance, or inductance. Any of these could be used for velocimeters by differentiating the displacement signal with respect to time. Such transducers will not be discussed here. However, shaft encoders are used in an important category of velocimeters—namely, vane anemometers—as well as being used more generally for any shaft angular velocity measurement. A shaft encoder is one of the few transducers that directly outputs a signal in digital form. Shaft (or linear) encoders come in two types: incremental and absolute. Incremental encoders can be further subdivided into single channel and two channel. The latter have two tracks of encoding bits offset by 90° of the mechanical pitch. Two digital signals are thus generated in quadrature. This allows bidirectional information to be obtained, as well as quadrupling the angular resolution of the encoder. Shaft encoders make use of proximity detectors to determine incremental changes in shaft rotation. This can be done by optical transmission through a disc with alternating opaque and transparent windows, with ferromagnetic gears and a variable reluctance or magnetoresistive transducer that “counts” the gear teeth (5),

or with a conductive gear and an eddy-current device. A variation of this technique is to use a reflective target mounted on the shaft of a rotating machine. A photodetector then emits a pulse for each rotation, and the pulses are counted for a fixed duration of time.

Vane anemometers always rotate in the same direction and the absolute angular position is not required; therefore, single-channel incremental shaft encoders are well suited to this application. The angular velocity, ω , can be determined by counting the output pulses for a fixed interval of time. Then the sensitivity of the anemometer is determined by the number of pulses per revolution, N , and the counting interval, T . The range of the anemometer is determined by the size of the counter.

$$\omega = \frac{\text{count}}{NT}$$

Long stroke length (>1 m) linear displacement transducers are difficult to manufacture. Converting the linear displacement to angular displacement is often used for such measurements. This technique is used with cable extension-type displacement transducers. In these transducers, a light, stainless steel line is wound on a reel with a spring return. The shaft of the reel may be connected via a gear train to a potentiometer for an analog output or a two-channel incremental shaft encoder for a digital output (6). These devices are often the transducer of choice for measurements requiring long stroke length (up to 19 m) (7). Some models are available that have a dc tachometer built into the transducer, so that the output is directly proportional to velocity. Typical sensitivities are 30 to 130 mV·m⁻¹·min.

A change in displacement can also be measured from successive images of a moving object or fluid field. This is the basis for high-speed cinematography or video recording and for particle velocimeters. High-speed cinematography requires manual digitization of limb or body centroid positions and joint positions on a frame-by-frame basis. Prior to the development of video recordings, there were no effective means of doing this automatically. Video cameras with high frame rates can replace photographic film to capture the images. This lends itself to digitization of the image and subsequent automatic or human-assisted determination of the position of anatomical landmarks. The advantage of cinematography for biomechanical studies is that the positions of individual limbs as well as the whole body can be determined. Thus, forces acting around joints can be calculated from the velocities. However, it is very difficult to acquire images for motions in three dimensions; thus errors are introduced since the velocity vectors are assumed to always lie in the plane of the two-dimensional image. Some systems [e.g., SelSpot II (Selective Electronics, Partille, Sweden)] with markers placed on the body have been developed in an attempt to overcome this limitation.

Particle image velocimeters take successive images of particles in a fluid field. Then two-dimensional correlation is used to find the displacement of greatest match of particle position between the two images. Any motion out of the plane of the image can introduce substantial error in subsequent velocity calculations (8). Various techniques have been used to acquire the two images. The time interval between exposures is usually less than 100 μs in order to get reliable estimates of high-speed

irregular flow profiles. Mechanical shutters and mechanical film advancing are too slow for such short interexposure timing. Pulsed lasers are frequently used as a stroboscopic light source to create two images at short time intervals. The image may be a double exposure on a single sheet of film or charge coupled device (CCD) camera. In this case, autocorrelation is used to find the position of the second largest peak, which represents the location of the shifted particles. To determine a complicated flow profile over a large image area, the correlation must be done in subregions of the image. The peaks of interest represent the average shift in particle position within that region.

Autocorrelation results in two such peaks, as there is 180° directional ambiguity in the correlation process. Crosscorrelation of two independent images does not have this ambiguity. However, this requires two independent images, which is a technically challenging proposition. Two exposures at different optical wavelengths can be used to create two separable images, or high-speed liquid crystal shutters can be used to create two images on different halves of a single photographic negative (9).

Acceleration-Based Techniques

In a manner analogous to displacement-based velocimeters, it is possible to create a velocity proportional signal by integrating the output of any accelerometer. In the field of vibration measurement on rotating machines, displacement, velocity, or acceleration may be the preferred parameter to measure. The lowest vibration frequency of interest is usually determined by the angular velocity of the rotating machine. For machines rotating at less than 500 rpm, displacement is usually the preferred measurement because of the frequency response of seismic vibration velocimeters (see the section title in "Linear Electromagnetic Induction Velocimeters") is limited to >5 Hz. For vibration frequencies above approximately 1 kHz, acceleration is usually the preferred measurand. In the midrange (5 Hz to 1 kHz), velocity used to be the preferred measurand using seismic velocimeters. With the development of piezoelectric accelerometers, the practical lower limit for acceleration measurement is now as low as for electromagnetic seismic velocimeters. However, because of the historical use of seismic velocimeters in the midfrequency range, some manufacturers [e.g., (10) and (11)] build velocity sensors based on piezoelectric accelerometers that have an integrator built into the transducer package.

Thermal Convection

Hot-wire anemometry is based on convective heat transfer from a (usually) cylindrical heated wire to a flowing medium. That is, the rate of heat transfer will depend on the fluid's mass flow rate, which is, of course, dependent on the velocity of the fluid. The probes can also be made by depositing a thin film of resistive material on a substrate. This allows shapes other than cylinders to be readily manufactured. These are referred to as hot-film probes. Hot-wire probes can be operated in either a constant temperature mode or a constant current mode. The latter is usually used for measuring temperature rather than mass flow rate. The reader is referred to H. H. Bruun (10) for a full development of the behavior of hot-wire probes.

For a finite length of hot wire,

$$d\dot{Q}_e = d\dot{Q}_{fc} + d\dot{Q}_c + d\dot{Q}_r + d\dot{Q}_s$$

where $\dot{Q}_e$ is the electrical heat generation rate, $\dot{Q}_{fc}$ is the forced convective heat transfer rate, $\dot{Q}_c$ is the convective heat transfer rate from the heated wire to the prongs supporting it, $\dot{Q}_r$ is the radiation heat transfer rate, and $\dot{Q}_s$ is the heat storage rate. This assumes that the forced convection rate is much greater than the natural convection rate. For hot-film probes, which have a thin layer of conductive material on a substrate, an additional term must be added for the heat loss to the substrate as this greatly affects its frequency response. The convective loss to the supporting prongs can represent a substantial proportion (~15%) of the total heat loss. For a finite length wire, the preceding relationship can be expressed in terms of the fluid velocity as

$$\frac{I^2 R_w}{R_w - R_a} = A + BU^n$$

where I is the current through the wire, R_w is the resistance of the wire at its operating temperature T_w , R_a is the resistance of the wire at the temperature of the fluid T_a , U is the velocity of the fluid, and A , B , and n are constants. A and B depend on the temperature coefficient of the wire; the density, the viscosity, and the specific heat of the fluid; and the diameter and the length of the wire. A , B , and n must be determined through calibration of a specific probe design. The voltage across the hot-wire element is

$$\frac{E_w^2}{R_w} = (A + BU^n)\alpha_0 R_0 (T_w - T_d) \quad (7)$$

where α_0 is the average temperature coefficient of the wire and R_0 is its resistance at 0°C.

In the constant temperature mode, the probe is connected in a feedback circuit to maintain a constant temperature difference ($T_w - T_a$) as shown in Fig. 6. The voltage across the probe, E , is then used as a measure of the fluid velocity. From Eq. (7), it can be seen that the voltage across the wire, and hence across the probe, depends on both the velocity of the fluid and the

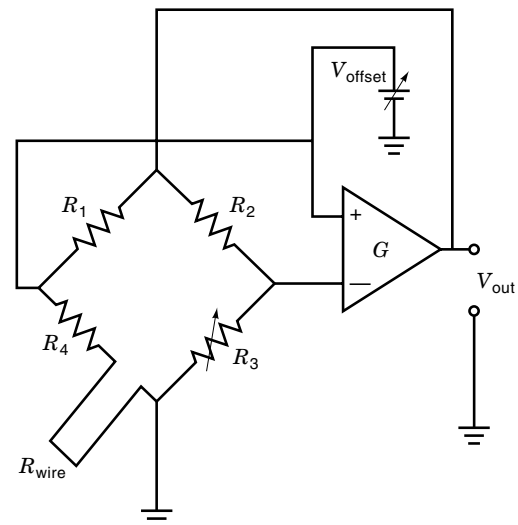


Figure 6. Constant temperature circuit for a hot-wire anemometer. R_3 may be a sensing element for ambient temperature rather than for balancing the bridge. The differential amplifier has gain G .

temperature difference. However, the sensitivity to changes in velocity is greater than the sensitivity to small changes in the ambient temperature (12). Furthermore, the sensitivity to velocity increases with temperature difference and the sensitivity to ambient temperature decreases. Therefore, constant temperature, hot-wire anemometers (HWA) are operated at the largest temperature difference that the wire can sustain. The exponential nature of the voltage-velocity relationship, while being highly nonlinear, means that the relative sensitivity stays approximately constant over a wide measurement range. Hence, HWA systems can be designed to operate at low air velocities as well as supersonic air velocities. The sensitivity will change in the presence of dust, smoke, oil vapor, and other contaminants.

PRACTICAL CONSIDERATIONS

Within the limitations of the transducer, direct measurement of a variable is preferable to indirect measurement. However, the measurement situation may impose constraints that require such indirect measurements. For example, a low-frequency velocity signal may result in unacceptable signal resolution for a direct velocity transducer, and a displacement transducer followed by differentiation may be preferred. However, high-frequency noise introduced on the displacement signal may result in high apparent velocities after differentiation. Care must be taken in filtering the displacement signal to reduce the noise bandwidth without removing signal information. Similarly, at high frequencies, accelerometers may be the transducers of choice followed by integration. Here offset errors can result in a drift in the calculated velocity. Digitization of the signal prior to integration will remove the drift associated with analog integrators. However, offset voltages can occur in the analog amplification of the signal and in the analog to digital conversion. The mechanical resonance of the transducer and the apparatus on which it is mounted must also be known in order to avoid errors whenever step or impulse forces are present or when measuring high-frequency velocities such as occur in vibration.

As in any measurement situation, loading effects must also be taken into consideration. Mechanical loading occurs when the transducer adds inertia to the system, through frictional losses, or when aerodynamic drag is increased. In flow measurement, the transduction process may introduce a pressure drop similar to a voltage drop when measuring electric current. This pressure drop must be kept sufficiently small such that the reduction in velocity is within an acceptable error limit for the measurement. Flow disturbances upstream of the transducer can cause measurement errors, but the transducer may also introduce flow disturbances that affect the process under measurement. This is particularly true when the velocity profile in a boundary layer is being investigated. Particles present in the fluid may cause abrasion of the transducer, or the transducer may cause accumulation of the particles, which then results in an excess pressure loss.

LINEAR ELECTROMAGNETIC INDUCTION VELOCIMETERS

Electromagnetic transducers (and sensors based on variable inductance and variable reluctance) tend to be mechanically robust and relatively immune to contamination from humidity, dirt, and grease. This type of transducer generates electrical energy from the motion of the system under measurement.

Therefore, mechanical loading effects and energy available must be taken into consideration when choosing a suitable transducer.

The coil and magnetic field can be arranged such that linear motion of the coil with respect to the field induces a voltage across the coil. One way of doing this is to have the coil in motion, as illustrated in Fig. 1(a). As long as the coil is not completely over the magnet, the number of flux linkages will change with position, and hence a voltage will be generated. Since the mass of the coil is usually less than the mass of the magnet, this arrangement has a higher mechanical resonant frequency than a moving core transducer. A moving core transducer is typically arranged as shown in Fig. 1(b), with two coils wound and connected in series opposing fashion. In this case, the coil is substantially longer than the magnet. As the flux linkages increase in one coil, they will decrease in the other, but with the series opposing connection, the sensitivity will be doubled.

Two sources of error affect the output of an electromagnetic velocimeter (13). The linear range (or stroke length) of the transducer is that distance where the core is fully within the two coils. As the core moves through the coils at a constant velocity, some variation in the output amplitude will occur. This is referred to as the linearity error. However, the transfer function of the transducer—that is, output voltage versus velocity input—will also have some nonlinearity—the characteristic linearity error. Decalibration can occur due to exposure to high temperature or mechanical shock as these affect the magnetization of the permanent magnet core. The stroke length may be 10 mm to 600 mm, and the sensitivity can be as high as $20 \text{ mV} \cdot \text{mm}^{-1} \cdot \text{s}$ (14).

A specialized version of this type of velocimeter is the seismic vibration transducer (15). In this device, the coils are attached directly to the case of the transducer, and the permanent magnet core is suspended within the coils by two springs. The core and springs form a second-order system with a high-pass frequency response. The components are chosen to provide a low resonant frequency ($<1 \text{ Hz}$) with a critically damped or slightly underdamped frequency response. Above resonance, the core will stay fixed in space with the case (and coils) moving with the vibrating structure to which it is attached. The magnitude of the induced voltage will be constant with respect to vibration frequency for frequencies above approximately three times the resonant frequency of the system. The voltage magnitude will be directly proportional to the velocity of the case. The electrical load impedance of any circuitry connected to the coil must be at least 10 times the coil resistance. A load impedance less than this will affect both the frequency response and the sensitivity of the transducer. The practical upper frequency limit for such devices is approximately 2 kHz (16).

ELECTROMAGNETIC FLOW METERS

As mentioned previously, electromagnetic flow meters are effectively Hall effect devices. In this case, the electric current is in the form of ions in a liquid. Both insertion-type and in-line type meters are commercially available; these are often referred to as magnetic flow meters in commercial literature. For an in-line device, the flow meter is a section of plastic or ceramic-lined pipe that is installed in the fluid path. Two electrodes of an inert metal (e.g., stainless steel or platinum) are

mounted diametrically opposite each other and flush to the inner surface of the plastic liner. Outside the plastic liner, coils are formed to create as uniform a magnetic field as possible mutually orthogonal to the flow direction and the electrodes. If the internal diameter of the flow meter is matched to the inlet and outlet pipe, then the pressure drop across the meter will only be that of a similar length of straight pipe. The coils may be excited with a dc current, a sinusoidal current, or a square-wave ac current. A dc magnetic field has the advantage of not inducing a voltage directly in the lead wires of the electrodes. However, polarization of the electrodes occurs due to creation of a space charge layer as positive ions tend to build up along one wall and negative ions along the other.

An alternating magnetic field can be used to reduce this polarization. A sinusoidal magnetic field is easy to generate, and a continuous reading of voltage and hence flow can be made by rectifying and filtering the resulting signal. However, it is difficult to minimize the voltage that will be directly induced in the lead wires by the electromagnet, since the lead wires and solution will form a single-loop coil. This induced voltage artifact is at 90° electrically to the flow-induced voltage. Thus, a phase-sensitive detector can be used to attenuate the interference signal. The alternating field has the advantage of eliminating the reduced signal amplitude due to polarization. An additional error is introduced if a resistive coating builds up on the electrode surface. This creates a phase shift in the flow-induced signal. The phase-sensitive detector may then reject the flow signal if the phase shift is too large.

Square-wave, ac excitation (sometimes referred to as dc pulsed excitation) can also be used to overcome the polarization error that arises in a steady dc magnetic field. In this case, a large transient voltage will be induced when the magnetic field changes polarity, but once the transient decays, a voltage proportional to the flow will remain. This signal can then be sampled and rectified to determine the flow velocity. Since phase-sensitive detection is not required, nonconductive coating of the electrodes does not introduce an error. Gating of the signal is required to avoid saturating the input amplifier with the transient voltage. The measured signal will be equivalent to that for a dc type, but the alternating polarity substantially reduces the polarization effect.

Electromagnetic flow meters measure the average velocity of the fluid. Therefore, they must be placed in the stream at a point where the flow profile is fully developed. A rule of thumb is that there should be 10 pipe diameters of straight pipe upstream of the flow meter and 5 pipe diameters of straight pipe downstream of the flow meter. The pipe must be full of the fluid being measured. The fluid conductivity must be greater than approximately 0.5 $\mu\text{S}/\text{mm}$. The dynamic range (or turn-down ratio) is typically 100:1 but may be as high as 3000:1 (17). Insertion-type flow meters are capable of accuracy of $\pm 2\%$, and in-line-type flow meters are capable of $\pm 1\%$.

TACHOMETERS

Tachometers are rotary or angular velocity transducers. Although angular velocity may be of interest itself, tachometers also find use in measuring linear velocity. Long stroke length (>0.25 m) velocity transducers are usually impractical or uneconomic, but a wheel or gear can usually be set up to convert the linear motion to rotary motion.

Generating tachometers are based on the same principles as electric power generators, and as such can be either ac or dc generators. A dc tachometer comprises a permanent magnet to establish a constant magnet flux through a rotor that has multiple coils each separately connected to a commutator. The advantage of a dc tachometer is that its output is rectified by the commutation of the rotor windings. The polarity of the output voltage depends on the direction of rotation. The output is linear with respect to angular velocity, although there will be some ripple (6% to 8%) (18) in the signal due to the commutation process. The ripple frequency is directly proportional to the angular velocity and the number of poles on the commutator. The brushes require some maintenance and eventual replacement over time. The sensitivity of dc tachometers is typically 2.5 to 10 V/1000 rev/min (19).

Generating ac tachometers, on the other hand, do not require commutators and brushes, as the permanent magnet is the rotor and the voltage is taken from the stator winding. This output is sinusoidal with a frequency proportional to the angular velocity of the rotor. The amplitude of the generated voltage is also proportional to the angular velocity; however, the direction of rotation cannot be determined from the output signal. By using the frequency of the output rather than the amplitude, errors due to temperature variation and loading effects can be minimized. Squaring of the output also lends itself to digital processing of the signal.

More commonly, ac tachometers are not of the generating type, but they are based on variable mutual induction between two stator windings. The rotor may be a squirrel cage, and the two windings are at 90° to each other. One winding is excited sinusoidally at frequency ω_s . Then the output will be of the form

$$e = k\omega_s\omega_r \sin \omega_s t$$

where ω_r is the angular speed of the rotor. Although the mains distribution frequency is often used for the excitation field, in electrically noisy environments an alternative frequency can be chosen so that a tuned filter can be used to eliminate mains interference. Ac tachometers typically have sensitivities similar to those of dc tachometers. Compensating thermistors with a negative temperature coefficient may be used to reduce temperature effects on winding resistance.

Digital tachometers use proximity detectors that give a pulse output as a sensed object passes the transducer. Tachometers with a pulse output may use any of the following: electromagnetic transduction, Hall effect, eddy current, variable reluctance, optical reflectance, or shaft encoder. An electromagnetic induction tachometer has a coil wound around a permanent magnet placed next to a toothed ferromagnetic rotor (i.e., a steel gear or fan) or key way in a rotor. The reluctance of the magnetic path will be different for a gear tooth passing close to the coil compared to when a gap in the gear is near the coil. This change in reluctance will create a change in magnetic flux and hence will generate a voltage pulse across the coil. A variable reluctance proximity detector is similar in physical layout; however, the coil is energized with a constant current. The voltage across the coil changes as the reluctance of the magnetic path changes. A Hall effect transducer can be placed between a stationary permanent magnet and the rotating gear teeth. Again, the change in reluctance will cause a voltage pulse to occur. However, in this case, the magnitude of the pulse will not be

affected by the angular velocity as occurs for the electromagnetic induction and variable reluctance types. Eddy current proximity transducers can also be used. In this case, a coil in close proximity to an electrically conducting rotor such as a gear or fan is excited at radio frequency (RF) frequencies. The resulting ac magnetic field will create eddy currents in the rotor teeth (or blades) when they pass close to the coil. The induced currents change the self-inductance of the coil and, therefore, will change the voltage across the coil. Reflective digital tachometers usually require a special reflective target be mounted on the rotating component. A light emitting diode (LED) or other light source is aimed at the target, and a photodetector emits a pulse each time the target rotates into its field of view. Angular velocities up to 1,000,000 rpm can be measured with this type of tachometer (20). Optical shaft encoders are also used in pulse-type tachometers. Here a LED/phototransistor pair is aligned with a rotating mask revolving between them. The mask has alternately opaque and transparent sections evenly spaced around its perimeter. Except for the shaft encoder, the digital tachometers do not require mechanical contact with the rotating component and have a minimal mechanical loading effect.

Two signal processing schemes can be used with pulse-type tachometers once the output has been squared. As a direct replacement for analog output tachometers, the output can be fed to a frequency to voltage converter (5). In this case, the pulses must be of constant duration independent of the rotational velocity. A completely digital technique can also be employed. The velocity is determined by counting the output pulses over a fixed duration of time. Very low angular velocities can be measured with better resolution by counting the number of "clock ticks" of a high-frequency clock that occur between successive pulses of the tachometer output.

VELOCITY MEASUREMENT USING ULTRASOUND

Flow Meters

Industrial ultrasonic flow meters may use either the Doppler shift technique or the time of flight. The former has the potential to provide information about the velocity profile across the flow pattern; the latter has the advantage of being usable in "clean" liquids and gases. Doppler shift techniques require reflective targets. Frequently, the targets are bubbles or particles that are either inherently present or are introduced into the fluid. It is also possible to use turbulent swirls within the fluid as the reflectors (21,22). Because of the need for reflective targets, Doppler techniques are not usable in gases.

Ultrasonic flow-meter transducers are mounted in one of three ways: clamped on the outside of the pipe, inserted through an opening in the side of the pipe, or installed in-line via a spool piece. Clamp-on transducers can be mounted on most existing pipe—steel, iron, hard PVC, or glass. Some transducers may also work with concrete lined pipes (23). The alignment of clamp-on transducers must take into account the diffraction of the wave at the two wall-fluid interfaces. The signal will also be substantially reduced in amplitude because of the reflections that occur at these interfaces due to the large acoustic impedance mismatch between the pipe wall and the fluid. For this reason, clamp-on transducers are most frequently used with liquids rather than gases. Clamp-on transducers may be

used where the fluid is corrosive, abrasive, or at high temperature, since the transducer does not come into direct contact with the fluid.

The measurement of flow velocity requires that a laminar flow profile be developed upstream of the flow transducer. This may require as much as 10 pipe diameters of pipe upstream of the transducer and 5 pipe diameters of pipe downstream of the transducer. [However, where vortices are the reflective targets, the flow meter must be located 1 to 3 pipe diameters below a 90° elbow (21).] Doppler ultrasound flow transducers require particles (or vortices) to be present. These may be present by the nature of the fluid, or they may be seeded into the flow. The size and concentration of the particles required will depend on the frequency of the ultrasonic signal. For example, particles must be at least 25 $\mu\text{L/L}$ and 30 μm for 1 MHz ultrasound. Lower frequencies require higher concentrations and larger particles. However, high concentrations of particles can result in reduced signal to noise ratio due to excessive scattering of the signal.

Water in rivers, lakes, and ocean has enough naturally occurring particles to allow Doppler velocity measurements. Several approaches have been developed to measure water currents. Small sample volumes can be used to measure boundary layer velocities, or multiple transmitter/receivers can be used to monitor the current at several depths simultaneously. A third technique is to use range gating of the returned signal from a single transducer to scan the depth of the channel being measured. In this case, a pulse of ultrasonic energy is emitted, and the return signal is monitored for only a short time. The time between pulse emission and gating of the return signal is varied to sample at different depths. The transmitted frequency may be as low as 300 kHz for depths of 100 m or more. Range gating has also been applied to medical ultrasound systems and for flow in pipes and tanks (24).

Flow meters and anemometers based on acoustic or ultrasonic transit time techniques are finding increased use industrially (25). One example is a wind speed system that uses three pairs of transmitter/receivers that are oriented 120° with respect to each other (26). This combination allows the velocity—that is, both magnitude and direction—to be calculated, assuming that the flow is in the plane of the transceiver pairs. Ultrasonic anemometers can measure down to zero velocity, whereas cup anemometers have a threshold speed of about 2 $\text{km}\cdot\text{h}^{-1}$. Where improved accuracy is needed, such as in custody transfer of natural gas, several transmitter/receiver pairs may be used across different chords of the pipe to improve the estimate of the average flow velocity. Four beam paths may be sufficient if the flow profile is not turbulent, or two crossed sets of four paths may be used where some turbulence occurs (27).

Medical Ultrasound Systems

Ultrasound is used extensively in medicine, not only for flow measurement but also flow visualization and imaging more generally. The reader is referred to Jensen (1) for a more complete mathematical development of the techniques described here. A shorter explanation is also available in an article by Routh (28). Medical ultrasound systems use transmitted frequencies in the range of 2 MHz to 10 MHz. Higher frequencies give better spatial resolution but have less penetration due to increased attenuation. In velocity measurement, the most common application is blood flow, but moving structures such as heart valves are also of clinical interest. Flow velocity varies

from $0.01 \text{ m}\cdot\text{s}^{-1}$ to $10 \text{ m}\cdot\text{s}^{-1}$. Amplifiers for the returned ultrasonic signal have a gain function that increases with time after pulse emission. This compensates for the attenuation that occurs as the signal passes through more tissue.

The ultrasound signal is emitted by a one-dimensional array of elements whose drive signals are controlled electronically to steer and focus the signal at the desired position in the tissue or blood vessel. These transducer arrays are either linear or convex. The size and shape of the array depends on the intended site of use: transcutaneous, transesophageal, transvaginal, or transrectal. With the speed of sound in tissue (approximately $1540 \text{ m}\cdot\text{s}^{-1}$), 3500 traces can be scanned each second at a penetration depth of 200 mm. Most ultrasound systems use pulsed rather than continuous wave signals. Pulsed wave systems have the advantage of allowing a two-dimensional image to be obtained with the velocity image overlain on it. The operator can then indicate, by means of a cursor on the display, the approximate direction of flow so that the actual velocity can be calculated. The display may be color enhanced with flow toward the transducer being red and flow away from the transducer blue. Pulsed wave signals are usually referred to as pulsed Doppler; however, Jensen (1) provides a strong mathematical derivation to show that the velocity information is actually contained in the time shift of successive pulses rather than the Doppler frequency shift. The classical approach to signal processing has been to use coherent demodulation of the returned echo with the transmitter oscillator followed by low-pass filtering. The mean frequency of the demodulated signal is determined by autocorrelation. An alternate approach is to sample directly the return signal and to crosscorrelate successive pulses. This method has become feasible with development of analog to digital converters capable of sampling at the frequencies used for ultrasound imaging.

A more recent development in medical flow imaging is power or energy flow imaging, sometimes referred to as power Doppler imaging. In this case, an estimate is made of the power in the return signal. The power is not dependent on the angle of incidence of the ultrasound beam and the moving target. However, the result is only an image of where flow is occurring without any ability to estimate the velocity of the flow. Clinically, the advantage to power imaging is that smaller arteries can be resolved. This allows a better understanding of tissue perfusion. Another improvement currently undergoing investigation is three-dimensional ultrasonic velocimetry [for example, see (29)]. Two approaches for this have been identified: to use multiple receivers whose beam axes are not co-planar and to use crosscorrelation of two successive images (similar to particle image velocimetry below).

MICROWAVE VELOCIMETRY

Perhaps the most notorious velocimeter is the microwave Doppler velocimeter or radar gun—used primarily for traffic speed enforcement. The target must be electrically conductive to behave as a reflector. A continuous wave microwave signal, at about 10 GHz, is generated by a Gunn diode and varactor in a resonant cavity and is emitted from a horn antenna. The return signal is received by the same antenna, where it is led by means of a waveguide to a Schottky mixer diode, where it is mixed with a portion of the oscillator output (30). The output of the mixer is the low-frequency, Doppler shift signal. The

frequency, and hence the velocity of the target, can be obtained from this signal, for example, by using a zero crossing detector (31). More complex Doppler radar systems are used in aircraft navigation systems to determine aircraft velocity (32). These are particularly useful in helicopter navigation and control because the helicopter may have long periods of low velocity when hovering. During these periods, drift in an inertial guidance system based on gyroscopes may affect the accuracy of position calculations.

The global positioning satellite (GPS) system and its Russian counterpart (GLONASS) are being used increasingly for navigation purposes. Collectively these are referred to as the global navigation satellite system (GNSS). The standard civilian GPS coded information channel (C/A on the L1 carrier frequency) has a resolution of less than 100 m 95% of the time (33). For short time intervals, this level of repeatability leads to highly erroneous velocity estimates based on successive positions and the elapsed time. However, it is possible to use the carrier signal rather than the coded information to improve both position and velocity calculations. Carrier phase tracking (33,34) takes advantage of the 1000-fold improvement in spatial resolution by using the wavelength of the carrier (approximately 190 mm) to obtain position information. Surveying instruments are available (34,35) that provide 10 mm resolution. They do not suffer from the dithering of the coded information channel referred to as selective availability. The most recent instruments do not require a base station at a known location to provide this level of accuracy. Thus, successive estimates of position can provide velocity information. Since GNS receivers must have highly stable clocks, time information is inherently readily available.

A third use of electromagnetic waves in velocimetry is measuring high-latitude ionospheric convection, which are currents of ionized particles that flow due to the interaction of the solar wind with the earth's magnetic field (36). The velocity of the ionized particles flowing in these field aligned currents can be measured using the Doppler shift that arises when an electromagnetic beam is reflected off the particles. The radar frequencies are in the range of 8 MHz to 20 MHz with corresponding wavelengths of approximately 10 m. Therefore, the reflective target cannot be individual charged particles, but irregularities in the electron density that have a size of approximately one-half of the wavelength. Particle drift velocity is on the order of $300 \text{ m}\cdot\text{s}^{-1}$ to $400 \text{ m}\cdot\text{s}^{-1}$ and may exceed $1000 \text{ m}\cdot\text{s}^{-1}$. With two radars separated by several thousand kilometers, it is possible to calculate the magnitude and direction of the currents in the plane of the radar beam.

OPTICAL VELOCIMETRY

Particle Image Velocimetry

Particle image velocimetry (PIV) is used for imaging complex flow profiles (8). Because of the cost and complexity of the apparatus, this technique is used primarily in research applications. Areas of application include combustion flows, unsteady aerodynamic flow around rotors, turbulent water flow, flows in internal combustion engines, and natural convection.

A laser beam is passed through a cylindrical lens to create a sheet of light that illuminates the fluid field. The thickness of the laser sheet can be controlled by interposing a spherical

lens after the cylindrical lens. The flow must be in the plane of the illuminating sheet, and the fluid must be transparent to the wavelength of the illuminating source. The fluid must have naturally occurring particles, or it must be seeded with particles appropriate to the fluid and the desired frequency response. Particles can range from $0.2\ \mu\text{m}$ to $40\ \mu\text{m}$, depending on the particle dynamics of the flow. Particular care must be taken in determining the particle size required to image swirling flows. Light from the particles is imaged onto a camera, and two successive images are taken. If irregularities in the flow pattern are smaller than the area being imaged, then subsequent processing is done on overlapping subsets of the total image. The light source is usually a pulsed laser with short duration (10 ns) pulses. In a dual laser system, the time between pulses may be adjusted from 200 ns to 0.5 s (37). The time between pulses must be known, of course, for the speed to be calculated. If both pulses are captured on the same image, then there will be an ambiguity as to the direction of flow. The image may be captured on a photographic transparency or a CCD video camera. The former has the advantage of high resolution; the latter has the advantage of representing the image electronically so that it can be readily converted into a digital representation.

The displacement information contained in the double exposure image can be extracted in a number of ways. If the image is on a transparency, this can be illuminated by a laser beam. The particle image pairs will cause Young's fringes to be formed from which the image displacement can be inferred. A transparency can be made of the Young's fringes. When this is, in turn, illuminated by a laser source, the autocorrelation function is created. The two first-order side peaks represent the displacement distance between the two images. Alternatively, the Young's fringes can be digitized via a CCD camera, and a digital autocorrelation can be calculated. If a CCD camera is to be used, then the original two PIV images can be captured directly on a single frame, and digital autocorrelation can be performed. If the velocity is sufficiently slow, successive images may be on the even and odd fields of a frame. However, this reduces the spatial resolution. Where separate images have been obtained, crosscorrelation or maximum quadrature difference can be used to determine the displacement (38).

A variety of techniques has been used to separate the two images to avoid the directional ambiguity. A rotating mirror can be used before the camera to move the second image by more than the largest expected displacement. This value can then be subtracted from that calculated by one of the aforementioned methods. This is particularly useful where turbulence prevents a priori knowledge of the direction of flow. In a dual laser system, the two beams can be given different optical polarization, and a birefringent crystal, such as calcite, can be used to shift one beam. Wirth and Baritaud (9) describe a method of using a ferroelectric liquid crystal retarder on the imaging path (rather than the illuminating path). Their technique resulted in one image with only the first laser pulse and a second image with the usual double image. A CCD camera that has storage cells associated with each pixel element of the camera is capable of acquiring two full frames of information at short time intervals (39). In this case, the image information is transferred in parallel from the light-sensitive elements to the storage cells prior to the second laser pulse. This technology has allowed on-line calculation of velocity vectors for flow velocities up to supersonic speeds. Such high-speed image calculations are usually performed by dedicated computers configured to maximize the

throughput of parallel calculations. Grant (8) provides a review of developments in the field, which are under development at the time of this writing.

Laser Doppler Anemometry

Laser Doppler anemometry (LDA) is also known as laser Doppler velocimetry, laser anemometry, or optical anemometry because its main application is in complex aerodynamic flow fields. The reader is referred to Durst, Melling, and Whitelaw (40) for a more complete explanation of the physics of laser Doppler anemometry. As with particle image velocimetry, no flow disturbance occurs; particles must be present; and there must be optical access to the flow. The Doppler frequency shift is detected as a phase shift by creating an interference pattern between the Doppler-shifted beam and a reference beam. The interference fringes will have a beat frequency because of the Doppler shift. The beat frequency is proportional to the flow velocity perpendicular to the source of illumination, unlike Doppler ultrasound, which measures the component parallel to the insonation beam. The beat frequency can then be detected with a photomultiplier tube or other photodetector. The technique has the advantages of high accuracy, wide dynamic range (a few microns per second to Mach 8) (37), and fast response time.

Three modes of operation are possible: dual beam (or fringe or differential Doppler), reference beam, and two scattered beam. The dual beam technique uses a beam splitter to create two beams from a single laser. By means of appropriate optical elements, including optical fibers, the two beams are focused on the same volume region of the flow but from different angles. A light-collecting lens and mask are used to image the interference fringes from a single particle onto a photodetector. The optical axis of the collecting side is symmetrical with respect to the two illuminating beams. This approach has a large solid angle but requires relatively few particles in the flow. The reference beam technique has a smaller solid angle and can be used in particle-rich flows. Both the reference beam and the scattering beam illuminate the flow region, as before. However, in this case, the optical axis of the collecting system is coincident with the axis of the reference beam. The two-scattered-beam configuration has a single beam illuminating the flow volume, but there are two collecting axes. The scattered light along these axes is then combined to create the desired interference fringes. The collected light can be backscattered rather than forward scattered, which means that optical access is required from only one side of the flow field. This last approach has the further advantage of being readily adapted to measure two- or three-dimensional flow components by having two or three mutually orthogonal collecting systems.

As with other Doppler shift techniques, LDA will have directional ambiguity. Where this is problematic, such as with turbulent flow, a rotating diffraction grating on one of the two illuminating beams can be used to create a frequency preshift. Alternately, the two beams can have their wavelengths shifted in opposite direction by being passed through a Bragg cell (41).

This technique has also been applied to the study of perfusion of soft tissues in medicine (42). In this application, a laser beam is scanned across the surface of the tissue by means of a rotating mirror. The reflected light is then processed in a manner similar to the aforementioned techniques. This is one of the

few methods that can image the low blood flow in peripheral vasculature. However, only relative flow measurements can be made, not absolute measurements.

Time-of-Flight Optical Velocimeters

If the radar gun is the most notorious velocimeter, then the lidar (or ladar) is quickly gaining a similar status for the same reason. This is the optical replacement for the radar gun in traffic speed enforcement. Lidar (light or laser distance and ranging) is based on the time of flight of a short pulse of laser light emitted and received by a single unit. Since the speed of light is constant with respect to temperature and atmospheric conditions, the time taken will be directly proportional to the distance. In fact, commercial units (43,44) include the ability to measure range as well as speed. Successive measurements can then be used to calculate the speed of the target. Because of the short duration of the laser pulses, up to 60 pulses may be used in a single speed measurement, which takes 0.3 s to perform. The infrared laser diodes have a beam divergence of 3 mrad (that is, 3 m in 1000 m) so individual vehicles can be targeted as reflectors. A geometric correction must be made if the target is not moving in line with the laser beam path.

Optical Gyroscopes

Gyroscopes measure rotational velocity usually for navigation purposes or stabilization of a platform. The propagation velocity of electromagnetic waves is not affected by motion of the medium through which they are passing, as was the case for acoustic waves. That is, electromagnetic waves, and photons particularly, can be thought of as moving in an inertial frame of reference. However, suppose that three mirrors are mounted on a platform that may be rotated as shown in Fig. 7. A photon traversing the path in the clockwise direction will experience a longer path length than one traversing in the opposite direction if the whole apparatus rotates clockwise. This is known as the Sagnac effect (45). Two types of optical gyroscopes have been

developed based on the Sagnac effect: ring laser gyroscopes and fiber optic gyroscopes.

In a ring laser, the rotating ring is filled with the lasing medium. The output is two independent beams—one in each direction. The two beams will be at different frequencies (that is, wavelengths) since there must be an integer number of wavelengths in the resonant path (46). The two path lengths will be different by the amount

$$\Delta p = \frac{2\Omega A}{c}$$

where p is the path length, Δp is the difference in path length, Ω is the rotational velocity, A is the projected area of the ring, and c is the speed of light. This leads to a difference in time it takes for light to travel around the ring:

$$\Delta t = \frac{2\Delta p}{c}$$

The wavelength shift can be measured through interference fringes created by mixing the two output beams. Ring lasers are used in research applications such as measuring the rotation of the earth.

Fiber optic gyroscopes (FOG) have counterpropagating light that is injected into opposite ends of a coil of single-mode optical fiber. The light source may be a laser or a superluminescent light emitting diode. The latter has the advantage of reducing one source of phase error—namely, the Kerr effect (47). If the optical fiber undergoes rotation in the plane of the coil, then the two beams will have had a different time of flight, as mentioned previously. This leads to a phase shift that can be detected through interference fringes. Multiple turns of the coil increase the phase shift and hence the sensitivity of the gyroscope:

$$\Delta\phi = \frac{8\pi\Omega NA}{\lambda c}$$

where $\Delta\phi$ is the phase shift, Ω is the angular velocity, N is the number of turns, A is the area, λ is the wavelength, and c is the speed of light. Note that in both the FOG and the ring laser, it is the area of the closed path that is important, not the path length. A detector placed at the output of the interfering beams has an amplitude that is dependent on the phase shift (48):

$$P_D = \frac{P_{in}(1 + \cos \Delta\phi)}{2 \text{ Loss}}$$

The maximum sensitivity will occur when the beams are 90° optically with respect to each other. A phase modulator is, therefore, usually placed in the path of one of the incoming beams. The phase modulator has the additional advantage of causing the interferometer to oscillate $\pm 90^\circ$. Thus, the otherwise dc signal becomes an ac signal at the frequency of the phase modulator. A tuned filter is then used on the output of the detector to reduce 1/f noise.

FOGs are manufactured to cover a wide range of performance requirements. Inertial navigation quality FOGs have maximum sensing rates of up to 400° per second and stability of 0.0001° per hour. At the low end of the performance scale, the maximum rate is 100° per second and the bias stability is

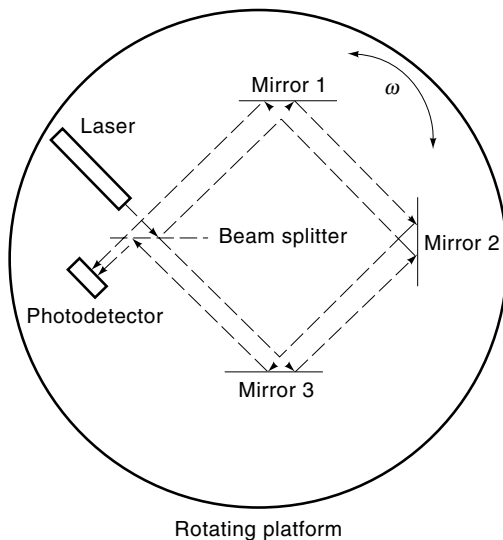


Figure 7. A Sagnac interferometer consists of a light source, a closed optical path, and a means of detecting the fringes that are generated as the platform rotates.

100° per hour. A typical FOG might have 1000 m of optical fiber on an 84 mm diameter spool (49).

Hot-Wire Anemometry

Hot-wire anemometry (HWA) is based on the physical principle that heat dissipation from a heated object is dependent on the mass flow of the medium in which the object is immersed. The heat dissipation is dependent not only on velocity but also on temperature, pressure, specific heat, and density of the medium. In most cases, HWA is used for air velocity measurements, in which case the relative humidity of the air affects its density and specific heat. For environmental air speed monitoring, such as in building ventilation systems, the atmospheric pressure and relative humidity must be known, or assumed, to calculate velocity from the mass flow rate. Since the technique is based on a differential temperature measurement, many instruments built for environmental monitoring also include the capability of measuring the ambient air temperature (20). Measurement of atmospheric pressure may also be included (50). Air speed from 0 to 60 m·s⁻¹ can be measured by such instruments. In instruments designed for mass flow measurements rather than velocity measurements, multiple sensors may be included on a single mount to obtain a better estimate of the flow profile (50).

HWA is also used to measure complex flow fields involving high air speeds and/or turbulence. Although HWA inherently alters the flow field, it is often used instead of laser Doppler anemometry or particle image velocimetry because of its lower cost and no requirement for optical access to the flow field or an optically clear fluid. In turbulent flow, the frequency response is an important figure of merit as well as the maximum velocity. Because of their small size, hot-film probes can be designed with a frequency response up to 300 kHz. Cone- or wedge-shaped hot-film probes are preferred over cylindrical probes due to greater strength (50). The film material is usually platinum and the substrate material is quartz. If the probes are coated, they can be used in conductive liquids, although this will increase the response time of the probe.

BIBLIOGRAPHY

1. J. A. Jensen, *Estimation of Blood Velocities Using Ultrasound: A Signal Processing Approach*, Cambridge, England: Cambridge Univ. Press, 1996.
2. L. C. Lynnworth, Ultrasonic flowmeters, in W. P. Mason and R. N. Thurston (eds.), *Physical Acoustics: Principles and Methods*, Vol. XIV, New York: Academic Press, 1979.
3. J. Fraden, *Handbook of Modern Sensors: Physics, Design, and Applications*, 2nd ed., Woodbury, NY: American Institute of Physics, 1997.
4. F. Gao, *An experimental feasibility investigation of a low cost piezoelectric film ultrasound transducer*, M.Sc. thesis, University of Saskatchewan, 1996.
5. Dynapar Corporation, *Fundamentals of digital measurements, Rotopulsers*, Gurnee, IL, 1989.
6. UniMeasure, Inc., *Linear Position and Velocity Transducers* [Online], Corvallis, OR, 1998. Available www: <http://www.unimeasure.com/>
7. Celesco Transducer Products, Inc., Canoga Park, CA [Online], 1998. Available www: <http://www.celesco.com/>
8. I. Grant, Particle image velocimetry: A review, *Proc. Inst. Mech. Eng. (Part C)*, **211** (C1): 55–76, 1997.
9. M. Wirth and T. A. Baritaud, A cross correlation PIV technique using electro-optical image separation, *Experiments in Fluids*, **21** (6): 410–416, 1996.
10. H. H. Bruun, *Hot-Wire Anemometry: Principles and Signal Analysis*, Oxford, England: Oxford University Press, 1995.
11. Wilcoxon Research, Inc., *PVT® PiezoVelocity Transducers*, Gaithersburg, MD [Online], 1998. Available www: <http://www.wilcoxon.com/velocity.htm>
12. Bently Nevada, Corp., *Case Mounted Transducer Systems*, Minden, NV [Online], 1998. Available www: <http://www.bently.com/pds/pdslist.htm#case>
13. Schaevitz Engineering Ltd., *Linear Velocity Transducers*, Technical Bulletin 1101A, Pennsauken, NJ.
14. TransTek, Inc., *Series 100 linear velocity transducers*, Ellington, CT [Online], 1998. Available www: <http://www.transtek-inc.com/lvt.htm>
15. C. L. Nachtigal, *Instrumentation and Control: Fundamentals and Applications*, New York: Wiley, 1990.
16. F. Oliver, *Practical Instrumentation Transducers*, New York: Hayden Book Co., 1971.
17. Danfoss A/S, *MAGFLO® electromagnetic flowmeters*, Nordborg Denmark [Online], 1998. Available www: <http://www.danfoss.com/Business/Instrumentation/electro/>
18. API Portescap Inc., *escap® tachogenerators and motor-tacho units*, La Chaux-de-Fonds, Switzerland [Online], 1998. Available www: <http://www.portescap.com/base/Group-prod.htm>, 1998.
19. R. Pallas-Areny and J. G. Webster, *Sensors and Signal Conditioning*, New York: Wiley, 1991.
20. Monarch Instruments, *Tachometers*, Amherst, NH [Online], 1998. Available www: <http://www.monarchinstrument.com/tachs.htm>
21. Omega Engineering Inc., *The Flow and Level Handbook*, Stamford, CT, 1995.
22. Alphasonics, Inc., *Ultrasonic flowmeters*, Austin, TX [Online], 1998. Available www: <http://www.isadirectory.org/online/directory/data-center/p273000/m076000p.cfm>
23. Panametrics, Inc., *New fixed-installation ultrasonic liquid flowmeter offers two measurement modes*, Waltham, MA [Online], 1998. Available www: <http://www.panametrics.com/pci/flowmeters/fixflow.shtml>
24. Dantec Measurement Technology, Inc., *Ultrasonic velocity profile monitor*, Skovlunde, Denmark [Online], 1998. Available www: <http://www.dantecmt.com/scientific/uvp/uvpmain.htm>
25. Automation Research Corp., *Ultrasonic flowmeter outlook study*, Dedham, MA [Online], 1998. Available www: <http://www.arcweb.com/public/studies/docs/ultrasonic.htm>
26. Handar, *425 Ultrasonic wind sensor*, Sunnyvale, CA [Online], 1998. Available www: <http://www.tsi.com/handar/color/html/425spec.htm>
27. L. C. Lynnworth, *Ultrasonics Measurements for Process Control*, Boston: Academic Press, 1989.
28. H. F. Routh, Doppler ultrasound: The ability to measure and image blood flow, *IEEE Eng. Med. Biol. Mag.*, **15** (6): 31–40, 1996.
29. Y. Ogura, K. Katakura, and M. Okujima, A method of ultrasonic 3-D computed velocimetry, *IEEE Trans. Biomed. Eng.*, **44**: 823–830, 1997.
30. Alpha Industries, Inc., *An introduction to the theory, operation and application of microwave motion sensing modules*, Woburn, MA, 1987.
31. R. E. Gander et al., Sprint Start Instrumentation, *IEEE Trans. Biomed. Eng.*, **43**: 637–643, 1994.
32. Litton Guidance and Control Systems Division, *Doppler radar navigation systems*, Woodland Hills, CA [Online], 1998. Available www: <http://www.littongcs.com/gcs/products/4doppler/overview.html>

33. P. H. Dana, *The Geographer's Craft Project*, Department of Geography, University of Texas at Austin [Online], 1998. Available www: <http://www.utexas.edu/depts/grg/gcraft/notes/gps/gps.html>
34. Trimble Navigation Limited, *All about GPS*, Sunnyvale, CA [Online], 1998. Available www: <http://www.trimble.com/gps/>
35. Ashtech, *About global positioning: The basics of GPS & GPS+GLONASS*, Sunnyvale, CA [Online], 1998. Available www: <http://www.ashtech.com/>
36. R. A. Greenwald et al., DARN/SuperDARN: A global view of the dynamics of high-latitude convection, *Space Science Rev.*, **71**: 761–796, 1995.
37. TSI, Inc., *Particle Image Velocimetry (PIV)*, 5, St. Paul, Minnesota, [Online], 1998. Available www: <http://www.tsi.com/fluid/piv/cat/contents.htm>
38. L. C. Gui and W. Merzkirch, A method of tracking ensembles of particle images, *Experiments in Fluids*, **21**: 465–468, 1996.
39. Dantec Measurement Technology, Inc., *Particle image velocimetry (PIV)*, Skovlunde, Denmark [Online], 1998. Available www: <http://www.dantecmt.com/scientific/piv/pivmain.htm>
40. F. Durst, A. Melling, and J. H. Whitelaw, *Principles and Practice of Laser-Doppler Anemometry*, 2nd ed., London: Academic, 1981.
41. Dantec Measurement Technology, Inc., *Laser-Doppler anemometry (LDA)*, Skovlunde, Denmark [Online], 1998. Available www: <http://www.dantecmt.com/scientific/lda/ldamain.htm>
42. K. Forrester, M. Doschak, and R. Bray, *In vivo* comparison of scanning technique and wavelength in laser Doppler perfusion imaging: Measurement in knee ligaments of adult rabbits, *Med. Biol. Eng. Comput.*, **35** (6): 581–586, 1997.
43. Kustom Signals Inc., ProLaser, Lenexa, KS.
44. Laser Technology Inc., *LTI 20-20 Marksman*, Englewood, CA [Online], 1998. Available www: <http://www.lasertech.com/laserproducts/marksman.html>
45. R. Anderson, H. R. Bilger, and G. E. Stedman, “Sagnac” effect: A century of Earth-rotated interferometers, *Amer. J. Phys.*, **62** (11): 975–985, 1994.
46. W. M. Macek and T. T. M. Davis, Jr., Rotation rate sensing with traveling-wave ring lasers, *Appl. Phys. Lett.*, **2** (3): 67–68, 1963, as reprinted in P. Hariharan (ed.), *Selected Papers on Interferometry*, SPIE Milestone Series, Vol. MS 28, SPIE Optical Engineering Press, 1991.
47. C. R. Pollock, *Fundamentals of Optoelectronics*, Chicago: Richard D. Irwin Inc., 1995.
48. Litton Guidance and Control Systems Division, *An overview of fiber optic technology* [Online], 1998. Available www: <http://www.littongcs.com/gcs/products/2guidance/fiber/01fibopttech/overview.html>
49. Fibersense Technology Corp., *High accuracy FOG, 1K/80*, Canton, MA [Online], 1998. Available www: <http://www.fibersense.com/>
50. Kurz Instruments, Inc., Monterey, CA [Online], 1998. Available www: <http://www.kurz-instruments.com/index.htm>

ROBERT E. GANDER
University of Saskatchewan

ACCELERATION MEASUREMENT

When we use the term *acceleration*, we usually implicitly mean *instantaneous acceleration*. However, when acceleration is measured, we actually obtain an estimate of *average acceleration* within some time window. The average acceleration a_a during a certain time interval Δt is equal to the change in velocity Δv per unit time during that interval. The instantaneous acceleration

(or just acceleration) a is the limiting value of the average acceleration, when the observation interval approaches zero (1). This can be expressed mathematically as

$$a = \lim_{\Delta t \rightarrow 0} a_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta v}{\Delta t} \quad (1)$$

Acceleration measurements are frequently needed for automatic control, protective supervision, and condition monitoring in applications like vertical and horizontal transportation, robotics, aerospace navigation, and technical diagnostics (2,3). Acceleration belongs to the important chain of kinematic quantities: position, velocity, acceleration, and jerk. They all have a linear connection to their neighbor quantities; for example, acceleration is obtained by differentiating the corresponding velocity or integrating the jerk. Therefore, in principle, all the kinematic quantities could be derived from a single quantity. In practice, however, only integration is widely used to process these kinematic quantities, since integration always provides advantageous noise attenuation. Differentiation, on the other hand, is noise amplifying by its nature. Therefore it is seldom utilized in practical applications when the input signal contains noise or other disturbances.

This possible noise results from various measuring and processing errors, as well as external disturbances like electromagnetic interferences (EMI). The measuring errors are caused by sensor nonidealities and by cumulative effects of the entire measuring instrumentation (4). Quantization, finite-precision computations, sampling, and approximative algorithms are typical sources of processing errors. External disturbances may enter a measuring system due to inadequate grounding, shielding, isolation, or poor cabling.

REFERENCE AXES AND DIFFERENT TYPES OF MOTION

Measurement of acceleration, as well as all kinematic quantities, are made with respect to some system of reference axes. The basic frame of reference used in mechanics is known as the *primary inertial system* (or astronomical frame of reference). It consists of an imaginary set of rectangular axes that neither translate nor rotate in space. Measurements made with respect to this primary inertial system are said to be *absolute*. In most earth-bound engineering applications, measurements made relative to the earth can also be considered absolute (at least the introduced error is negligible).

There are many engineering problems for which the analysis of motion is simplified by measuring kinematic quantities with respect to a moving coordinate system. These measurements, when combined with the observed motion of the moving coordinate system, permit the determination of the absolute motion. This approach is known as a *relative motion analysis* (2).

Direct and Indirect Measuring Techniques

There exist two classes of acceleration measurement techniques: *direct* measurements by special sensors or accelerometers and *indirect* measurements where velocity is differentiated using some differentiator circuitry or a computational algorithm (2). The applicability of these techniques depends on whether one is measuring rectilinear, angular, or curvilinear motion or equilibrium-centered vibration, which is a special class of acceleration.

Direct measuring using an accelerometer is usually preferred when the motion is either rectilinear or curvilinear. Techniques used to measure linear unidirectional acceleration include spring mass, stretched wire, pendulum, piezoelectric, strain gauge, and force balance (3). Vibration measurements are based almost solely on direct techniques because there usually exists no actual velocity signal that can be differentiated successfully. Angular acceleration, on the other hand, is usually measured indirectly because the rotation range of existing angular accelerometers is severely limited due to their mechanical structures (5). Angular velocity, a necessary basis for indirect measurement of angular acceleration, can easily be measured by ac or dc tachogenerators or pulse encoders with some post-processing electronics (6–8). Besides, rectilinear motion is often first converted mechanically into angular motion, and the corresponding angular acceleration is then measured using some indirect technique.

Differentiators needed in indirect measuring have two important requirements: They must provide simultaneously *adequate noise attenuation* and *appropriate delay characteristics*. The delay properties are particularly critical in real-time control and protective supervision applications. Even a small delay in an acceleration signal that is used for feedback control can reduce the control performance drastically. Besides, a considerably delayed acceleration curve is not a sufficient basis for time-critical actions.

RECTILINEAR, ANGULAR, AND CURVILINEAR ACCELERATION

Before we go into specific measuring techniques, we need to formulate the different types of acceleration in the corresponding coordinate systems. The applicable measuring techniques are closely connected to these formulations.

Rectilinear Acceleration

Rectilinear motion is illustrated in Fig. 1. This kind of motion occurs, for example, in vertical and horizontal transportation as well as in various positioning servo applications. The average velocity v_a during a time interval Δt is the displacement Δs divided by the time interval. The instantaneous velocity (or just velocity) v can be defined in a similar way as the instantaneous acceleration of Eq. (1):

$$v = \lim_{\Delta t \rightarrow 0} v_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt} \quad (2)$$

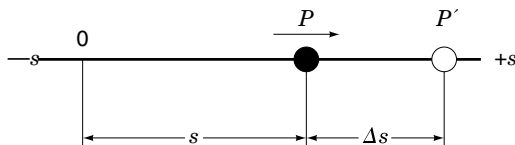


Figure 1. Rectilinear motion of a particle. Its original position is $s(P)$, and the new position is $s + \Delta s(P')$. *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Rectilinear acceleration can now be written as the time derivative of velocity or the double time derivative of the corresponding distance

$$a = \lim_{\Delta t \rightarrow 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = \frac{d(ds/dt)}{dt} = \frac{d^2s}{dt^2} \quad (3)$$

Equation (3) is used frequently in indirect acceleration measurement. However, the double differentiation is seldom implemented in practice due to the noise amplification problem discussed earlier. Instead, the time derivative of measured velocity is somehow approximated. By combining Eqs. (2) and (3), we obtain another important formula to be potentially used in indirect acceleration measurement:

$$a = v \frac{dv}{ds} \quad (4)$$

Depending on the applied velocity measuring technique, it is sometimes more natural to compute the displacement derivative instead of the time derivative.

The roughness of motion is widely described by jerk. This roughness is directly related to ride comfort in vertical and horizontal transportation systems. Acceleration difference Δa at the beginning and end of a diminishing time interval Δt can be used to derive the instantaneous jerk k as

$$k = \lim_{\Delta t \rightarrow 0} \frac{\Delta a}{\Delta t} = \frac{da}{dt} \quad (5)$$

All the kinematic quantities can be either positive or negative. Throughout the preceding equations, the sign of a kinematic quantity follows the base convention of defining the rectilinear displacement Δs positive or negative.

Angular Acceleration

Angular motion is illustrated in Fig. 2. Here the counterclockwise direction is defined to be positive, and the signs of the other motion quantities are selected correspondingly. This type of motion exists, for example, in electric machines and industrial robots. Therefore the measurement of angular motion quantities is of great practical importance. In angular motion the

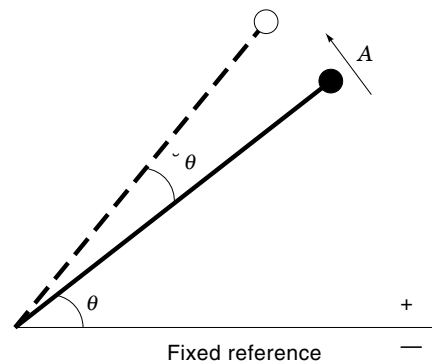


Figure 2. Angular motion of a rotating line. Its original angle is θ , and the new angle is $\theta + \Delta\theta$. *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

observed displacement is an angle $\Delta\theta$ instead of a linear distance. Thus the SI unit of angular acceleration is rad/s^2 instead of m/s^2 .

The average angular acceleration α_a during a defined time interval Δt is equal to the change in angular velocity $\Delta\omega$ per unit time during that interval. Now the instantaneous angular acceleration (or just angular acceleration) α can be expressed as

$$\alpha = \lim_{\Delta t \rightarrow 0} \alpha_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt} \quad (6)$$

The equation of angular velocity is analogous to the rectilinear velocity of Eq. (2) and can be written as a time derivative of the angular displacement

$$\omega = \lim_{\Delta t \rightarrow 0} \omega_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt} \quad (7)$$

By combining Eqs. (6) and (7), we obtain another useful formula for indirect measurement of angular acceleration:

$$\alpha = \omega \frac{d\omega}{d\theta} \quad (8)$$

Angular jerk is a seldom used quantity, but it can be computed indirectly using an equation analogous to Eq. (5).

Curvilinear Acceleration

When a particle is moving along a curved path, the motion is curvilinear. Curvilinear motion can be two-dimensional, plane curvilinear or three-dimensional, space curvilinear. Next we discuss the curvilinear acceleration, referring to Figs. 3 and 4.

Plane Curvilinear Acceleration. Plane curvilinear motion equations are very similar to those of rectilinear motion. Now we only need a two-dimensional vector instead of a scalar to

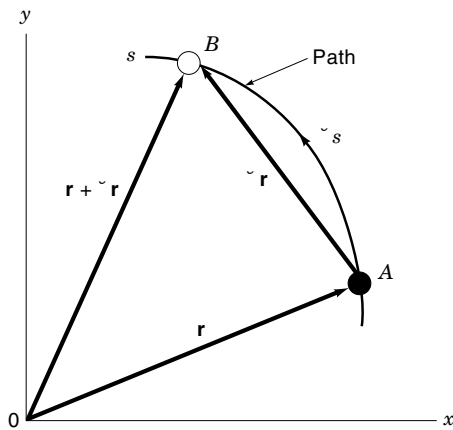


Figure 3. Plane curvilinear motion of a particle. Its original position is A, and the new position is B. The moving path is denoted by s , and Δs is the traveled distance. Now the particle can be located by a position vector $\mathbf{r}$ (and $\mathbf{r} + \Delta\mathbf{r}$). *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

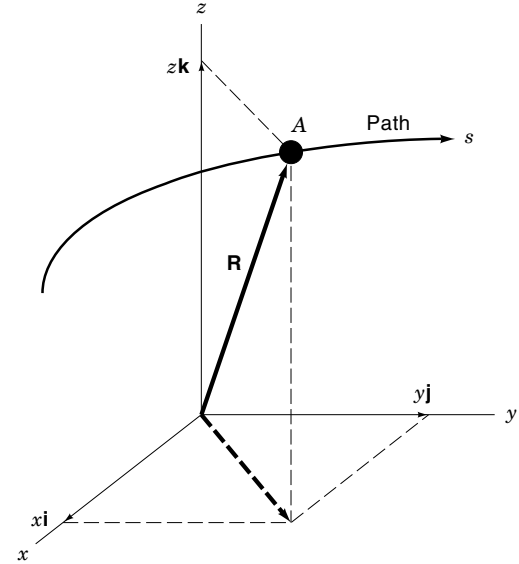


Figure 4. Space curvilinear motion of a particle along the path s . The position is determined by rectangular coordinates (x, y, z) . The position vector $\mathbf{R}$ can be expressed as $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

define the position $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$ of a particle. All the other kinematic quantities are correspondingly of vector value. Figure 3 illustrates such a motion type. The instantaneous velocity $\mathbf{v}$ can be expressed as the limiting value of the average velocity $\mathbf{v}_a$ when the observation interval Δt approaches zero:

$$\mathbf{v} = \lim_{\Delta t \rightarrow 0} \mathbf{v}_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta\mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} \quad (9)$$

The vector displacement $\Delta\mathbf{r}$ is called *linear displacement*, while the scalar displacement Δs is the traveled distance. The magnitude of the instantaneous velocity is called the *speed* of the particle. Next we define the instantaneous acceleration $\mathbf{a}$ using the limiting value of average acceleration $\mathbf{a}_a$ as the observation interval approaches zero:

$$\mathbf{a} = \lim_{\Delta t \rightarrow 0} \mathbf{a}_a = \lim_{\Delta t \rightarrow 0} \frac{\Delta\mathbf{v}}{\Delta t} = \frac{d\mathbf{v}}{dt} = \frac{d^2x}{dt^2}\mathbf{i} + \frac{d^2y}{dt^2}\mathbf{j} \quad (10)$$

Different coordinate systems (rectangular, normal, tangential, and polar) are commonly used to describe plane curvilinear motion. A detailed discussion of these is given by Dally et al. (2).

Space Curvilinear Acceleration. Rectilinear acceleration of Eq. (3) is a special case of plane curvilinear acceleration of Eq. (10). Further, plane curvilinear acceleration is a special case of space curvilinear acceleration (see Fig. 4). The most general form of motion, space curvilinear motion occurs in a three-dimensional space. We use the rectangular coordinates in defining the kinematic quantities.

The spatial position $\mathbf{R}$ can be expressed as

$$\mathbf{R} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \quad (11)$$

The corresponding spatial velocity $\mathbf{v}$ is a time derivative of the spatial position

$$\mathbf{v} = \frac{d\mathbf{R}}{dt} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} \quad (12)$$

Finally, we can write the spatial acceleration $\mathbf{a}$ as a time derivative of the spatial velocity

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2x}{dt^2}\mathbf{i} + \frac{d^2y}{dt^2}\mathbf{j} + \frac{d^2z}{dt^2}\mathbf{k} \quad (13)$$

Cylindrical and spherical coordinates are considered in Dally et al. (2). The choice of coordinates (rectangular, cylindrical, or spherical) for a particular application, involving space curvilinear motion, depends on the nature of the motion and the choice of measurement techniques.

VIBRATORY ACCELERATION

Vibratory acceleration is a distinct form of planar or spatial acceleration. It can be one-, two-, or three-dimensional. The principal characteristic of vibratory motion is that it occurs around some equilibrium position. Therefore the longer-term time average of the vibratory displacement is zero.

Vibratory motion occurs, for example, in all kinds of machines, as well as in almost all natural and artificial structures. The considerable interest in vibration analysis has three main motivations: (1) persistent vibration causes long-term wearing and possible breakdown of vibrating structures, (2) even slight vibration can degrade the performance of vibrating machines, and (3) medium-frequency vibration can also cause disturbing audible noise. For monitoring the condition of machines and installations, diagnostic procedures based on frequency-domain vibration analysis are in everyday use. Piezoelectric accelerometers have become widely used in machine vibration acquisition. Small dimensions and rigid design allow their utilization in various fields of technology. In contrast to displacement pickup and velocity pickup, these acceleration sensors can be applied into a wider range of frequencies (9).

A simple vibration can be modeled as a periodically repeated motion about the position of equilibrium (2). Figure 5 illustrates a rotating line representation of a simple type of vibratory motion that occurs commonly in various physical systems. This two-dimensional vibration can be separated into vertical (y -axis) and horizontal (x -axis) vibration components. The $x(y)$ position of the vibrating particle $P(Q)$ can be expressed as a function of time as

$$\begin{cases} x_P = A_0 \cos \omega t \\ y_Q = A_0 \sin \omega t \end{cases} \quad (14)$$

By differentiating the separate xy -position components of Eq. (14), we obtain the corresponding horizontal and vertical velocities of the vibratory motion:

$$\begin{cases} v_P = \frac{dx_P}{dt} = -A_0\omega \sin \omega t \\ v_Q = \frac{dy_Q}{dt} = A_0\omega \cos \omega t \end{cases} \quad (15)$$

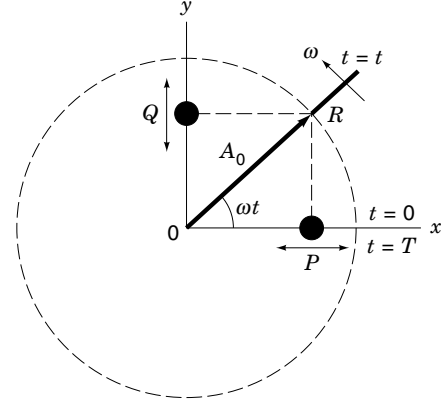


Figure 5. Rotating line representation of a simple type of vibratory motion. A_0 is the amplitude of vibration, T is the vibration period, and ω is the angular velocity of the rotating line OR . *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Finally, after differentiating the vibratory velocity components, we can write the vibratory acceleration components as

$$\begin{cases} a_P = \frac{dv_P}{dt} = -A_0\omega^2 \cos \omega t \\ a_Q = \frac{dv_Q}{dt} = -A_0\omega^2 \sin \omega t \end{cases} \quad (16)$$

Given the basic trigonometric equivalencies of Eq. (17), we can easily see that the vibratory velocities, Eq. (15), can be calculated straightforwardly by multiplying the corresponding displacement, Eq. (14), by ω and increasing the angle by $\pi/2$ radians:

$$\begin{cases} \sin \omega t = -\cos \left(\omega t + \frac{\pi}{2} \right) \\ \cos \omega t = \sin \left(\omega t + \frac{\pi}{2} \right) \end{cases} \quad (17)$$

Similarly, given the equivalencies of Eq. (18), we obtain the vibratory accelerations, Eq. (16), by multiplying the displacements, Eq. (14), by ω^2 and increasing the angle by π radians:

$$\begin{cases} \cos \omega t = -\cos(\omega t + \pi) \\ \sin \omega t = -\sin(\omega t + \pi) \end{cases} \quad (18)$$

Now we can write the vibratory acceleration as a simple function of the vibratory position:

$$\begin{cases} a_P = -\omega^2 x_P \\ a_Q = -\omega^2 y_Q \end{cases} \quad (19)$$

Any motion for which the acceleration is proportional to the displacement from a fixed point on the path of motion and always directed toward that point is defined as *simple harmonic motion* (2). Periodic motion that is not simple harmonic motion can be modeled as a sum of several simple harmonic

motions with different displacement amplitudes and harmonically related angular frequencies. This natural generalization is discussed further by Dally et al. (2).

ACCELERATION MEASUREMENT TECHNIQUES

After defining the different types of motion, we are ready to go into the details of the available measurement techniques. The main emphasis of our discussion is in the indirect measuring methods of acceleration. The direct methods, based on commercially available linear or angular accelerometers, are usually straightforward instrumentation applications where low absolute or differential voltages or currents are measured from a specific acceleration sensor. For a comprehensive presentation of accelerometers see Accelerometers.

Indirect Acceleration Measurement

Indirect acceleration measuring techniques are based on analog or digital postprocessing of position or velocity signals. These are typically measured by pulse encoders or tachogenerators. Therefore some kind of differentiator needs to be constructed to provide the acceleration signal. As discussed earlier, this complete differentiator is not trivial because the differentiation operator is noise amplifying by nature. To loosen the noise attenuation requirements of the final differentiator that produces the acceleration signal, the measuring noise problem must be tackled already in all the functional blocks of the preceding measuring chain. High-performance velocity measuring techniques are presented, for example, by Brown et al. (9), Pasanen et al. (8), and Laopoulos and Papageorgiou (10). The articles contain velocity measuring techniques that can be parametrized to provide low output noise. The applied sensors are pulse encoders. Thus the proposed methods can be used to measure both angular and rectilinear velocities, which can then be postprocessed by a differentiator.

In principle, it is easy to attenuate any noise when the primary signal and the disturbing noise are clearly separated in the frequency domain. However, the filtering task becomes considerably more difficult if there also are strict delay constraints for the filtering process. Simple nonrecursive and recursive smoothing techniques to enhance the quality of the differentiator output were suggested by Jaritz and Spong (11). The z -domain transfer functions of their smoothers are given as

$$S_n(z) = \frac{1}{N} \sum_{i=0}^{N-1} z^{-i} \quad (20)$$

$$S_r(z) = \frac{1/N}{1 - \sum_{i=1}^{N-1} z^{-i}/N} \quad (21)$$

These smoothers still suffer from a notable tracking error or lag. This forces the designer to use a small value of N that, on the other hand, leads to poor noise attenuation capabilities.

Ultimately we would prefer a delayless lowpass filter because any additional delay degrades the overall performance when the filtered acceleration signal is used for feedback control or time-critical supervision. Unfortunately, there exists no general-purpose lowpass filter that does not delay the frequencies on its passband. At present we can choose between two

approaches to solve this difficult problem: predictive filtering or state observing. In the first case, an application-specific, predictive lowpass filter is cascaded with a differentiator, and in the second, a linear (or nonlinear) stochastic model to represent acceleration is developed and used as an acceleration estimator. Under greatly time-varying conditions, the predictive filter as well as the stochastic estimator should be adaptive to maximize the noise attenuation capabilities and to minimize the harmful lag of the primary acceleration curve.

Predictive Postfiltering. Predictive filters (8) as well as state observers (12) are usually implemented in the discrete time domain either because they have their theoretical origins in digital signal processing or because an analog implementation would be unfeasible due to the mathematical operations required. This assumption is valid throughout the following discussion. Predictive (or phase-advancing) lowpass filtering is possible if we can place easing constraints for the incoming signal characteristics. These constraints may be explicit either in the time domain or in the frequency domain. There are many practical applications where the kinematic quantities can be approximated with sufficient accuracy by piecewise low-degree polynomials, as illustrated in the experimental sections of Refs. 12, 8, and 5. This is due to the mechanical inertia that necessarily smoothens the movements of masses. Typical rectilinear position curves, such as those related to vertical and horizontal transportation, can be approximated piecewise by third- and fourth-degree polynomials, velocity curves by second- and third-degree polynomials, acceleration curves by first- and second-degree polynomials, and jerk curves by zeroth- and first-degree polynomials. For instance, a third-degree polynomial model can be expressed as

$$f(k) = \beta_0 + \beta_1 k + \beta_2 k^2 + \beta_3 k^3 \quad (22)$$

where k is the discrete time index, and $\beta_i, i = 0, 1, 2, 3$, are the signal-dependent curve-fitting parameters. A thorough discussion of polynomial modeling is given by Williams (13).

By using the polynomial signal model, we can easily design predictive filters that provide the desired forward prediction behavior with the mandatory lowpass characteristics (14). One-step-ahead prediction is typically adequate to compensate for the delay caused by the differentiation algorithm as well as the data acquisition and processing delays. Thus we are performing on-line curve fitting. For this we only need to select an appropriate polynomial degree; the signal-dependent parameters of the polynomial model are handled implicitly.

The difference equation of a general n -step-ahead predictive filter can be expressed as

$$\hat{u}(k+n) = \sum_{i=1}^N \chi_i \hat{u}(k+n-i) + \sum_{j=0}^M \delta_j u(k-j) \quad (23)$$

where the coefficients χ_i and δ_j are real-valued constants, $\hat{u}(k+n)$ is the n -step-ahead output of the predictive filter, and $u(k)$ is the corresponding input sample. The prediction step n is an application-specific parameter, and it depends on the cumulative measuring and processing delays of the entire instrumentation system. An important consequence of the polynomial model is naturally that the result of predictive filtering is less satisfactory for other signal classes, like vibratory

acceleration. This is due to a narrow prediction bandwidth (a frequency range where the group/phase delay of the filter is negative) that is a principal restricting characteristic of polynomial predictive filters. There exist several finite impulse response (FIR) and infinite impulse response (IIR) polynomial predictors as reviewed by Ovaska (15).

Recursive linear smoothed Newton (RLSN) predictors form a class of computationally efficient IIR predictors (14), which are particularly attractive for postprocessing of the noisy output of a differentiator. Their applicability is an immediate consequence of the simple design process and efficient noise attenuation capabilities. The z -domain transfer functions of the first- and second-degree [$H_1(z)$ and $H_2(z)$, respectively], one-step-ahead RLSN polynomial predictors are given as

$$H_1(z) = \frac{[c + (1/N)] - (z^{-N}/N)}{1 - (1 - c)z^{-1}} \quad (24)$$

$$H_2(z) = \frac{[2c + (1/N)] + [c^2 - 2c - (1/N)]z^{-1} - (z^{-N}/N) + (z^{-(N+1)}/N)}{1 - (2 - 2c)z^{-1} + (1 - 2c + c^2)z^{-2}} \quad (25)$$

The only adjustable parameters, c and N , control primarily the stopband attenuation and the passband gain peak, respectively. In acceleration measurement applications, typical values for these parameters are $c = 0.01$ – 0.05 and $N = 16$ – 64 .

Hence we can postfilter the noisy output of the simple backward-difference differentiator of Eq. (26) by a polynomial predictive filter without harmfully delaying the low-degree polynomial component. This polynomial component approximates our desired acceleration signal. The delay introduced by the difference operation of Eq. (26) is half a sampling period.

$$a(k) = \frac{v(k) - v(k-1)}{T_s} \quad (26)$$

In Eq. (26), $a(k)$ is the average acceleration, $v(k)$ the instantaneous velocity, and T_s the constant sampling period. The accuracy of Eq. (26) can be improved by decreasing the sampling period as suggested in the original definition, Eq. (3), of this time derivative. In practical applications the selection of the sampling period is based on the bandwidth requirements of the closed acceleration control loop, the latency requirements of some supervision action such as emergency shutdown, and the computational capacity of the implementation environment. Figure 6 depicts the indirect acceleration measuring scheme based on differentiation and predictive postfiltering. Three alternative implementation techniques are presented: fully analog, analog-digital, and fully digital. Analog polynomial predictors needed in the fully analog alternative were introduced in Ref. 16.

Linear State Observing. Instead of differentiation and (predictive) postfiltering, an optimized Kalman filter-based state observer can be used for estimating acceleration. An angular acceleration estimator is proposed by Bélanger (12), but the same principle can be used in estimating rectilinear acceleration. First, a stochastic state-space model is developed to rep-

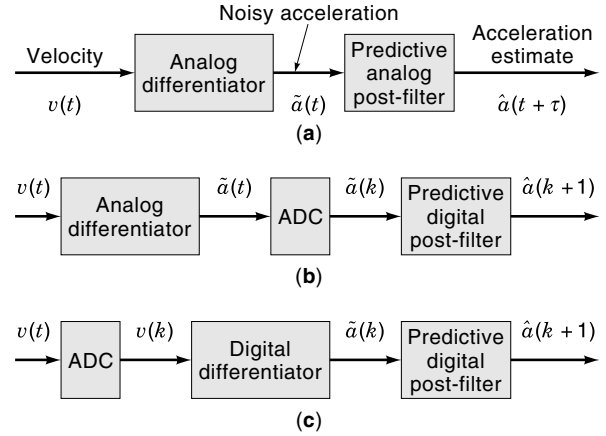


Figure 6. Alternative implementations of indirect acceleration measurement using a cascade of a differentiator and a polynomial predictive postfilter. The prediction step is τ with the analog predictive filter and one sampling period with the digital ones.

resent the rotation angle $\theta(t)$:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \Gamma w(t) \\ y(t) = \theta(t) = \mathbf{C}\mathbf{x}(t) + e(t) \end{cases} \quad (27)$$

where $\mathbf{x}(t)$ is the state vector containing the angle, angular velocity, and angular acceleration; $w(t)$ is zero mean, white Gaussian noise with covariance q . When the modeled motion is not characterized by such a stochastic process but is merely deterministic, the parameter q may be considered as a pure filter parameter to be adjusted empirically. Further $\mathbf{A}$ is a (3×3) matrix, Γ is a (3×1) vector, and $\mathbf{C}$ is a (1×3) vector. The scalar output $y(t)$ is the actual angle measurement, $\theta(t)$, and $e(t)$ is the additive quantization error. The versatile Kalman filter (11,17) provides an optimal (minimum variance) solution to this observing problem (12). Before the model of Eq. (27) can be implemented by some digital processor, it must be discretized using one of the available continuous-time to discrete-time transformations (18).

For estimating the angular acceleration, Bélanger (12) postulates the fixed numerical model:

$$\left\{ \begin{array}{l} \frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \\ y(t) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x}(t) + e(t) \end{array} \right. \quad (28)$$

The application discussed in that article is a single robot joint under proportional-derivative (PD) control with all the evaluated estimates corresponding to one-step-ahead prediction. Although only a marginal improvement of the angle

estimates is reported over those provided directly by the pulse encoder, the velocity estimate standard deviations are improved by a factor of 2 to 4 over the standard deviations of the plain backward-difference differentiation. In the estimate of the angular acceleration, there is an order of magnitude improvement. Hence even the fixed Kalman filter approach offers clear benefits by remarkably improving the accuracy of acceleration measurements. On the other hand, the computational complexity of the Kalman filter may be a limiting factor when the estimator must be adapted on-line with a high sampling rate. Nevertheless, such an adaptive approach offers satisfactory accuracy even under unknown and time-varying conditions.

An advantageous characteristic of such a linear state observer is that it provides estimates of all the state variables *simultaneously*. Therefore no explicit cascade processing structure is needed as in the case of differentiation and predictive postfiltering. When direct measurements of the velocity signal are available, the number of states, namely the dimension of the state vector $\mathbf{x}(t)$, is reduced from three to two. This simplifies the implementation of the optimized Kalman filter.

Conclusions of the Indirect Methods. As a conclusion, we can state that backward-difference differentiation cascaded with a polynomial predictive filter provides a computationally efficient indirect method for acceleration measurement. However, the method is applicable solely when the acceleration curve can be approximated piecewise by a low-degree polynomial. Practical degrees of the polynomial model are less than three or four because the available polynomial predictive filters for higher degree polynomials can offer only marginal noise attenuation capabilities (14).

Linear state observing is an attractive method when no such assumption on the polynomial nature of the acceleration curve can be made. By using an adaptive Kalman filter, estimators for time-varying acceleration behavior can be developed if only the required computational complexity can be supported by the available implementation hardware. Also, nonlinear state observers can be developed to handle the possible nonlinearities of the underlying dynamic system, as shown in the case of velocity estimation by Jaritz and Spong (11).

Direct Acceleration Measuring

Direct acceleration measuring is based on sensors that transform either linear or angular acceleration into an electrical variable: charge, voltage, or current. Acceleration measurement without a fixed reference requires the use of a seismic transducer (2). Those transducers detect relative motion between a fixed mounting base and a moving seismic mass. The seismic mass tends, due to inertia, to resist any changes in the movement. Acceleration sensors require an extremely small mass, which is connected to the frame through a stiff spring. This makes it possible to provide a wide operating bandwidth.

The currently available accelerometers usually need an external high-sensitivity preamplifier with a high-input impedance to amplify the weak primary signal (proportional to acceleration) to a suitable level for the following signal processing or data acquisition electronics. In recent designs, however, the critical preamplifier is often incorporated into the transducer housing. This is an obvious advantage because an application engineer can concentrate on higher-level

instrumentation electronics instead of highly sensor-dependent solutions. Now a high-level voltage output with moderate or high signal-to-noise ratio (SNR) is obtainable. On the other hand, some design flexibility is always lost with such integrated components.

Linear and Vibratory Acceleration. Compact piezoelectric accelerometers are widely applied to the measuring of linear acceleration due to their wide operating bandwidth, usually from a few Hz to several kHz. This wide bandwidth is particularly useful in precise inertial navigation and the measurement of vibratory acceleration. In inertial navigation the spatial acceleration of Eq. (13) is measured with three accelerometers, one accelerometer for each of the three dimensions (x, y, z). Vibratory acceleration may also have more than one dimension, but all these spatial dimensions are usually represented by individual linear components.

Piezoelectric accelerometers are charge-generating devices, and after the necessary charge-to-voltage conversion and preamplification, they can produce typical output voltages of 10 mV/g to 30 mV/g (where g is the acceleration of gravity, 9.8 m/s^2) with accuracies of a few percent (2,3). Hence the electrical measuring task is necessarily more demanding with low acceleration levels than it is with moderate or high accelerations. The operating range of such sensors is typically from zero to a few hundred or thousand g : s. Manufacturers of piezoelectric accelerometers include, e.g., the following companies (in alphabetic order): Endevco Corporation, PCB Piezotronics, Kistler, and Murata. For a comprehensive presentation of various types of accelerometers see Accelerometers.

We can model a piezoelectric accelerometer as a charge generator, which is connected in parallel with an internal capacitor C_T . The terminals of this capacitor are the actual output pins of the transducer. Therefore we can measure the voltage difference over these pins. When we connect a measuring cable to the terminals of the piezoelectric accelerometer, the cable and possible connectors introduce an additional capacitance C_C , which is summed to the internal capacitance. Further the input capacitance of the preamplifier C_I is an additional component of the total capacitance $C_\Sigma = C_T + C_C + C_I$. An ac-coupled voltage follower or an instrumentation amplifier is preferred to amplify the low voltage that is observed over the total input capacitance. Figure 7 illustrates the voltage follower-based measuring circuitry. Typical values of the total capacitance C_Σ range from 300 pF to 10 nF, and the value of the ac-coupling capacitor C_A is usually about 100 nF (2).

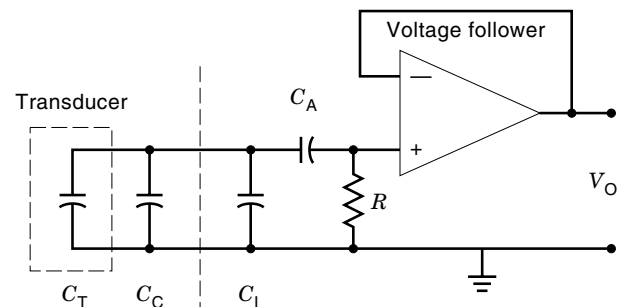


Figure 7. Voltage follower-based acceleration measuring scheme using a piezoelectric transducer.

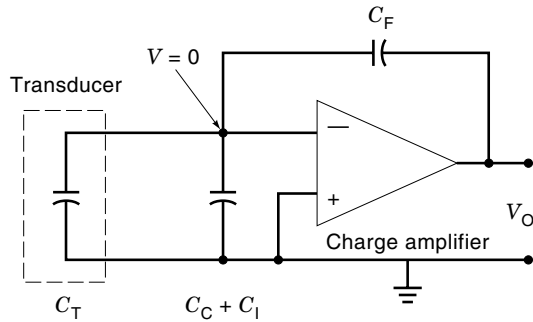


Figure 8. Charge-to-voltage converter-based acceleration measuring scheme using a piezoelectric transducer.

Unfortunately, the voltage sensitivity of such a straightforward measuring system is inversely proportional to the total capacitance C_{Σ} (2). This may cause accuracy problems because, in an industrial operating environment, the cable capacitance C_C can change remarkably due to varying environmental conditions, such as humidity and dirt. Besides, the variation of C_C causes changes in the transfer function of the dynamical system formed by the sensor, cable, and preamplifier. These changes necessarily affect both the magnitude and phase (or delay) responses of the acceleration sensor.

Thus there is a natural demand toward integrated sensor modules, where the critical wiring would be of minimal length and the entire circuitry could even be hermetically sealed. Also the critical component values would be trimmed by the manufacturer during fabrication of the accelerometer module to provide constant voltage sensitivity and dynamical characteristics.

Charge amplifiers are widely used to preamplify the output of a piezoelectric acceleration transducer. They make use of an operational amplifier having a high open loop gain. A complete charge amplifier consists of two cascaded sections: a charge-to-voltage (C/V) converter and a trimmed voltage amplifier. Figure 8 illustrates a basic inverting C/V-converter, which is formed by using a capacitive feedback C_F . The voltage gain of this circuit is proportional to C_F/C_T . Here the possibly varying cable capacitance C_C appears between the summing point and circuit common. Because the voltage at the summing point is zero, C_C does not affect the provided voltage gain. However, the cable capacitance naturally affects the noise gain of the C/V-converter, which is proportional to $(C_T + C_C)/C_F$. Morrison (19) gives a short practical presentation of the implementation aspects of charge amplifiers. His discussion concentrates on the noise characteristics and the required component tolerances of charge-to-voltage converters. Hence the voltage sensitivity is subject solely to negligible variations due to changes in environmental conditions. On the other hand, a complete charge amplifier (C/V-converter cascaded with a standardization amplifier) is more complex than a simple voltage follower. Charge amplifiers can handle frequencies down to about 1 Hz (3-dB point). A detailed presentation and analysis of charge amplifier-based measurement circuits is given by Dally et al. (2).

Angular Acceleration. Although the angular acceleration can be measured indirectly using either a rotating angle sensor or a velocity sensor, the cumbersome noise-amplification problem associated with differentiators has motivated the efforts to develop transducers for direct sensing of angular acceleration.

Direct measuring of linear acceleration is in wide everyday use, but the angular acceleration sensors, particularly those with unlimited rotation angle, can still be considered as emerging devices. Therefore the measuring techniques to be discussed below are not yet used widely in practical applications.

Godler et al. (5) proposed a rotary acceleration transducer that senses the angular acceleration independently of the rotation velocity and has an unlimited rotation angle. This *mechanic-opto-electronic* sensor is intended for motor control and vibration control in robotics applications. It has an obvious advantage when compared with indirect measuring techniques: The bandwidth of the output filter can be made wider than that of the (predictive) postfilter attenuating the noise of the differentiator. This is due to the more advantageous shape of the introduced noise spectrum. Therefore the new sensor offers a wider acceleration bandwidth that could directly improve the performance of an acceleration control loop. Considering the structural details and laboratory experiments presented in Ref. 5, it can be concluded that the mechanic-opto-electronic angular acceleration sensor can provide comparable accuracy with the indirect time derivative methods when the base velocity ω of Eq. (6) is measured, for example, by a low-ripple dc tachogenerator. The sensor can provide a buffered voltage output, and hence its usage is very simple. No demanding preamplifiers are needed as with the piezoelectric accelerometers. Unfortunately, only evaluation prototypes were available when Ref. 5 was published in 1995.

Furukawa et al. (20) proposed a *piezoresistive* angular acceleration sensor for robotics applications. This sensor is manufactured by micromachining a small wafer of silicon. The micro acceleration sensor manufactured by integrated circuits (IC) technology has the advantage that the necessary measuring electronics can be integrated in the same chip design as the sensor itself. Therefore a compact, low-noise, and possibly even an intelligent accelerometer unit could be developed. The final accelerometer is so small that it can be mounted at most required places and will not disturb the primary motion of a robot arm. From the instrumentation point of view, the electronics required simply measures the change of sensor resistance, which is proportional to the angular acceleration. This could be accomplished using the Wheatstone bridge (21). In contrast to the proposed sensor of Godler et al. (5), this sensor operates solely within 0° to 360° angles. This accelerometer was still in the prototype phase when Ref. 20 was published.

Conclusions of the Direct Methods. Direct methods to measure linear acceleration are widely used because of the availability of accurate, compact, and low-cost accelerometers. They are far more accurate than indirect methods for wide acceleration bandwidths. However, in cases where a narrow bandwidth is adequate, either a direct or an indirect method can be applied with comparable results. The selection of an appropriate technique depends on the presence of distance or velocity sensors; if no other sensor type is needed in the specific application, it is natural to use only an accelerometer.

Angular acceleration sensors are not yet widely available as commercial products. The demand for such transducers will rise steadily as the performance requirements of various motion control applications increase. The basic components are presented in Refs. 5 and 20. Therefore, the field of direct angular acceleration measurement will become an important area of future research and development activities.

ACCELERATION MEASUREMENT ERRORS

Acceleration measurement errors are due to three primary sources: sensors, acquisition electronics, and signal processing algorithms. For an economical and robust measurement system, the total measurement error should preferably be evenly distributed into different parts of the entire instrumentation chain.

Errors of the Direct Acceleration Measurement

In direct acceleration measurement, the primary error sources are usually the acceleration sensor and the immediate acquisition electronics. This combined error is typically less than 2% to 5% with piezoelectric accelerometers when the measurement electronics design is based on the application notes of the corresponding sensor manufacturer. Within this moderately accurate range the errors related to sampling and analog-to-digital (A/D) conversion are easily controlled by careful selection of multiplexers, sample-and-hold (S&H) circuits, and A/D-converters. Garrett gives a thorough presentation of the different error types of computerized instrumentation systems in Ref. 4. Those error components are not acceleration measurement-specific but exist in all computerized data acquisition systems.

Errors of the Indirect Acceleration Measurement

In indirect acceleration measurement the approximative differentiation is the main source of noiselike error. On the other hand, when a state observer is used instead of differentiation and postfiltering, the accuracy and bandwidth of the state-space model of Eq. (27) become especially important. Pulse encoders and tachogenerators provide the base quantities, linear or angular displacement and velocity, for indirect acceleration measurement. The measurement error of such a base sensor is typically no more than 1%. However, there is a complicated dependence between the error of the base sensor and the final acceleration error, particularly when the velocity estimate is first calculated from encoder pulse information. This is due to the usual nonlinearity of velocity estimation algorithms (8,10).

To quantify the noise problem of the backward-difference differentiator, Eq. (26), we consider a simple numerical example. Let us assume that we are measuring linear motion with constant velocity of 1.00 m/s, and our dc tachogenerator has a 1% ripple in its output voltage. The output is sampled with a constant 1-ms sampling period T_s . Let $v(k) = 0.995$ m/s and $v(k-1) = 1.005$ m/s be two consecutive measurements of the linear velocity containing sensor-originated ripple. Now the simple estimate of the instantaneous acceleration $a(k) = -10$ m/s². Although the base velocity with the 1% ripple could be considered adequate for many velocity control applications, even this small error makes the direct utilization of the differentiator output impossible. Therefore postfiltering is needed to reduce the acceleration ripple down to an acceptable level. In off-line data acquisition a conventional analog or digital low-pass filter can be used to achieve a suitable SNR. However, in real-time applications a predictive filter can offer indisputable benefits as discussed earlier.

To get insight into the required bandwidth of the low-pass postfilter, let us consider the simple case where the constant velocity signal v_c is deteriorated by an additive sinusoid

ripple $A \sin 2\pi ft$. The available measurement signal $v_r(t)$ can be expressed as

$$v_r(t) = v_c + A \sin 2\pi ft \quad (29)$$

This signal is sampled using a sampling period of T_s . When we apply the differentiator of Eq. (26), the maximum erroneous acceleration a_m has the value

$$a_m = \frac{2A}{T_s} \sin \pi f T_s \quad (30)$$

If the amplitude A of the sinusoid ripple is 0.005 m/s, and the sampling period is the same as in the example above, we can calculate the maximum acceleration value a_m for different sinusoid frequencies. With $f = 500$ Hz, $a_m = 10.00$ m/s²; with $f = 50$ Hz, $a_m = 1.56$ m/s²; with $f = 5$ Hz, $a_m = 0.16$ m/s²; and with $f = 0.5$ Hz, $a_m = 0.02$ m/s². This shows clearly the trade-off between the bandwidth and the corresponding erroneous acceleration level.

Hence a very narrow-band lowpass filter is needed to keep the disturbing acceleration ripple moderate or low. This is an obvious motivation for the usage of polynomial predictive filters in real-time applications because they can provide an adequate prediction bandwidth even when a narrowband magnitude response is required.

Quantization of the velocity signal during the sampling process is also a potential source of remarkable acceleration error because the quantization noise is amplified similarly as the sensor-originated ripple by the differentiator. There are two basic alternatives to alleviate this problem: Either perform the differentiation in the analog domain or use high-resolution quantization. In practice, however, the analog alternative is less attractive due to the large dynamic range required. If an analog differentiator is employed, a scaling amplifier with a programmable gain is suggested to be placed in front of the A/D-converter. Besides the sensor ripple and quantization noise, the coefficient word length and precision of arithmetic operations have effects on the noise level which is observed in the output of the composite differentiator. All these are very critical in real-time applications because of the greatly limited filtering freedom.

Since there are many potential sources of significant acceleration error in indirect acceleration measurement, the total error (or signal-to-error ratio, SER) is largely dependent on implementation. Thus a careful error analysis is important when any indirect technique is used. Typically even in well-designed systems the true maximum errors are between 5% and 10% depending on the required acceleration bandwidth and the allowed delay. In some applications this error is mainly due to the overshoot of a postfilter output under transient conditions. Such error percentages are obtainable in vertical and horizontal transportation as well as in robotics applications.

SPECIFIC ACCELERATION MEASUREMENT METHODS AND APPLICATIONS

Our presentation has, for the present, concentrated on the main principles of the available acceleration measurement methods. The characteristics and implementation considerations were presented mostly for the pure "main-line" techniques: direct

measuring using accelerometers, as well as indirect measuring based on either differentiating and (predictive) postfiltering, or state-observing. Besides these straightforward and multipurpose methods, there exist a vast number of application-specific and application-tailored techniques that can provide an excellent base for future development and engineering activities.

Selective Review of the Advanced Literature

A collection of diverse acceleration measurement methods will be reviewed briefly here. The assortment consists solely of indirect measuring techniques, particularly techniques for indirect measuring of angular acceleration. Our emphasis is set intentionally on the angular acceleration measurement. Since there exist only a few commercial (and economical) angular accelerometers with unlimited rotation range, the indirect techniques are naturally of utmost importance.

Hoffmann de Visme (22) introduced a purely digital method for obtaining acceleration information of rotating shafts in 1968. His method is based on a rotating pulse encoder, and it gives good sensitivity and accuracy when a high pulse rate N (pulses per revolution) is available and a long measuring interval T can be tolerated. The estimated angular acceleration α (rad/s²) is given by

$$\alpha = \frac{2\pi(n_1 - n_2)}{NT^2} \quad (31)$$

where n_1 and n_2 pulses are counted in two successive intervals of duration T . This method suits well both for hardware and software implementations.

Dunworth (6) proposed a sophisticated digital instrument for the measurement of angular acceleration, in which a rotating pulse encoder is used. This method comprises a digital frequency register controlling a variable-rate pulse generator that tracks the incoming pulse rate in a closed-frequency control loop. The rate at which the pulse generator frequency needs to be corrected (incremented or decremented) corresponds to the present acceleration if the control loop is just locked on the input pulse rate. Acceleration resolution of 1% is easily achievable. Although Dunworth's instrument shares the possibly troublesome long measuring interval of Ref. 22, it still provides an interesting measuring procedure that is particularly well suited for application-specific integrated circuit (ASIC) implementation.

Smith et al. (23) reported on a direct software implementation of Eq. (31) in 1973. A pulse encoder with 10,800 pulses per revolution was used. They utilized the angular acceleration estimate for computing the electromagnetic torque of an electric motor. When the inertia of the rotating loaded component is known and assumed to stay constant, the torque can be calculated as the product of acceleration and inertia. The presented experimental results are in close agreement with the corresponding theoretically computed quantities, the maximum errors being no more than 5% to 10%.

Hancke and Viljoen (7) presented an acceleration measurement technique for a condition-monitoring application of a turbo generator where only one pulse per shaft revolution is available. This naturally causes serious constraints for the obtainable performance, namely for resolution and measuring interval. The time interval between two consecutive pulses is measured using a counter that is clocked by a high-frequency square wave. An interesting characteristic of the developed

method is that it can estimate the angular acceleration between pulse instants satisfactorily by extrapolation of the past two acceleration samples. This extrapolation is based on a ramp assumption, namely a first-degree polynomial model. Here the applicability of the first-degree polynomial model is justified by the obvious fact that the inertia of the rotating components is so high that the angular velocity cannot vary greatly between any two pulses. If the angular acceleration estimates corresponding to the latest two pulse instants are $\alpha(k-1)$ and $\alpha(k)$, the next acceleration sample, $\alpha(k+1)$, can be extrapolated as

$$\alpha(k+1) = \alpha(k) + [\alpha(k) - \alpha(k-1)] \quad (32)$$

Now linear interpolation can simply be used to compute the acceleration estimates at arbitrary instants between the discrete time indices k and $k+1$.

Kadhim et al. (24) presented a straightforward method for the measurement of steady-state and transient acceleration of a rotating motor shaft. They suggested the use of a slight modification of Eq. (31); three total pulse counts (c_1, c_2, c_3) are used instead of the two pulse count differences (n_1, n_2):

$$\alpha = \frac{2\pi(c_3 - 2c_2 + c_1)}{NT^2} \quad (33)$$

Also here the increase of the measuring interval T improves the accuracy of average acceleration. However, it increases as well the harmful delay between the instant when a measurement becomes available and the moment at which it applies. Kadhim et al. have suggested that for off-line applications, low-degree polynomial curve-fitting techniques can improve the signal-to-error ratio. Also on-line curve-fitting techniques are readily available for such a purpose, as discussed above in "Predictive Postfiltering."

Laopoulos and Papageorgiou (10) proposed an angular acceleration measurement instrument based on a 24-bit signal processor in 1996. They computed the difference of the pulse encoder's output frequency ΔF for two consecutive pulse cycles with the measuring interval of ΔT . Hence the angular acceleration is directly proportional to $\Delta F / \Delta T$. In this division both the frequency difference and the measuring interval should have a high resolution to keep the amplified quantization noise low. To provide this necessity, a 24-bit binary counter was applied to count the period of an encoder pulse cycle using a 20 MHz clock. The angular acceleration is calculated from

$$\alpha = \frac{2\pi \Delta F}{N \Delta T} \quad (34)$$

Notice the varying length of the measuring interval ΔT which depends on the availability rate of fresh encoder pulses. With this method the accuracy of the (nonuniformly sampled) acceleration estimate is determined largely by the random jitter corrupting the nominal encoder pulse length. A high-performance, uniformly sampled acceleration measurement method can be constructed by combining the predictive synchronization and restoration method of Pasanen et al. (8) with the accurate acceleration measurement procedure of Ref. 10.

All the methods discussed above are based on the noise-amplifying difference function: either a pulse count difference or a frequency difference. Schmidt and Lorenz (25) use another possible indirect approach; they propose a computationally

efficient acceleration observer for a dc servo drive. Their acceleration observer uses two measured input quantities: the shaft angle and the actual current of the dc motor. The entire state-observer contains a velocity observer block that provides the estimate of the angular velocity for the following acceleration observer. It should be noted that this state-observer is intimately connected to the overall controller structure. A sampling period of 1.25 ms was successfully applied, and all the computations were performed using high-precision floating-point arithmetic. The observer-generated acceleration feedback was shown to improve the control performance remarkably.

Evaluative Discussion of the Specific Methods

The development of indirect measuring instruments of angular acceleration began in the late 1960s. Then there were two important advancements that made the classic work of Hoffmann de Visme (22) and Dunworth (6) possible: the emerging availability of rotating encoders with high-pulse rates and the rapid development of solid-state digital electronics and integrated circuits. The early instrumentation techniques were strongly technology driven without a strict connection to any specific application.

At the beginning of the 1970s no outstanding innovations occurred in the field while available methods continued to evolve, such as that of Smith et al. (23) which was implemented in a general-purpose minicomputer. Nevertheless, with the development of high-resolution optical encoders came an increase in the number of successful industry applications, and acceleration feedback provided opportunities to increase functional performance.

In the late 1970s and during all of the 1980s, the availability of microprocessors and microcontrollers turned the instrumentation engineers' interests almost entirely from pure hardware implementations to software implementation of acceleration measurement procedures. During that productive period the entire control and instrumentation industry began to introduce novel microprocessor-based products, pushing the need for acceleration feedback mostly to the background.

Finally, at the very beginning of the 1990s, it became evident that the performance of motion control systems in the well-established robotics and servo industry could be improved economically by using additional acceleration feedback. Acceleration feedback is advantageous because it allows higher overall stiffness without requiring higher bandwidths of the velocity and position control loops in servo applications. Essentially, as is pointed out by Schmidt and Lorenz (25), the acceleration feedback acts like an "active inertia." Along with this application-pull factor in the resurgence of interest in acceleration feedback, there were the traditional technology-push cases. The recent development of implementation techniques (signal processors and application-specific integrated circuits), signal-processing algorithms, and estimation methods has opened a new era of growth for indirect techniques in acceleration measurement. Therefore it is foreseeable that both the availability of high-performance methods (10,12,25), and their successful applications are going to increase steadily.

CONCLUSION

A comprehensive variety of direct and indirect techniques exist for acceleration measurement. Application-tailored measurement techniques currently predominate in real-time systems,

and computerized instrumentation, in general, is becoming widespread. However, mixed analog-digital processing offers some benefits over pure digital signal processing in cases where the primary measurement quantity is available as an analog voltage or current.

The field of linear and vibratory acceleration measurement is stable and well developed. A complete assortment of instruments for vibration measurement is available, for example, from Bruel and Kjaer Instruments. The main focus of future development is in the intelligent postprocessing of accelerometer data and the automatic interpretation of measurement results. The performance and capabilities of integrated accelerometer modules will continue to increase as low-cost accelerometers and more novel applications are introduced such as is presently being seen in the field of technical diagnostics.

The whole field of angular acceleration measurement is expanding as both direct and indirect techniques are evolving to new applications. There exist two principal challenges for the research and development community: to develop economical and accurate angular accelerometers for unlimited rotation range and to create wideband indirect techniques with a low signal-to-error ratio. It is likely that the entire technology profile of acceleration measurement will advance in the early twenty-first century.

ACKNOWLEDGMENTS

The sections on Reference Axes and Different Types of Motion, Rectilinear, Angular, and Curvilinear Acceleration and Vibratory Acceleration are partly based on the original content of *Instrumentation for Engineering Measurements*, J. W. Dally, W. F. Riley, and K. G. McConnell, Copyright © 1984, by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

BIBLIOGRAPHY

1. M. Alonso and E. J. Finn, *Fundamental University Physics, Volume One: Mechanics and Thermodynamics*, 2nd ed., Reading: Addison-Wesley, 1980, Ch. 5.
2. J. W. Dally, W. F. Riley, and K. G. McConnell, *Instrumentation for Engineering Measurements*, New York: Wiley, 1984, Ch. 7.
3. G. C. Barney, *Intelligent Instrumentation: Microprocessor Applications in Measurement and Control*, London: Prentice-Hall International, 1985, Ch. 6.7.
4. P. H. Garrett, *Advanced Instrumentation and Computer I/O Design: Real-Time System Computer Interface Engineering*, New York: IEEE Press, 1994, Ch. 7.
5. I. Godler et al., A novel rotary acceleration sensor, *IEEE Control Syst.*, **15**: 56–60, 1995.
6. A. Dunworth, Digital instrumentation for angular velocity and acceleration, *IEEE Trans. Instrum. Meas.*, **18**: 132–138, 1969.
7. G. P. Hancke and C. F. T. Viljoen, The microprocessor measurement of low values of rotational speed and acceleration, *IEEE Trans. Instrum. Meas.*, **39**: 1014–1017, 1990.
8. J. Pasanen, O. Vainio, and S. J. Ovaska, Predictive synchronization and restoration of corrupted velocity samples, *Measurement*, **13**: 315–324, 1994.
9. R. H. Brown, S. C. Schneider, and M. G. Mulligan, Analysis of algorithms for velocity estimation from discrete position versus time data, *IEEE Trans. Ind. Electron.*, **39**: 11–19, 1992.
10. T. Laopoulos and C. Papageorgiou, Microcontroller-based measurement of angular position, velocity, and acceleration, in *Proc. IEEE*

Instrumentation and Measurement Technology Conference, 73–77, 1996.

11. A. Jaritz and M. W. Spong, An experimental comparison of robust control algorithms on a direct drive manipulator, *IEEE Trans. Control Syst. Technol.*, **4**: 627–640, 1996.
12. P. R. Bélanger, Estimation of angular velocity and acceleration from shaft encoder measurements, in *Proc. IEEE International Conference on Robotics and Automation*, 585–592, 1992.
13. C. S. Williams, *Designing Digital Filters*, Englewood Cliffs, NJ: Prentice-Hall, 1986, Ch. 7.
14. S. J. Ovaska and O. Vainio, Recursive linear smoothed Newton predictors for polynomial extrapolation, *IEEE Trans. Instrum. Meas.*, **41**: 510–516, 1992.
15. S. J. Ovaska, Predictive signal processing in instrumentation and measurement: A tutorial review, in *Proc. IEEE Instrumentation and Measurement Technology Conference*, 48–53, 1997.
16. O. Vainio and S. J. Ovaska, A class of predictive analog filters for sensor signal processing and control instrumentation, *IEEE Trans. Ind. Electron.*, **44**: 565–570, 1997.
17. P. S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. 1, San Diego: Academic, 1979.
18. P. Katz, *Digital Control Using Microprocessors*, London: Prentice-Hall International, 1981.
19. R. Morrison, *Solving Interference Problems in Electronics*, New York: Wiley, 1995, Ch. 6.
20. N. Furukawa and K. Ohnishi, A structure of angular acceleration sensor using silicon cantilevered beam with piezoresistors, in *Proc. IEEE IECON*, 1524–1529, 1992.
21. M. U. Reissland, *Electrical Measurements: Fundamentals, Concepts, Applications*, New Delhi: Wiley Eastern, 1989, Ch. 7.3.
22. G. Hoffmann de Visme, Digital processing unit for evaluating angular acceleration, *Electronic Engineering*, **40**: 183–188, 1968.
23. I. R. Smith, M. J. Hajiroussou, and J. F. Miller, Precision digital tachometer, *IEEE Trans. Instrum. Meas.*, **22**: 278–279, 1973.
24. A. H. Kadhim, T. K. M. Babu, and D. O'Kelly, Measurement of steady-state and transient load-angle, angular velocity, and acceleration using an optical encoder, *IEEE Trans. Instrum. Meas.*, **41**: 486–489, 1992.
25. P. B. Schmidt and R. D. Lorenz, Design principles and implementation of acceleration feedback to improve performance of dc drives, *IEEE Trans. Ind. Appl.*, **28**: 594–599, 1992.

SEPPO J. OVASKA
Helsinki University of Technology

DENSITY MEASUREMENT

Density of a substance is defined as the mass per unit volume ($\rho = m/V$) under fixed conditions. The term is applicable to solids, liquids, and gases. The density depends on temperature and pressure. This dependence is much greater in gases. Although there are many different units, usually the values of density are given in terms of grams per cubic centimeter.

Specific gravity (SG) is an abstract number expressing the ratio of the density of one substance to the density of another reference substance, both obtained at the same temperature and pressure. For solids and liquids, water is taken as the reference substance, whereas air is the reference for gases. The specific gravities of solids, liquids, and gases under reference

conditions may be expressed by

$$\text{Liquid (or solid) SG} = \frac{\text{density of liquid (or solid)}}{\text{density of water}}$$

$$\text{Gas SG} = \frac{\text{density of gas}}{\text{density of air}}$$

Care must be taken to define standard conditions under which the densities or specific gravities are measured, so as not to introduce errors due to variations in measurement conditions. Commonly accepted sets of conditions are Normal Temperature and Pressure (NTP) and Standard Temperature and Pressure (STP). NTP is usually used for solids at the temperature 0°C and a pressure of 760 mm mercury. The STP is used for solids and fluids at a temperature of 15.6°C and a pressure of 1 atm (or 101.325 kPa).

Density measurements are a significant part of instrumentation systems. In many processes, the density is taken as the controlling parameter for the rest of the process; therefore, accurate measurements are necessary. Density measurements are made for at least two important reasons: (1) to determine the mass and volume of products, and (2) to assess the quality of products. In many industrial applications, density measurement ascertains the value of the product. A list of manufacturers offering different types of density meters is provided in Table 1.

Table 1. List of Manufacturers

ABB K-Flow Inc Drawer M Box 849 Millville, NJ 08332 Tel: 800-825 3569	McGee Engineering Co., Inc. Tujunga Canyon Blvd. Tujunga, CA 91042 Tel: 800-353 6675
American Density Materials Inc. Rd 2, Box 38E Belvidere, NJ 07823 Tel: 908-475 2373	Porous Materials, Inc. Cornell Business & Technology Park Ithaca, NY 14850 Tel: 800-825 5764
Anton Paar USA 10201 Maple Leaf Court Ashland, VA 23005 Tel: 800-722-7556	Princo Instruments Inc 1020 Industrial Hwy., Dept L Southampton, PA 18966-4095 Tel: 800-496 5343
Arco Instrument Company, Inc. 1745 Production Circle, Riverside, CA 92509 Tel: 909-788 2823 Fax: 909-788 2409	Quantachrome Corp. 1900-T Corporate Drive Boynton Beach, FL 33426 Tel: 800-966 1238
Cambridge Applied Systems, Inc. 196 Boston Avenue, Medford, MA 02155 Tel: 617-393 6500	Tricor Systems, Inc. 400-T River Ridge Rd. Elgin, IL 60123 Tel: 800-575 0161
Dynatron Automation Products, Inc. 3032 Max Roy Street Houston, TX 77008 Tel: 800-231 2062 Fax: 713-869 7332	X-rite, Inc. 3100-T 44th St. S. W Grandville, MI 49418 Tel: 800-545 0694
Kay-Ray/Sensall, Fisher-Rosemount 1400 Business Center Dr. Mount Prospect, IL 60056 Tel: 708-803 5100 Fax: 708-803 5466	

In many modern applications, densities are obtained by sampling techniques. However, there are two basic approaches: static density measurements and dynamic (on-line) density measurements. Within each concept, there are many different methods available, depending on physical principles and process characteristics. In many cases, application itself and process conditions determine the best suitable method to be employed. Generally, static methods are well developed, lower in cost, and more accurate. Dynamic samples are expensive, highly automated, and use microprocessor-based signal processing devices. Nevertheless, today, many static methods are also computerized, offering easy to use, flexible, and self calibrating features.

There is no single, universally applicable density measurement technique available. Different methods must be employed for solids, liquids, and gases. Here, some of these measurement techniques and the devices involved are introduced in the relevant sections of solid, liquid, and gas density measurements. Further detailed discussions are given for specific types of instruments.

DENSITY AND DENSITY MEASUREMENT OF SOLIDS

Theoretically, the density of solids may vary both with operating pressure and temperature. However, in many applications, the effect of pressure may be neglected in solids and incompressible fluids. In some cases, the effect of temperature can also be neglected if the operational temperature is not significantly higher from the temperature at which density measurements are made, or a high degree of accuracy is not required.

The volumetric effect of temperature on the density of liquids or solids may be expressed as

$$V = V_0(1 + \beta \Delta t) \quad (1)$$

where β is the coefficient of expansion of solid or liquid which is consistent with the temperature units used.

As the mass is same before and after temperature rise, the change in density is inversely proportional to the change in volume and can be expressed as

$$\rho/\rho_0 = V/V_0 \quad (2)$$

If the solid sample has a regular shape and uniform, the determination of its density is a simple task. Once the volume and mass of the solid are known, the density may be found by using the basic ratio: density = mass/volume (kg/m^3). In order to avoid errors, the weights and volumes must be determined by using accurate instruments.

However, in many applications, solids have different constituents and are made up from the mixture of different materials. The volumetric ratios of constituents may also change. A common method of determining the density of irregular and nonuniform samples is the hydrostatic weighting. In some cases, dynamic methods are employed, such as radioactive absorptions and ultrasonic methods.

Powdered solids occlude air between or inside individual particles, giving rise to apparent, bulk, tap, effective, and true densities. The apparent density includes the air lodged in the cavities or pores, and the density is determined without filling up the pores. The numerical value of the density depends on

the amount of compacting employed (tap density) which can be achieved by moderate mechanical means and also embedded foreign particles (effective density). For true density, it is necessary to dislodge the air by means of suitable liquids or gases. Special pycnometers are developed for this purpose.

DENSITY AND DENSITY MEASUREMENT OF LIQUIDS

Overall density of fluid is the ratio of total mass to total volume. Point density is the ratio of molecular mass in a volume element centered at a point to the element's volume, the volume being much smaller than the total volume. As in the case of solids, the densities of liquids are affected by temperature and pressure. Most liquids are incompressible; therefore, pressure effects may be neglected. Nevertheless, in determination liquid densities, the effects of temperatures must carefully be monitored as indicated in equations (1) and (2).

The measurements of densities of fluids are much more complex than solids; therefore, there are many different techniques developed. Hydrometers, pycnometers, hydrostatic weighing, flotation methods, drop methods, radioactive methods, optical methods, and so on are typical examples of measuring liquid densities.

Difficulties in the measurement of densities of fluids is due to complexities in processes, variations of fluid densities within the process, and the diverse characteristics of the process and fluids themselves. Some of these methods are custom designed and applicable to special cases only. Others are very similar in principles and technology, and can be used for many different type of fluids. At the present, apart from conventional methods, many novel and unusual techniques are reported to be undergoing extensive development and research stages. For example, densitometers based on electromagnetic principles may be given as a part of intelligent instrumentation systems.

Depending on the application, fluid densities can be measured both in *static* or *dynamic* forms. In general, static density measurements of fluids are well developed, precise, and have greater resolution than most dynamic techniques. Pycnometers and buoyancy are examples of static techniques that can be adapted to cover small density ranges with a high resolution and precision. Nowadays, many manufacturers are offering dynamic instruments previously known to be static. Also, many static density measurement devices are computerized, coming with appropriate supporting hardware and software. In general, static type measurements are employed in laboratory conditions, and dynamic methods are employed for real time measurements where properties of fluids vary from time to time.

DENSITY AND DENSITY MEASUREMENT OF GASES

The density of a gas will vary significantly with absolute pressure. Increasing the pressure of a gas at a constant temperature causes the gas to be compressed to a smaller volume of the same mass, thus increasing the density. Boyle's Law states that for ideal gases or a mixture of ideal gases at a constant temperature, the volume, V , is inversely proportional to the absolute pressure; that is

$$V = \text{Constant}/P \quad (3)$$

The following formula may be written for comparing the volumes of an ideal gas at a constant temperature but at different pressures, P :

$$V/V_0 = P_0/P \quad (4)$$

Charles' Law states that the density of gas will vary significantly with absolute temperature, T . Increasing the temperature of a gas at constant pressure causes the gas molecules to increase their activities and motions in relation to each other, whereby increasing the volume and decreasing the density of gas for the same mass. Charles' Law may be stated in the following form:

$$V/V_0 = T/T_0 \quad (5)$$

Charles' and Boyle's Laws can be combined to yield the Ideal Gas Law as

$$PV = nRT \quad (6)$$

where R is the universal gas constant in consistent units, and n is the number of moles.

The Ideal Gas Law can also be expressed in the following form:

$$V/V_0 = TP_0/T_0P \quad (7)$$

During gas density measurements, when variations in pressure and temperature are small, the temperature and pressure act almost independently of each other; thus, estimates of reasonable accuracy can be obtained by adding percentage temperature and pressure deviations from a given set of conditions.

Often, density measurements of Non-Ideal Gases are required which do not act as ideal gases at certain conditions, such as at high pressures, low temperatures, or under saturation. Their non-ideal behavior may be accounted for by modifying the Ideal Gas Law with a Z factor, as

$$V/V_0 = TP_0Z/T_0PZ_0 \quad (8)$$

The Z factor is numerically dependent on operating conditions and can be read from generalized compressibility charts, as shown in Fig. 1, with a reasonable degree of accuracy.

Due to the above reasons, extra care and further considerations are necessary in gas density measurements. For example, perfect gases contain equal numbers of molecules under same conditions and equal volumes. Therefore, molecular weights may be a better option in density measurements. Flask methods, gas balance methods, optical methods, x-ray methods are typical techniques employed for gas density measurements.

MAGNETIC AND VIBRATIONAL METHODS

Magnetic Method

This method is used for both liquids and gases. It allows the determination of effects of pressures and temperatures down to cryoscopic range. Basically, the device contains a small ferromagnetic cylinder, encased in a glass jacket. The jacket and ferromagnetic material combination constitutes a buoy

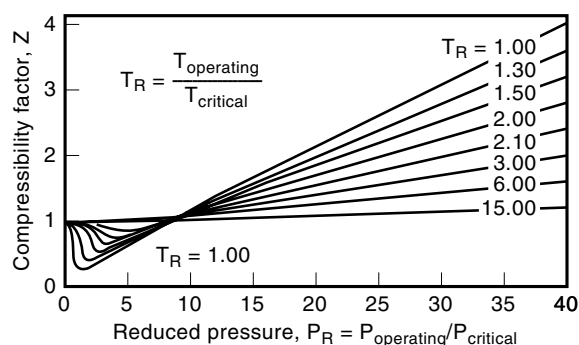


Figure 1. Generalized gas compressibility chart. In many applications, densities of Nonideal Gases are required. Nonideal gases do not act as ideal gases at certain conditions, such as at high pressures, low temperatures, or under saturation. Their behavior may be accounted for by modifying the Idea Gas Law with a Z factor. The Z factor is numerically dependent to different operating conditions and can be read from generalized compressibility charts such as this one.

or float. The cylinder is held at a precise height within the medium by means of solenoid which is controlled by a servo system integrated with a height sensor. The total magnetic force on the buoy is the product of the induced magnetic moment by the solenoid and the field gradient in vertical direction. The total magnetic force at a particular distance in vertical direction in the solution compensates the difference in the opposing forces of gravity (downwards) and buoyancy (upwards) exerted by the medium, through the Archimedes' principle. The magnetic force is directly proportional to the square of the current in the solenoid. If the buoyant force is sufficient to make the ferromagnetic assembly float on the liquids of interest, the force generated by the solenoid must be downward to add to the force of gravity for equilibrium.

By the use of precision resistors and accurate differential voltmeters, the measurements simply consist of reading and recording voltages. Calibrated curves obtained from standard solutions of known densities must be used for a given buoy, to be referred to solutions under test. Buoys cover a limited range of density; hence, many buoys must be provided to cover large density ranges. The complete measurement of a sample requires about 5 minutes for insertion, temperature, and buoy stabilization. The sensitivity of the devices is in the region of ± 0.01 mV with a tight temperature control. Their uses can also be extended into applications such as the specific gravity measurements under low pressures and density measurements of hazardous fluids.

In some instruments, a smaller buoy is used containing tiny permanent magnets. By this way, the sample volume and errors from adhering bubbles are reduced, and the current in the solenoid is linearized as a function of density. This provides a greater precision in calibrations. In other instruments, a rotating magnetic field is superimposed on a standard device via remotely situated field coils in such a way that controlled rotation of the immersed buoy can be accomplished. Since the rate of rotation is proportional to the viscosity of the medium, the density and viscosity are determined simultaneously.

Vibrational Methods

Devices based on vibrations are used for liquid and gas density measurements. They make use of the changes in natural

frequency of vibration of a body containing fluid in it or surrounded by it. The natural frequency of the vibrating body is directly proportional to the stiffness and inversely proportional to the combined masses of the body and the fluid. It also depends on the shape size and elasticity of materials, induced stresses, and the total mass and mass distribution of the body. Basically, the vibrations may be equated to the motion of a mass attached to a mechanical spring. Hence, expression for the frequency may be written as

$$\text{Resonant frequency} = \sqrt{(K/(M + k\rho))} \quad (9)$$

where K is the system stiffness, M is the transducer mass, k is the system constant, and ρ is the fluid density.

A factor common to all types of vibrating element densitometers is the problem of setting the element in vibration and maintaining its natural resonance. There are two drives available for this purpose, these being the magnetic drives and the piezoelectric drives.

Magnetic Drives. Magnetic drives of the vibrating element and the pick up sensors of vibrations are usually achieved using small coil assemblies. Signals picked up by the sensors are amplified and fed back for maintaining the disturbing forces on the vibrating body.

In order to achieve steady drives and minimize external effects, vibrating elements may be made from nonmagnetic materials. In this case, small magnetic armatures are attached to maintain vibrations.

The main advantage of the magnetic drive and pick up system is that they are noncontact methods. They use conventional copper windings, are simple to construct, and are reliable within the temperatures of -200 to $+200^\circ\text{C}$.

Piezoelectric Drives. The piezoelectric drives are mechanically fixed on the vibrating element by adhesives. Therefore, attention must be paid for the careful placement of the mount in order to reduce the strain experienced by the piezo elements, due to thermal and pressure stresses while the instrument is in service. A wide range of piezoelectric materials are available to meet the requirements. These materials demonstrate good temperature characteristics as in magnetic drive types. They have high impedances, thus making the signal conditioning circuitry relatively easy and cost effective.

A number of different types of densitometers are developed which utilize properties of vibrating bodies in liquids and gases. The three main commercial types are introduced here. In all cases, vibrations are maintained either through magnetic drives or piezoelectric drives. In recent years, the coriolis effect is used for density measurements yielding a new type of density meters.

In general, vibrating element meters have advantages and disadvantages. The advantages are:

1. They are suitable for both liquid and gas density measurements under static and dynamic conditions, with reasonable accuracy.
2. Real-time measurements may easily be interfaced to computers since they operate on frequencies and are inherently digital.
3. They are relatively robust and easy to install.

4. Programmable and computerized versions are available. Programmable versions make all the corrections automatically. They provide the output of process density, density at reference conditions, relative density, specific gravity, concentration, solid contents, etc.

The main disadvantage is that they do not relate directly to primary measurements; therefore, they have to be calibrated. They also have problems of measuring densities of multi-phase fluids.

Vibrating Tube Densitometers

These devices are suitable in highly viscous liquids or slurry applications. The mode of operation of vibration tube meters is based on the transverse vibration of single or twin tubes as shown in Fig. 2. The tube and the driving mechanisms are constrained to vibrate on a single plane. As the liquid moves inside the tube, the density of the entire mass of liquid is measured. The tube length is approximately 20 times greater than the tube diameter.

One major design problem with the vibrating tube method is the conflict to limit the vibrating element to a finite length, and also the accuracy of fixing the nodes. Special attention must be paid to avoid any exchange of vibrational energy outside the sensory tube. The single tube has the disadvantage of presenting obstructing to the flow, thus having some pressure losses. The twin tubes, on the other hand, offer very small blockage, and they can easily be inspected and cleaned. Their compact sizes are another distinct advantage. In some densitometers, the twin tubes are designed to achieve a good dynamic balance with the two tubes vibrating in anti-phase. Their nodes are fixed at the ends, demonstrating maximum sensitivity to installation defects, clamping, and mass loadings.

The main design problems of the vibrating tube sensors are in minimizing the influence of end padding and overcoming the effects of pressure and temperature. Bellows are used at either ends of the sensor tubes to isolate the sensors from external vibrations. Bellows also minimize the end loadings due to differential expansions and installation stresses.

The fluid runs through the tubes; therefore, it does not need pressure balances. Nevertheless, in some applications, the pressure stresses the tubes, resulting in stiffness changes.

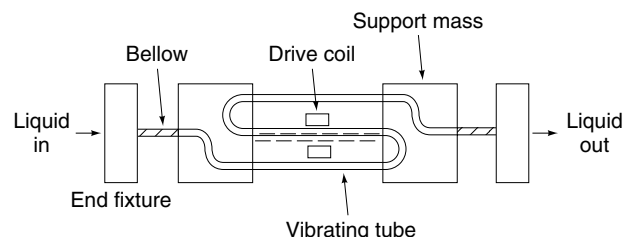


Figure 2. Vibrating tube densitometer. A tube containing fluids is vibrated at resonant frequency by electromagnetic or piezoelectric vibrators. The tube and the driving mechanisms are constrained to vibrate on a single plane. The resonant frequency is a function of the density of the fluid moving inside the tube. The tube is isolated from the fixtures by carefully designed bellows. These devices are not suitable in highly viscous liquids or slurry applications. The tube length is approximately 20 times greater than tube diameter.

Some manufacturers modify the tubes to minimize the pressure effects. In these cases, corrections are necessary only when high accuracy is needed. The changes in the Young's modulus with temperature may be reduced to almost zero by using Ni-span-C materials whenever corrosive properties of fluids permit. Usually, manufacturers provide pressure and temperature correction coefficients for their products.

It is customary to calibrate each vibration element densitometer against other methods as a transfer of standards. Often, the buoyancy method is used for calibration purposes. The temperature and pressure coefficients are normally found by exercising the transducer over a range of temperatures and pressures on some liquid of well known properties. Prior to calibration, the vibration tube densitometers are subjected through a programmed burn-in cycle for stabilization against temperatures and pressures.

Vibrating Cylinder Densitometers

A thin walled cylinder, with a 3 to 1 ratio of length versus diameter, is fixed with stiff ends. The thickness of the wall of the cylinder varies from 25 to 300 μm , depending on the density range and type of fluid used. The cylinder can be excited to vibrate in a hoop mode by magnetic drives mounted either in or outside the cylinder.

For good magnetic properties, the cylinder is made from corrosion-resistant magnetic materials. Steel such as FV520 is used for the purpose. Such materials have good corrosion resistance characteristics. Unfortunately, due to their poor thermoelectric properties, they need extensive temperature corrections. Often, nickel iron alloys such as Ni-span-C are used to avoid temperature effects. Especially since once correctly treated, the Ni-span-C alloy has near zero Young's modulus properties. Since the cylinder is completely immersed in the fluid, there are no pressure coefficients.

The change in the resonant frequency is determined by the local mass loading of the fluid in contact with the cylinder. The curve of frequency against density is nonlinear and has a parabolic shape, thus requiring linearization to obtain practical outputs. The resonant frequency range varies from 2 to 5 kHz depending on the density range of the instrument. The cylinders need precision manufacturing and, thus, are very expensive to construct. Each meter needs to be calibrated individually for different temperatures and densities to suit specific applications. In the case of gas density applications, gasses with well known properties such as pure argon or nitrogen are used for calibrations. In this case, the meters are subjected to gas environment with controlled temperature, and pressure and calibration curves are achieved by repetitions to suit requirements of individual customers for their particular applications. In the case of liquids, the meters are calibrated with liquids of known density, or they are calibrated against another standard such as pycnometer or buoyancy type densitometers.

Vibration cylinder type densitometers have zero pressure coefficients and they are ideal for liquidised gas products or refined liquids. Due to relatively small clearances between cylinder and housing, they require regular cleaning. They are not suitable with liquids or slurries with high viscous properties.

Tuning Fork Densitometers

These densitometers make use of the natural frequency of a

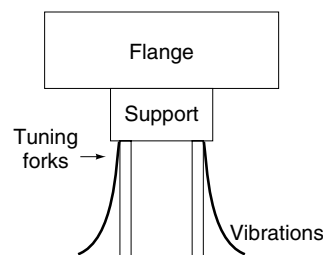


Figure 3. Tuning fork densitometer. Twin forks are inserted into the liquid or gas media whose density needs to be measured. Since the natural frequency of the forks is a function of the density of the media, small changes in the natural frequencies must be monitored accurately. Calibration is necessary in each application.

low mass tuning forks, shown in Fig. 3. In some cases, the liquid or gas is taken into a small chamber in which the electromechanically driven forks are situated. In the other cases, the fork is inserted directly into the liquid. Calibration is necessary in each application.

Coriolis Densitometers

Another type of vibrational density meter is based on the coriolis principle. The device is similar in vibrating tube methods with slight variations in physical design. They are comprised of sensors and a signal processing transmitters. Each sensor consists of one or two flow tubes enclosed in a sensor housing. The sensor tubes are fixed at one end and free at the other end, shown in Fig. 4. The sensor operates by applying Newton's Second Law of motion ($F = ma$).

Inside the housing, the tubes are vibrated at their natural frequencies using drive coils and a feedback circuit. This resonant frequency of the assembly is a function of the geometry of the element, materials of construction, and mass of the tube assembly. The tube mass is comprised of two parts: the mass of the tube itself and the mass of the fluid inside the tube. The mass of the tube is fixed for a given sensor. The mass of fluid in the tube is equal to the fluid density multiplied by the volume. Since the tube volume is constant, the frequency of oscillation can be directly related to the fluid density. There-

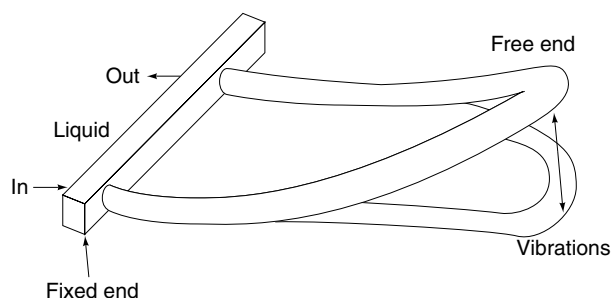


Figure 4. Coriolis densitometer. Each sensor consists of one or two flow tubes enclosed in a sensor housing. Inside the housing, the tubes are vibrated at their natural frequencies using drive coils and a feedback circuitry. This resonant frequency of the assembly is a function of the geometry of the element, material of construction, and mass of the tube assembly. Vibration of the tube is detected and related to the mass and flow rate of the fluid. They are manufactured in various shapes and sizes.

fore, for a given geometry of a tube and the material of the construction, the density of the fluid can be determined by measuring the resonant frequency of vibration. Temperature sensors are used for overcoming the effects of changes in the modulus of elasticity of the tube. The fluid density is calculated using a linear relationship between the density and the vibrations of the tube and calibration constants.

Special peripherals, based on microprocessors, are offered by manufacturer for a variety of measurements. However, all density peripherals use the natural frequency of the sensor coupled with the sensor temperature to calculate the on-line density of the process fluid. Optional communication, interfacing facilities, and appropriate software are also offered. These devices come in different shapes and sizes.

WEIGHT/WEIGHING SYSTEM/MASS MEASUREMENT METHODS

Pycnometric Densitometers

Pycnometers are static devices used for measuring densities of liquids and gases. They are manufactured as fixed volume vessels which can be filled with sample fluids. The density of the fluid is measured by weighing the sample with the vessel. The simplest version consists of a glass vessel in the shape of a bottle with a long stopper containing a capillary hole as shown in Fig. 5. The volume and variation of volume with temperature has been accurately determined. The capillary is used to determine the exact volume of the liquid, thus giving high resolution when filling the pycnometer. If the weight of the empty pycnometer is W_1 and the weight of the pycnometer, when containing a volume V of liquid at temperature t , is W_2 , the density of liquid ρ_1 may be calculated directly from

$$\rho_1 V = W_2 - W_1 \quad (10)$$

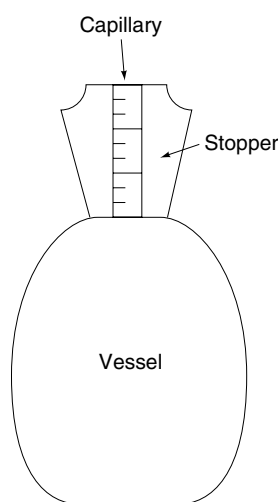


Figure 5. Pycnometer. A fixed volume container is filled with liquid and weighed accurately. The capillary is used to determine the exact volume of the liquid, thus giving high resolution when filling the pycnometer. For accuracy and good precision in results, the bottle must be cleaned after each measurement, the temperature must be kept constant, and precision balances must be used. It is necessary to apply accurate buoyancy and thermal expansion corrections.

For accuracy and good precision in results, ultimate care must be exercised during measurements. That is, the bottle must be cleaned after each measurement, the temperature must be kept constant, and precision balances must be used. For precise measurements, it is necessary to apply buoyancy and thermal expansion corrections. In some cases, to ensure filling of the pycnometer, twin capillary tubes are used. The two capillaries, made in glass, are positioned such that fluid can be driven into the vessel under vacuum conditions. Accurate filling to graduation marks on capillary is then made.

The pycnometers have to be light weight, strong enough to contain samples, and nonmagnetic for accurate weighing to eliminate possible ambient magnetic effects. Very high resolution balances have to be used to detect small differences in weights of gases and liquids. Although many pycnometers are made of glass, they are also made in metals to give enough strength for the density measurements of gases and liquids at extreme high pressures. In many cases, metal pycnometers are necessary for taking samples from the line of some rugged processes.

Pycnometers are used, Fig. 6, for measuring density of solid substances, such as carbon black, cement, fibers, ceramics, charcoals, fertilizers, pharmaceutical and powdered materials, *etc.* The density is computed from the mass difference between the pycnometer filled with sample together with a displacer fluid which penetrates into pores of the solid, and with fluid only. In some cases, mercury is used as the displacing liquid. The accuracy increases with the increasing quantity of solid which can be used. In powder technology, pycnometers are used to measure the true volume of the solid materials by employing Archimedes' principle of fluid displacement. The displaced fluid is a gas which can penetrate the finest pores to assure maximum accuracy. Usually, helium is used because of its small atomic dimension, thus penetrating into crevices and pores effectively.

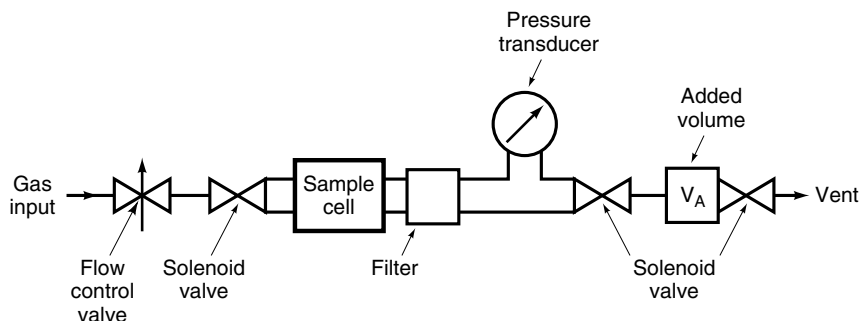
The pycnometers have advantages and disadvantages. The advantages are that if used correctly, they are accurate and can be used for both density and specific gravity measurements. The disadvantages are

1. Great care must be exercised for accurate results
2. The sample has to be taken off-line with consequent time lag in results. This creates problems of relating samples to the materials that exist in the actual process
3. High precision pycnometers are expensive. They require precision weighing scales and controlled laboratory conditions. Specialized techniques must be employed to take samples in high pressure processes and hostile conditions such as offshore installations.

Hydrostatic Weighing Densitometers

These devices are suitable for solid and liquid density measurements only. The density of a solid is often measured by weighing it first in air, and afterwards in a suitable liquid of known density. The latter weighing is done by suspending the solid under the pan of a precision balance by means of a very thin wire. A typical commonly used example of such devices is the Westphal balance, as shown in Fig. 7. If the weight of the plummet in air is W_a , its volume, V , and its weight when suspended in liquid,

Figure 6. Use of a pycnometer for porous or powdered solids. Pycnometers are used for measuring density of solid substances, such as carbon black, cement, fibers, ceramics, charcoals, fertilizers, pharmaceutical, and powdered materials, and so on. In powder technology, pycnometers are used to measure the true volume of the solid materials by employing Archimedes' principle of fluid displacement. The displacing fluid is usually helium gas which can penetrate the finest pores to assure maximum accuracy. In some cases, water, mercury, or other fluids are used as displacers.



W_1 , the equilibrium of vertical forces on the plummet gives

$$W_1 + \rho_1 V - W_a = 0 \quad (11)$$

from which the density ρ_1 at the temperature t may be calculated.

The inaccuracies of this type of device are mainly limited by the irregularity of meniscus around the wire, particularly in the case of water and aqueous solutions. The wetting of the wire may be improved by covering it electrolytically with platinum black.

A common device using hydrostatic weighing of liquids consists of a U-tube which is pivoted on flexible end couplings. A typical example is shown in Fig. 8. The total weight of the tube changes, depending on the density of fluid flowing through it. The change in the weight needs to be measured accurately, and there are a number of methods employed for it. The most common commercial meters use a force balance system. The connectors are stainless steel bellows. In some cases, rubber or other materials are used, depending on the process fluid characteristics and the accuracy required. There are temper-

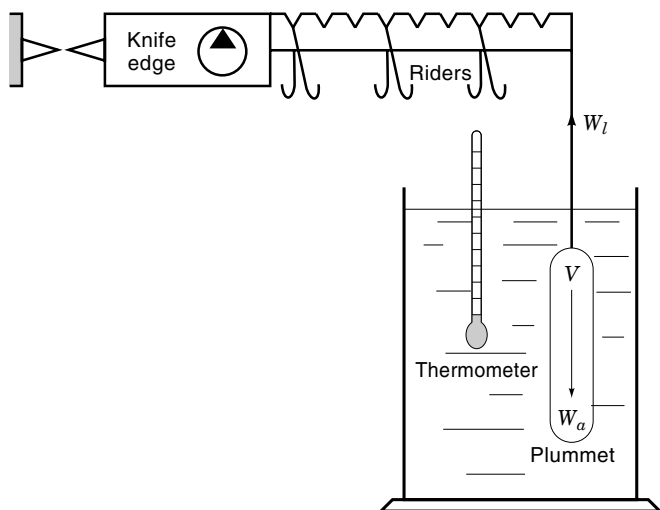


Figure 7. Hydrostatic weighing densitometers. The density of a solid is often measured by weighing it first in air, afterwards in a suitable liquid of known density. The latter weighing is done by suspending the solid under the pan of a precision balance by means of a very thin wire. A typical commonly used example of such devices is the Westphal balance. Riders are used for precision measurements. Accurate temperature measurement is necessary.

ature, flow rate, and pressure limitations due to bellows, and the structure of the system may lead to a reading offset. The meter must securely be mounted on a horizontal plane for best accuracy.

The hydrostatic weighing methods of liquids give continuous readings for two phase liquids such as slurries, sugar solutions, powders, etc. They are rugged, give accurate results, and are used for the calibration of the other liquid density transducers. However, they must be installed horizontally on a solid base; hence, they are not flexible enough to adapt for any process. Thus, the process must be designed around it.

Balance Type Densitometers

Balance type densitometers are based on gravity and/or weighing principles, and they are suitable for liquid and gas density measurements. Manufacturers offer many different types; four most commonly used ones are discussed below.

Balanced-Flow Vessel. A fixed volume vessel, as shown in Fig. 9, is employed for the measurements. While the liquid or gas is flowing continuously through the vessel, the total assembly is weighed automatically by a sensitive scale, a spring balance system, or a pneumatic force balance transmitter. Since the volume and the weight of the fluid are known, the density

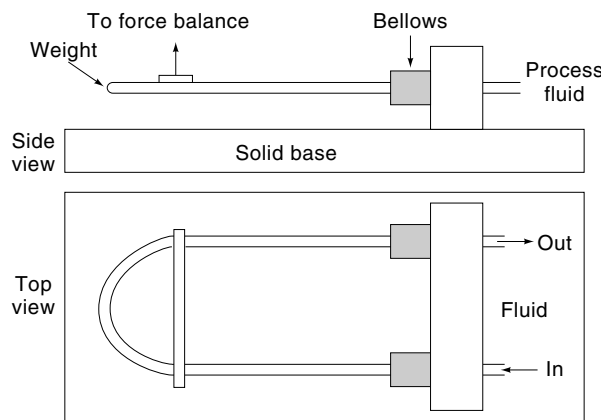


Figure 8. Hydrostatic weighing. This device consists of a U-tube which is pivoted on flexible end couplings. The total weight of the tube changes depending on the density of fluid flowing through it. The change in the weight needs to be measured accurately. There are temperature, flow rate, and pressure limitations due to bellows and the structure of the system may lead to a reading offset. The meter must securely be mounted on a horizontal plane for best accuracy.

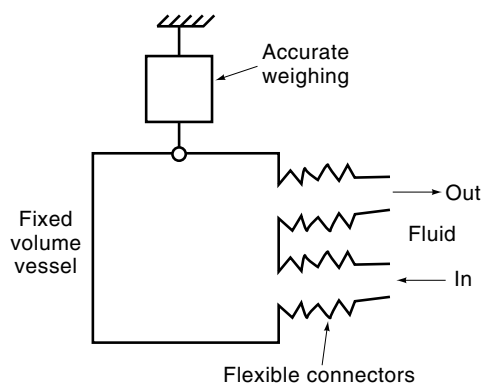


Figure 9. Balanced-flow vessel. Liquid or gas flows through a fixed volume vessel. The vessel is weighed continuously by a sensitive scale while the fluid is flowing through it. Since the volume and the weight of the fluid are known, the density or specific gravity can be calculated and scaled in respective units.

or specific gravity can easily be calculated and scaled in respective units. In the design process, extra care must be exercised for the flexible end connections.

Buoyancy Hydrostatic-Weighing Methods. The buoyancy method basically uses Archimedes' principle. A displacer with a known mass and volume is immersed in the liquid whose density is to be measured. Position of the displacer is kept constant by bellows, as exemplified in Fig. 10. Once the mass, the volume, the displacer, the bellow position, and pressure are known, the density of the liquid can be calculated. However, some corrections need to be made for cubicle expansion coefficients of the displacer and the temperature of the process. Buoyancy type densitometers give accurate results, and they are used for the calibration of the other liquid density transducers.

Chain Balanced Float. In this system, a self centering fixed volume submerged plummet is used for density measurements, as illustrated in Fig. 11. The plummet is located entirely under the liquid surface. At balance, the plummet operates without friction, assuming a stable position based on mass balance tech-

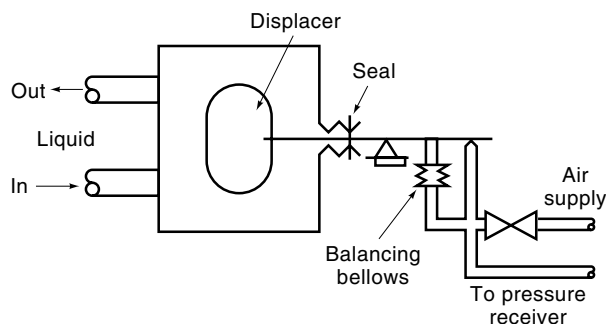


Figure 10. Buoyancy hydrostatic-weighing method. A displacer with a known mass and volume is immersed in the liquid whose density is to be measured. The displacer is kept in position by sensitive balancing bellows. Buoyancy type densitometers give accurate results, and they are used for the calibration of the other liquid density transducers.

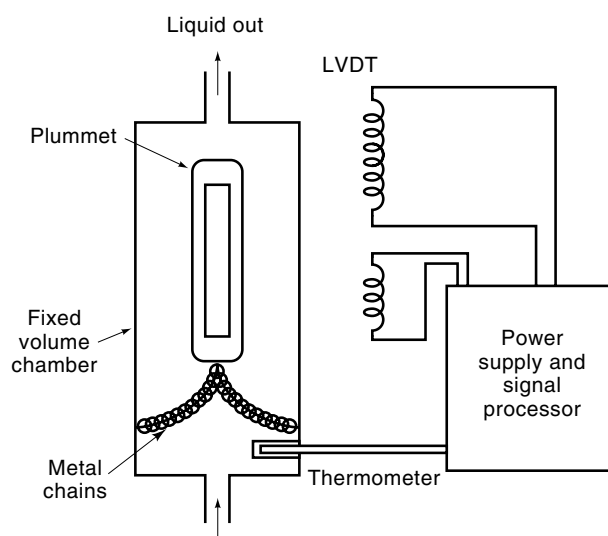


Figure 11. Chain balanced float. The fixed volume and weight plummet totally suspended in the liquid assumes equilibrium position depending on the density. The force exerted by the chains on the plummet is a function of plummet position; hence, the measured force is proportional to the density of the liquid. Any change in the density process causes the plummet to move to a new equilibrium point. In some cases, the plummet contains a metallic transformer core which transmits changes in the position to be measured by a pick up coil or a linear variable transformer (LVDT).

nique. The effective weight of the chain on the plummet varies, depending on the position of the plummet which, in turn, is a function of the density of the liquid. Any change in the density of the process causes the plummet to move to a new equilibrium point by transferring the chain weight between the reference point and the plummet. In some cases, the plummet contains a metallic transformer core which transmits changes in the position to be measured by a pick up coil or a linear variable transformer (LVDT). When LVDTs are used, the voltage differential, a function of plummet displacement, is calibrated as a measure of variations in density or specific-gravity. A resistance thermometer bridge is used for the compensation of temperature effects on density. The range of the instrument can be from 0.5 g/mL to 3.5 g/mL. Their accuracy is in the region of $\pm 3\%$ of the range. Viscosity limits (50 centipoise) and flow rates of the liquid (2 to 3 L/min) are the main limitations.

Gas Specific Gravity Balance. A tall column of gas is weighed by a float located at the bottom of the vessel. This weight is translated into the motion of an indicating pointer which moves over a scale graduated in units of density or specific gravity. This method can be employed in the density measurements of many gases.

Buoyancy Gas Balance. In this instrument, a displacer is mounted on a balance beam in a vessel, shown in Fig. 12. The displacer is balanced for air, and the manometer reading is noted at the exact balance pressure. The air is then displaced by gas, and the pressure is adjusted until the same balance is restored. The ratio of the pressure of air to pressure of gas is then the density of gas relative to air. This method is commonly applied under laboratory conditions and is not suitable

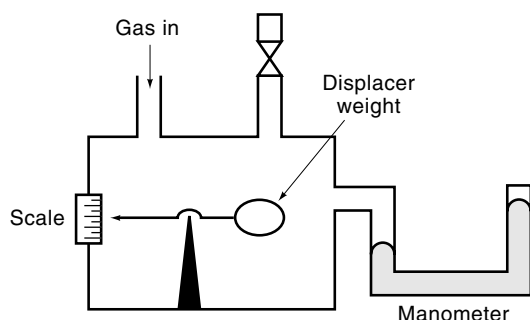


Figure 12. Buoyancy gas balance. Position of the balance beam is adjusted by a set pressure air; air is then displaced by gas of the same pressure. The difference in the reading of the balance beam gives the specific gravity (SG) of the gas. The pressures are read on the manometer. This method is commonly applied in laboratories and not suitable for continuous measurements.

for continuous measurements.

REFERENCE METHODS

Hydrometers

Hydrometers are direct reading instruments, the most commonly used devices for measurement of density of liquids. They are so commonly used that their specifications and procedure of use are described by national and international standards such as ISO 387. The buoyancy principle is used as the main technique of operation. Almost all hydrometers are made from high grade glass tubing. The volume of fixed mass is converted to a linear distance by a sealed bulb shaped glass tube containing a long stem measurement scale, shown in Fig. 13. The bulb is ballasted with a lead shot and pitch, the mass of which is dependent on the density range of the liquid to be measured. The bulb is simply placed into the liquid, and the density is read from the scale. The scale may be graduated in density units such as kg/m^3 . Hydrometers can be calibrated for different ranges for surface tensions and temperatures. Temperature corrections can be made for set temperature such as 15, 20, 25°C. ISO 387 covers a density range of 600 to 2000 kg/m^3 .

Hydrometers may be classified according to the indication provided by graduations of the scale: density hydrometers, specific gravity hydrometers, percentage hydrometers showing the percentage of solution, for example, sugar, and arbitrary scale hydrometers. Customized scales are also available, for example, lactometers for testing milk, alcoholometers for alcohol levels, *etc.* Many other alternative scales are offered by manufacturers such as specific gravity, API gravity, Brix, Brine, *etc.*

The best way to read a hydrometer in clear liquids is to start with the eyes slightly below the plane of the liquid surface, and slowly raise the eyes until the surface of liquid appears as a straight line. The place where the line crosses the scale is the reading. With opaque liquids, such as oils, it is necessary to read the hydrometer at the top of the meniscus. For accurate readings, the stem must be absolutely clean. Also, the surface of the liquid must be clean and free of dust. With precision grade hydrometers, with long small diameter stems, density values may be read to 0.0001. In general use of hydrometers,

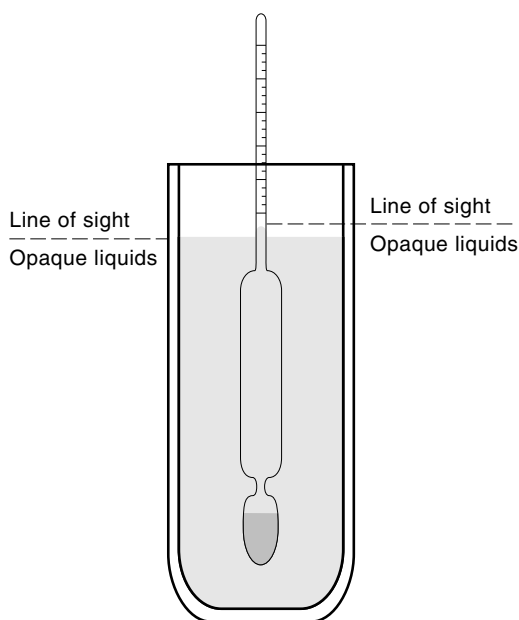


Figure 13. Hydrometer. A fixed weight and volume bulb is placed into the liquid. The bulb sinks in the liquid depending on its density. The density is read directly from the scale. They are classified according to the indication provided by graduations of the scale. The best way to read a hydrometer in clear liquids is to start with the eyes slightly below the plane of the liquid surface, and slowly raise the eyes until the surface of liquid appears as a straight line. The place where the line crosses the scale is the reading. With opaque liquids, such as oils, it is necessary to read the hydrometer at the top of the meniscus.

the uncertainty of readings may be in the region of ± 0.01 .

Hydrometers are low cost devices and are easy to use with a good resolution. However, they have a number of disadvantages, such as

1. They have small span; therefore, to cover a significant range, a number of hydrometers are required.
2. They are made from glass and are fragile. Metal and plastic versions are not as accurate.
3. The fluid needs to be an off-line sample, not representing the exact conditions of the process. There are pressure hydrometers for low vapour pressure hydrocarbons, but this adds a need for accurate determination of pressure.
4. If good precision is required, they are difficult to use, needing surface tension and temperature corrections. Further corrections may be required for opaque fluids.

Column Type Densitometers

These devices are used for liquid density measurements. There are a number of different versions of column methods. As a typical example, a reference column method is illustrated in Fig. 14. A known head of sample liquid and water from their respective bubbler pipes are used. A differential pressure measuring device compares the pressure differences, proportional to relative densities of the liquid and the water. By varying the depth of immersion of the pipes, a wide range of measurement may be obtained. Both columns must be maintained at the same temperature in order to avoid the necessity for corrections of

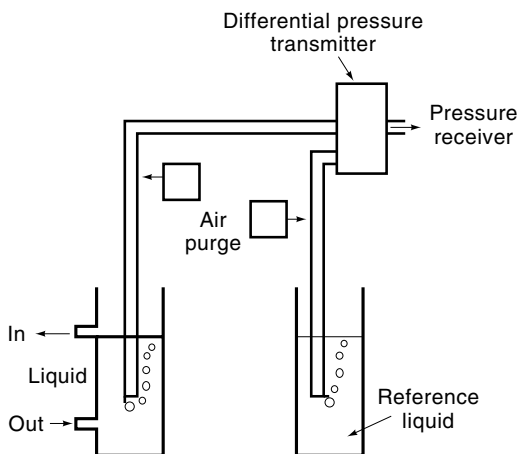


Figure 14. Reference column densitometer. Two identical tubes, having the same distance from the surface, are placed in water and liquid. Water with known density characteristics is used as a reference. The pressures necessary to displace the fluids inside the tubes are proportional to the densities of the fluids. By varying the depth of immersion of the pipes, a wide range of measurement may be obtained. Both columns must be maintained at the same temperature.

temperature effects.

A simpler and the most widely used method of density measurement is achieved by the installation of two bubbler tubes as illustrated in Fig. 15. The tubes are located in the sample fluid such that the end of one tube is higher than that of the other. The pressure required to bubble air into the fluid from both tubes is equal to the pressure of the fluid at the end of the bubbler tubes. The openings of the tubes are fixed; hence, the difference in the pressure is the same as the weight of a column of liquid between the ends. Therefore, the differential pressure measurement is equivalent to the weight of the constant volume of the liquid. Calibrations can be made with direct relationship to the density of the liquid. This method is accurate to within 0.1 to 1% specific gravity. It must be used with liquids that do not crystallize or settle in the measuring chamber.

U-Tube Method

The unknown density of a liquid may be obtained from the known density of another liquid by placing them in a U-tube as in Fig. 16. By measuring the lengths of liquid columns, l_1 and l_2 , and using the manometer principles, the following equation may be written

$$\rho_1 l_1 = \rho_2 l_2 \quad (12)$$

thus allowing ρ_1 to be calculated. In this method, the liquids must not be miscible. This method gives a quick idea about the density of the liquid and is not usually accurate because the various menisci prevent accurate measurements of the lengths of the liquid columns.

RADIOACTIVE METHODS

Radioactive density measurements are suitable for both liquids and solids undergoing dynamic processes. The principle relies

on the radioactive isotopes decay emitting radiation in the form of particles or waves which may be used for density measurements. For example, gamma-rays passing through the samples under test are absorbed depending on the volume, mass, and density of samples, as illustrated in Fig. 17. The rate of arrival of the rays after the absorption can be measured by using ion or scintillation based detection. By comparing the amount of rays entered into the known volume sample and with the rays detected at the end, the value of absorption may be determined accurately.

Generally, γ -ray mass absorption rate is independent of material composition; hence, they can be programmed for a wide range of materials. Densitometers based on radiation methods can provide accuracy up to ± 0.0001 g/mL. Many of these devices have self-diagnostic capabilities and are able to compensate for drift caused by source decay, thus pinpointing any signalling problems.

If a γ -ray of an intensity J_0 penetrates a material of a density ρ and thickness d , then the intensity of the radiation after passing through the material may be expressed by

$$J = J_0 \exp(-n\rho d) \quad (13)$$

where n is the mass absorption coefficient.

The accuracy of the density measurement is dependent on the accuracy of the measurement of the intensity of the radiation and the path length d . A longer path length through the material gives a stronger detection signal. Hence, for accurate operations, many different arrangements can be made for relative locations of transmitters and detectors, as illustrated in Fig. 18. Generally, the source is mounted in a lead container which is clamped onto the pipe or the container wall. In many applications, the detector is also clamped onto the wall.

The advantages of using radioactive methods are

1. The sensor does not touch sample; hence, there is no blockage on the path of the liquid.
2. Multiphase liquids can be measured.
3. They come in programmable forms and are easy to inter-

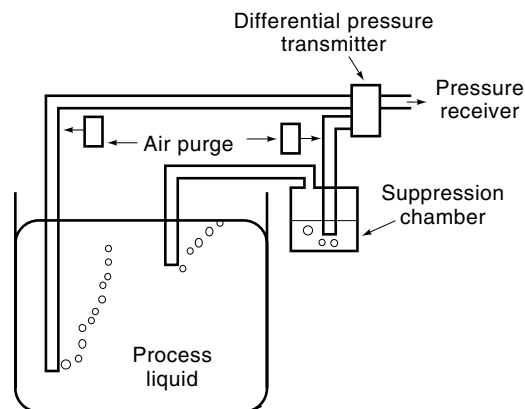


Figure 15. Two tube column densitometer. The pressure difference at the differential pressure transmitter depends on the relative positions of the openings of the pipes and the density of liquid. Once the relative positions are fixed, the pressure difference can be related to the equivalent weight of the liquid column at the openings of the pipes, hence the density of the liquid.

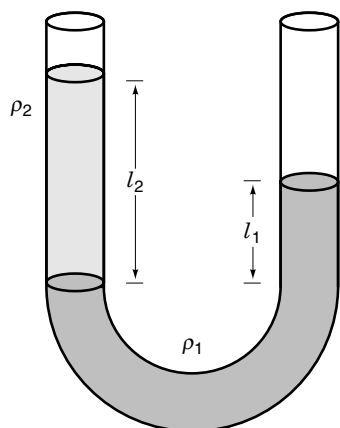


Figure 16. U-tube method. Unknown density of a liquid may be obtained from the known density of another liquid by placing them in a U-tube and measuring the lengths of liquid columns. In this method, the liquids must not be miscible. The method is not accurate but gives a quick idea about the density of the liquid.

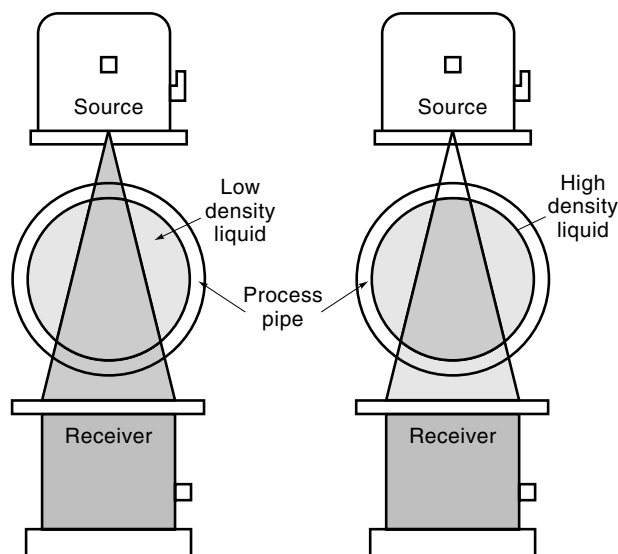


Figure 17. Radioactive density gage. The gamma-ray passing through samples is absorbed depending on the volume, mass, and density of the sample. The rate of arrival of the rays after absorption is determined by ion or scintillation based detection. The amount of absorption is proportional to the density of samples. Generally, γ -ray mass absorption rate is independent of material composition.

face.

The disadvantages are

1. A radioactive source is needed; hence, it is difficult to handle.
2. For a reasonable accuracy, a minimum path length is required.
3. There may be long time constants, making them unsuitable in some applications.
4. They are suitable only for solid and liquid density measurements.

OPTICAL METHODS

Refractometric Methods

This method is suitable for density measurements of gases and clear liquids. They are essentially optical instruments operating on the principles of refraction of light travelling in liquid or gas media. Depending on the characteristics of the samples, measurement of refractive index can be made in a variety of ways: critical angle, collimation, and displacement techniques are a few methods to mention. Usually, an in-line sensing head is employed, whereby a sensing window, commonly known as a prism, is wetted by the product to be measured. In some versions, the sensing probes must be installed inside the pipelines or in tanks and vessels. They are most effective in reaction type process applications where blending and mixing of liquids takes place. For example, the refractometers can measure dissolved soluble solids accurately.

Infrared diodes, lasers, and other lights may be used as sources. However, this measurement technique is not recommended in applications in processes containing suspended solids, high turbidity, entrained air, heavy colors, poor transparency and opacity, or extremely high flow rates. The readings are automatically corrected for variations in process temperature. The processing circuitry may include signal outputs adjustable in both frequency and duration.

Another version of refractometers is the index of refraction type densitometer. There are many versions of this type available. For example, in the case of position sensitive detectors, the index of refraction of a liquid under test is determined by measuring the lateral displacement of a laser beam. When the laser beam impinges on the cell at an angle of incidence, as in Fig. 19, the axis of the emerging beam is displaced by the cell wall and by the inner liquid. The lateral displacement can accurately be determined by position sensitive detectors. For maximum sensitivity, the devices need to be calibrated with the help of interferometers.

Refractometers are often used for the control of the adulteration of liquids of common use, such as edible oils, wines, and gasoline. They also find applications in industries such as pulp and paper, food and beverage, sugar, dairy, and other chemical

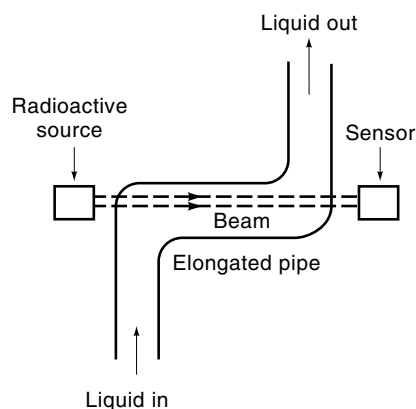


Figure 18. Fixed radioactive densitometer. An elongated path gives longer path length of the radioactive energy through the liquid, hence stronger attenuation. In some cases, the pipe may be enlarged to give longer beam length through the liquid.

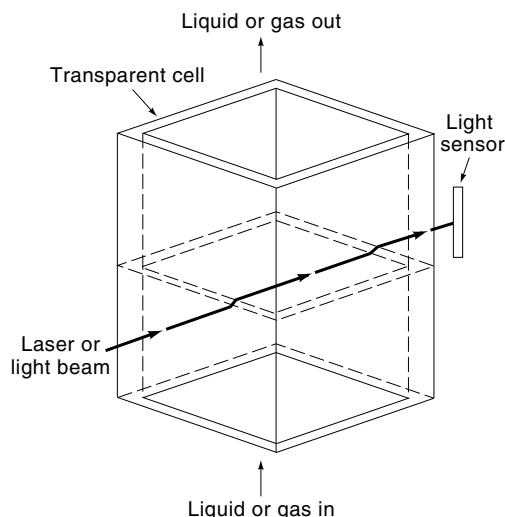


Figure 19. Index of refraction type densitometer. Angle of refraction of the beam depends on the shape, size, and thickness of the container, and the density of fluid in the container. Since the container has the fixed characteristics, the position of the beam can be related to density of the fluid. Accurate measurements of the position of the beam are necessary.

processes.

Absorption Type Densitometers

Absorption techniques are also used for density measurements in specific applications. X-rays, visible light, UV, and sonic absorptions are typical examples of this method. Essentially, attenuation and phase shift of a generated beam going through the sample is sensed and related to the density of the sample. Most absorption type densitometers are custom designed for applications having particular characteristics. Two typical examples are (1) UV absorption or x-ray absorptions are used for determining the local densities of mercury deposits in arc discharge lamps, and (2) ultrasonic density sensors are used in connection with difficult density measurements such as density measurement of slurries. The lime slurry, for example, is a very difficult material to handle. It has a strong tendency to settle out and to coat all equipment it contacts. An ultrasonic density control sensor can fully be emerged into an agitated slurry, thus avoiding the problems of coating and clogging. Since the attenuation of the ultrasonic beam is proportional to the suspended solids, the resultant electronic signal is proportional to the specific gravity of the slurry. Such devices can give accuracy up to 0.01%. The ultrasonic device measures the percentage of the suspended solids in the slurry by providing a close approximation of the specific gravity.

FALLING-SLUG METHOD

Drop Method

These methods are useful when only extremely small quantities of liquid or solid samples are available. Small drops of constant volume are put in a liquid which does not mix with the sample. The time necessary for sinking through a given

height is a measure of the density of the drops. In this method, the volume sample may be as low as 0.1 mL.

In another method, a density gradient is established by diffusion of two different liquids into one another. The drops reach an equilibrium position in the mixture, depending upon their density. In both cases, the calibration by drops of known density is necessary.

BIBLIOGRAPHY

- H. Bettin, F. Spieweck, and H. Toth, A computer-operated fluid-density measuring device using a balance and two permanent magnets, *Meas. Sci. Technol.*, **12** (11): 1036–1038, 1991.
- H. Eren, Particle concentration characteristics and density measurements of slurries using electromagnetic flowmeters, *IEEE Trans. Instrum. Meas.*, **44**: 783–786, 1995.
- Kay-Ray, Solution for Process Measurement*, Fisher-Rosemount, USA, 1995.
- P. Mason, Liquid and gas density measurements-Fluid density, *Meas. Control*, **25**: 297–302, 1993.
- J. A. McDowel, *Measuring density, and pressure-searching for structure*, Toronto: Holt, Rinehart and Winston, 1975.
- R. W. Miller, *Flow measurement engineering handbook*, 3rd ed., New York: McGraw-Hill, 1996.
- J. Moon, Monitor fluid density accurately, *Hydrocarbon Processing*, **73** (2): 87–89, 1994.
- Micro Motion Product Catalogue*, USA: Fisher-Rosemount, 1995.

HALIT EREN
Curtin University of Technology

LEVEL METERS

APPLIED ACOUSTICS: BASIC PRINCIPLES

Acoustics is the science that studies sound (1). Sounds are vibration phenomena (in gas, liquids, and solids) originally limited to the sense of hearing and subsequently extended to ultrasounds and infrasounds (2–5). In the following, the term *noise* will be used as both an undesired sound and as a statistically random oscillation, according to American Standard Acoustical Terminology (6).

Sound Pressure

A sound exists when a perturbation propagates in an elastic medium and causes a pressure variation and particle movement in the medium. When a particle of air is displaced from its mean position, there is a temporary variation in pressure. The pressure variation acts in two ways: by restoring the particle to its original position, and by passing on the disturbance to the next particle. The cycle of pressure increases (compressions) and decreases (rarefactions) propagates through the medium as a longitudinal sound wave, characterized by two important parameters: the pressure (the local increases and decreases with respect to the environment) and the velocity of the air particles that oscillate around a fixed position. The basic definitions of sound are in terms of the magnitude of the fluctuating component of pressure in a fluid medium. The sound pressure $p(t)$ is a scalar quantity, characteristic of the measurement points; it has a dimension of newtons per square meter or pascals. The sound pressure level (L_p) is measured in

decibels (dB):

$$L_p = 20 \log_{10} p/p_0 \text{ dB} \quad (1)$$

where p is the root-mean-square (rms) sound pressure, expressed in micro Pascal, and p_0 is equal to $20 \mu\text{Pa}$.

The rms value of the fluctuating component of pressure is used because most sound consists of random signals rather than pure sine waves. The value $20 \mu\text{Pa}$ is an accepted standard reference value of pressure against which other pressures are compared by Eq. (1). Note that when p equals $20 \mu\text{Pa}$, the sound pressure level is 0 dB. This value was selected somewhat arbitrarily, but it represents the average threshold of audibility for human beings if a 1000 Hz tone is used. That is, the 0 dB level was selected as the lowest pressure fluctuation normally discernible by human beings. The decibel (logarithmic) scale is used as a convenience because of the human ear, sensitivity to noise, which follows an approximately logarithmic law, and the great ranges of sound pressure level of interest in ordinary work (from 10^{-6} Pa to 10^3 Pa). For example, an office with tabulating machines may have an L_p of 60 dB to 65 dB. The average human threshold of pain is about 140 dB. Sound pressure close to large rocket engines is on the order of 160 dB. One atmosphere is 194 dB. The span from the lowest to the highest pressure of interest is thus on the order of 1 to 10^9 .

Sound Intensity

Sound pressure must not be confused with sound intensity, since the first is a scalar quantity, characteristic of the measurement point, and the second is a vector with its direction of propagation. Sound intensity $I(t)$ is a fundamental quantity for sound description (7): It is the product of particle velocity and pressure. As can be seen from the transformation of Eq. (2), it is equivalent for a given point and direction to the temporal average of the energy flux transmitted through a unitary surface perpendicular to the assigned direction in the considered point, and it is then measured in watts per square meter.

$$\begin{aligned} \text{Intensity} &= \text{Pressure} \times \text{Particle Velocity} \\ &= \text{Force/Area} \times \text{Distance/Time} \\ &= \text{Energy}/(\text{Area} \times \text{Time}) = \text{Power/Area} \end{aligned} \quad (2)$$

The acoustic intensity level L_I expressed in decibels is

$$L_I = 10 \log_{10} I/I_0 \text{ decibels (dB)} \quad (3)$$

where I is the time-averaged intensity, expressed in W m^{-2} , and I_0 is equal to 1 pW m^{-2} .

The relationship between intensity and pressure is known precisely only in the case of *free fields* and *diffuse fields*. Given that a sound field is a region where there is sound, fields are classified according to the manner and the environment in which the sound waves travel.

Free Field. The term *free field* describes sound propagation in an idealized free space where there are no reflections. These conditions hold in the open air or in an anechoic room where the sound striking the walls is totally absorbed. Free field

propagation is well characterized in terms of pressure and intensity level drop versus the distance from the source in the direction of sound propagation, as well as in terms of the relationship between sound pressure and acoustic intensity (8):

$$|I| = p^2/\rho c \quad (4)$$

where ρc is the acoustic resistance of the medium (ρ is the density and c the speed of sound in the medium). Having defined that, the acoustic impedance of the medium on a given surface lying in a wave front is the complex quotient of the sound pressure on the surface divided by the flux (volume velocity, or linear velocity multiplied by the area) through the surface; the acoustic resistance is defined as the real component of the impedance.

Diffuse Field. In a diffuse field, sound is reflected so many times that it travels in all directions with equal magnitude and probability. This field is approximated in a reverberant room. Although the net intensity $|I|$ is zero, there is a theoretical relationship that relates the pressure in the room to the one-side intensity, I_x (the intensity in one direction, ignoring the equal and opposite component) (88):

$$I_x = p^2/4\rho c \quad (5)$$

Sound propagation involves energy flow, but there can still be sound pressure even when there is no propagation. An *active field* is one where there is energy flow. In a *pure reactive field* there is no energy flow. A perfect correspondence, which gives rise to the use of the terms *active* and *reactive*, can be stated between pressure, particle velocity, and intensity, on the one hand, and voltage, current, and electrical power on the other hand. In an active field, pressure $p(t)$ and particle velocity $u(t)$ vary simultaneously. A peak in the pressure signal occurs at the same time as a peak in the particle velocity signal. They are therefore said to be in phase, and the product of the two signals gives a time-averaged intensity $I(t)$ (Fig. 1). In a pure reactive field the pressure and the particle velocity are 90° out of phase. One is shifted a quarter of wavelength with respect to the other. Multiplying the two signals together gives an instantaneous intensity $i(t)$ varying sinusoidally around zero. Therefore, the time-averaged intensity $I(t)$ is zero. In a diffuse field the pressure and particle velocity phase vary at random, so the time-averaged intensity is zero.

Sound Power

The sound power W is the energy emitted in the time unit, measured in watts, while the power level L_W expressed in decibels is

$$L_W = 10 \log_{10} W/W_0 \text{ decibels (dB)} \quad (6)$$

where W_0 is equal to 1 pW .

According to the law of energy conservation, the energy flow per time unit that crosses a surface that completely covers a source must be equal to the sound power of that source (except for eventual dissipation losses). The total sound power for a spherical propagation in an elastic media is linked to the sound

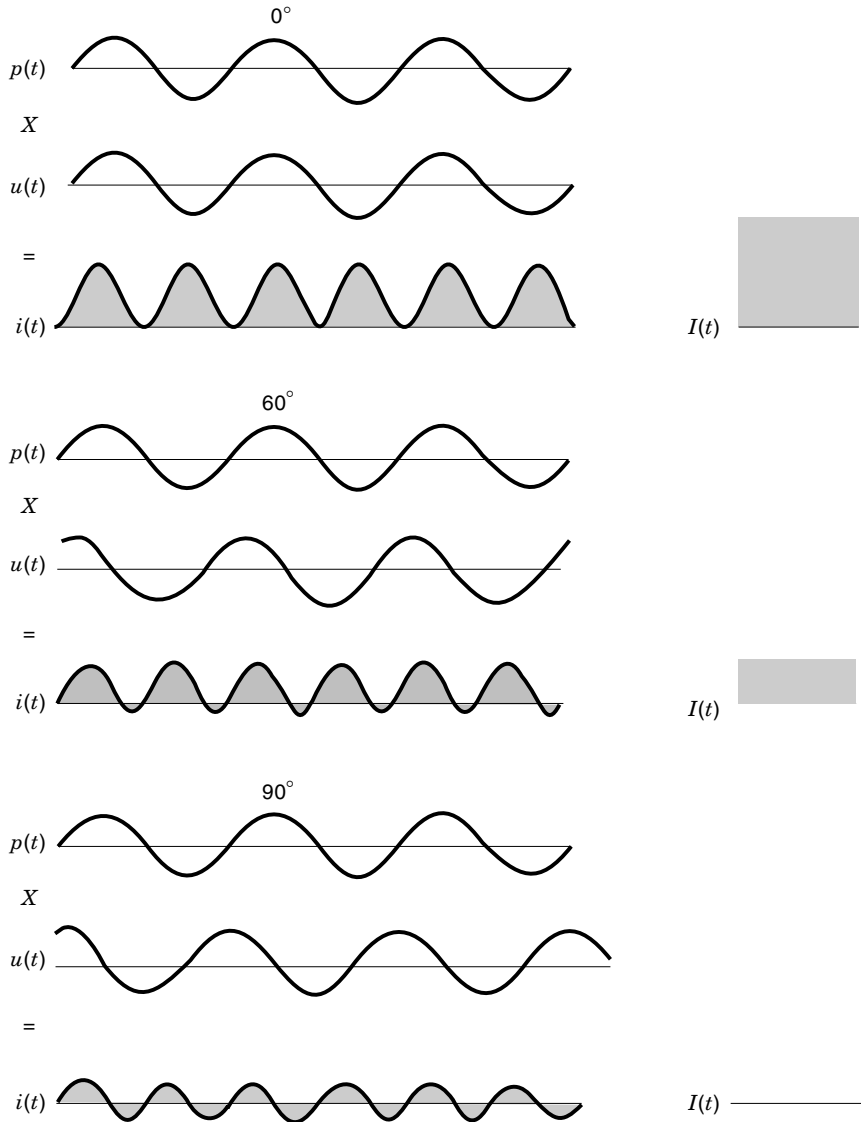


Figure 1. Instantaneous intensity $i(t)$ and time-averaged intensity $I(t)$ for different phase shifts between pressure $p(t)$ and particle velocity $u(t)$ highlight the differences between the two quantities and the effect of the phase shift on the $I(t)$ values.

intensity at distance r by the equation

$$W = 4\pi r^2 I = 4\pi r^2 (p_r^2 / \rho c) \quad (7)$$

where p_r is the sound pressure at a distance r .

Loudness

Loudness is a subjective quantity. It is defined as that aspect of auditory sensation in terms of which sounds may be ordered on a scale running from *soft* to *loud*. Loudness is chiefly a function of the level of sound pressure, but it is also dependent on the frequency and the composition of the sound. The range of loudness is divided subjectively into equal-unit steps called *sones*. The loudness of a sound at a given sound pressure level at one frequency may be quite different from the loudness of a sound of the same level at a different frequency. Nevertheless, listeners can adjust the level of one tone to match the loudness of another, and fair agreement among observers is usually obtained. Such experiments provide a useful objective scale of loudness, called loudness level. The loudness level of a tone in *phons* is

numerically equal to the sound pressure level of a 1000 Hz tone that sounds equally loud. The loudness in *sones* (S) can be related to the loudness level in *phons* (P) by the empirical formula

$$S = 2^{(P-40)/10} \quad (8)$$

The equal loudness contours for pure tones of Fig. 2 were obtained from measurements on human beings and show that the frequency response of the human ear is both nonflat and nonlinear (9). The numbers on the contour indicate the loudness level in phons, with 0 phon corresponding to the threshold of audibility. These curves were obtained under free-field conditions, and the ordinate of the curve is free-field sound pressure level. Thus, the ordinate indicates what pressure amplitude must be applied at any given frequency so that the human observer will perceive a sensation of equal loudness. For example, at a 50 phon loudness, a sound level of 58 dB L_p at 100 Hz sounds as loud as one of 50 dB L_p at 1000 Hz. This demonstrates the nonflatness of the ear's frequency response. Its nonlinearity is manifested by the need for a family

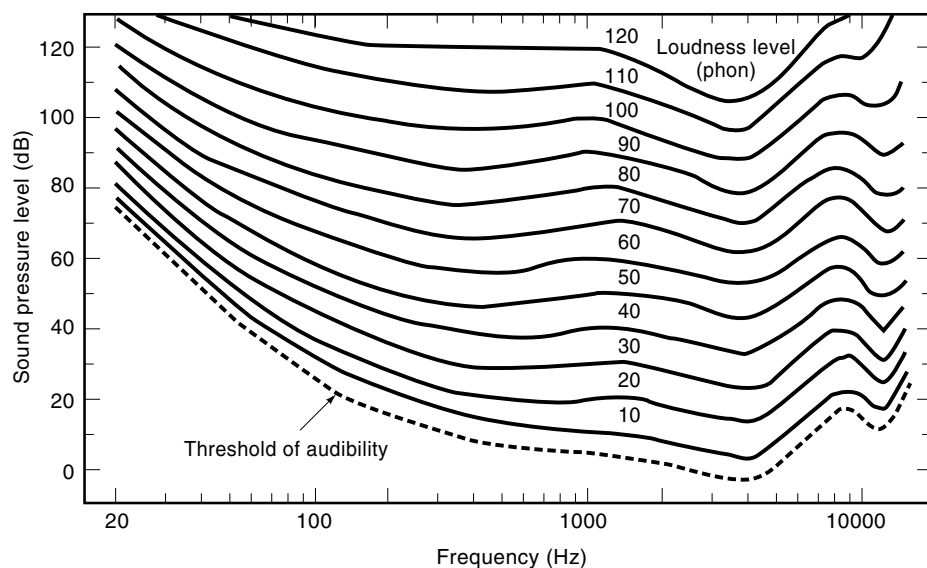


Figure 2. Normal equal loudness contours for pure tones. The levels were measured after the listener left the free plane-wave sound field. (ISO R226-1961)

of curves for various loudness levels, rather than just a single curve.

The calculation of loudness is crucial in characterizing loudspeakers, whose quality depends on how much the transduced music or speech is loud with respect to background noise. Namely, the loudness efficiency rating is defined as the ratio of the total effective acoustic power produced by the loudspeaker to the available electrical power; here the total effective acoustic power is measured so that it is nearly proportional to the loudness produced by the loudspeaker in the free field (10).

Reference Levels

Sound pressure, intensity, power, and particle velocity levels (L_p , L_I , L_W , and L_u , respectively), are all measured in decibels. Decibels are a ratio of the specified quantity measured against some reference. As previously stated, the pressure reference level is chosen so that it corresponds approximately to the threshold of audibility. Other reference levels have been approximately related to this by using the free field relations between pressure and intensity, and pressure and particle velocity. In the free field we will obtain the same decibel reading irrespective of whether we measure pressure, intensity, or particle velocity (measured in the direction of propagation). Actually, because round numbers have been chosen for the reference levels, there is a slight difference in levels. The actual difference depends on the value of the acoustic impedance, ρc , of the medium in which the difference is measured. The difference is usually negligible in the air except at high altitudes. To avoid possible confusion with pressure levels, sound power levels are sometimes given in bels (10 dB equals 1 bel).

In the free field the pressure and intensity levels in the direction of propagation are numerically the same. However, intensity measurements in the free field are not needed. In practice, measurements are not performed in the free field, so there will be a difference between the pressure and intensity levels. This difference is an important quantity. It is known variously as the reactivity index, pressure-intensity index, or phase index.

INSTRUMENTS FOR SOUND-LEVEL MEASUREMENTS

Sound-Level Meter

The most commonly utilized instrument for routine sound measurement is the sound-level meter (11,12). This is basic to all sound- and noise-level measurement, particularly that done outside the laboratory. It is actually a measurement system made up of a number of interconnected components. Figure 3 shows a typical arrangement.

The sound pressure $p(t)$ is at first transduced to a voltage by means of the microphone. Microphones generally employ a thin diaphragm to convert pressure to motion. The motion is then converted to voltage by suitable transducers, usually a capacitance, piezoelectric, or electret type.

The output voltage of the microphone generally is quite small and at a high impedance level; thus an amplifier of high input impedance and gain is used at the output of the microphone. This can be a relatively simple ac amplifier, since response to static or slowly varying voltages is not required. Capacitor microphones often use for the first stage a field-effect transistor (FET)-input amplifier built directly into the microphone housing. This close coupling reduces stray capacitance effects by eliminating cables at the high-impedance end.

The weighting networks follow the first amplifier. They are electrical filters whose frequency response is tailored to approximate the frequency response of the average human ear. In fact, the main use of a sound-level meter is not the accurate measurement of pressure but rather the determination of the loudness perceived by human beings; for this reason a flat instrument frequency response is not really desired. The weighting networks of Fig. 3 are electrical filters suitably designed to approximate the human ear's response at different loudness levels so that the instrument reading will respect perceived loudness. Usually three filters are provided: A, B, and C. Figure 4 shows the frequency response of these filters. Some meters also provide a *flat* setting if true pressure measurements are wanted; if a flat setting is not available, the C network is a good approximation of a flat response. Actually, many practical measurements are made by employing the A scale

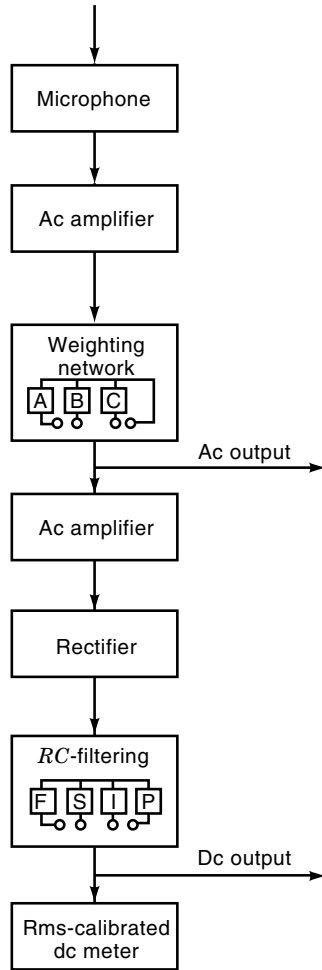


Figure 3. Simplified block diagram of a sound-level meter.

since it is a simple approach that has given good results in many cases and has been written into many standards and codes. The level of sound pressure measured with a sound-level meter and its weighting network is called sound level to distinguish it from the original sound pressure levels. When A-weighting is used, the sound level is given in dB(A) instead of dB.

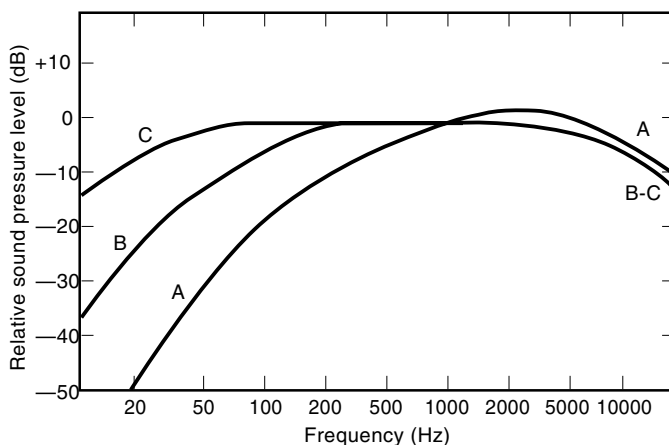


Figure 4. Normalized A, B, and C weighting curves.

The output of the weighting network is further amplified, and an output jack is provided to lead this ac signal to an oscilloscope (if observation of the waveform is desired) or to a wave analyzer (if the frequency content of the sound is to be determined). If only the overall sound magnitude is desired, the rms value of the voltage signal must be found. While true rms voltmeters are available, their expense is justifiable only in the highest-grade sound-level meters. Rather, the average value of the ac signal is determined by rectifying and filtering; then the meter measures the obtained dc value on a scale usually calibrated to provide the rms value. This procedure is exact for pure sine waves since there is a precise and known relationship between the average value and the rms value of a sine wave. For nonsinusoidal waves this is not true, but the error is generally small enough to be acceptable for relatively unsophisticated work. The filtering is accomplished by both a simple low-pass RC filter and low-pass meter dynamics. Moreover, the filtering time constant also acts on the meter response time. In particular, most meters have a switch that changes the filtering time constant. The *slow* position (1 s time constant) gives a steady, easy-to-read needle position but masks any short-term variations in the rms of the signal. If these short-term variations in the signal are of interest, they may be visually observed on the meter by switching it to the *fast* (125 ms), *impulse* (35 ms), or *peak* (20 μ s to 50 μ s) position. Figure 5 shows the effect of the different time constant on a sound-level measurement.

Finally, a dc meter calibrated in rms gives the measurement result. Since Eq. (1) establishes a definite relation between sound pressure in μ Pa and dB, the meter is often directly calibrated in decibel.

Integrating Sound-Level Meter

Integrating sound-level meters are designed to measure continuously the *continuous equivalent level* in a defined time interval T , $L_{eq,T}$ (13). The aim of this parameter is to characterize, with a unique measurement data, a variable noise. The $L_{eq,T}$ of a signal is the level of a hypothetical constant level noise that has the same acoustic energy as the original signal in the same period of time and, namely,

$$L_{eq,T} = 10 \log \frac{1}{T} \int_0^T \left(\frac{p(t)^2}{p_0^2} \right) dt \quad (\text{dB}) \quad (9)$$

where p_0 is the reference pressure (20 μ Pa) and $p(t)$ is the sound pressure level of the noise under analysis. Usually A-weighted sound levels, rather than sound pressure levels, are considered and then a $L_{Aeq,T}$ is obtained.

If the noise under analysis is constituted by a sequence of constant levels L_{A1} , L_{A2} , ..., L_{AN} , each one present for a time interval Δt_1 , Δt_2 , ..., Δt_N , respectively, or if the available measures are in this form, it is possible to evaluate the $L_{Aeq,T}$, with $T = \sum_i \Delta t_i$ by the equation

$$L_{Aeq,T} = 10 \log \frac{1}{T} \sum_{i=1}^N 10^{L_{Ai}/10} \Delta t_i \quad [\text{dB(A)}] \quad (10)$$

In the field of ear protection, a more complex formula was employed for a long time and can sometimes still be implemented; this formula has no physical explanation but allows the parameters q [exchange factor, representing the level increment in dB(A) that requires halving the exposure time to

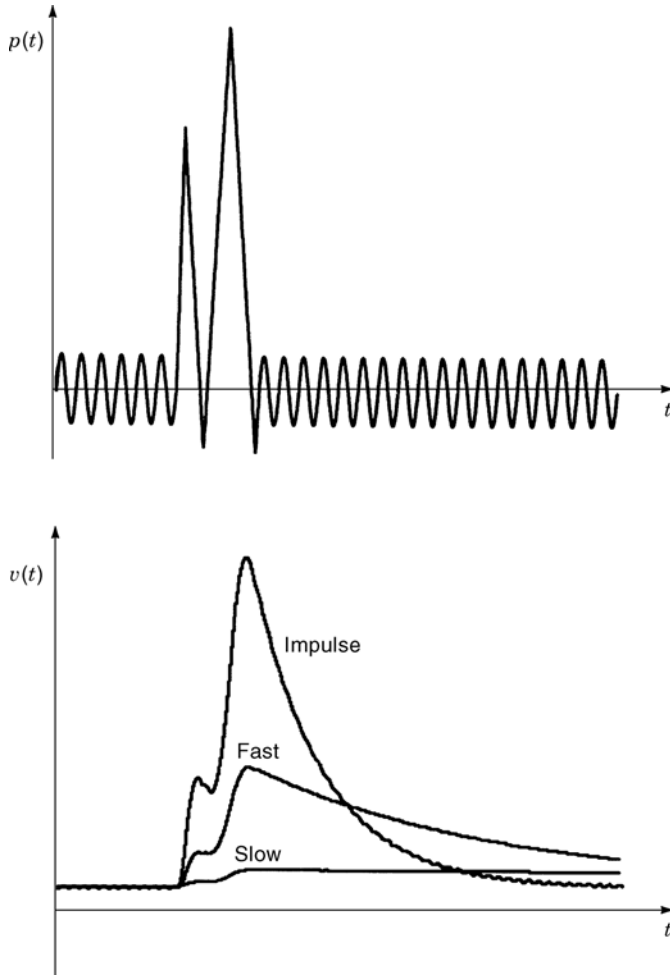


Figure 5. The RC-filtering outputs for different time constants in the case of an impulse noise.

obtain a constant risk] and L_t (threshold level, namely, a level under which the noise damage can be considered absent) to be introduced:

$$L_{eq,T} = L_t + \frac{q}{\log 2} \log \frac{1}{T} \sum_{i=1}^N 10^{\log 2(L_{Ai} - L_t)/q} \Delta t_i \quad (\text{dB}) \quad (11)$$

which gives back the previous equation for $q = 3$ and $L_t = 0$. The value of q is usually fixed on the basis of the reference standards: In the United States a $q = 5$ is suggested. In Europe a $q = 3$ is suggested. The value of L_t is often assumed to be equal to 70 dB.

It is important to emphasize that even if the $L_{eq,T}$ is often referred to as an average noise measurement, a simple arithmetic mean cannot be used to compute its value starting from partial values, since it is a logarithmic average. For example, an 8 h exposure time subdivided into 4 h at 100 dB and 4 h at 80 dB is characterized by an $L_{eq,8h}$ of 97 dB rather than 90 dB. As a further example, two 80 dB sources give a total $L_{eq,T}$ of 83 dB—namely, 3 dB higher than the one produced by each source.

Another useful parameter found on more elaborated integrating sound-level meters is the *sound exposure level*, $L_{EA,T}$,

also measured in dB(A) and often referred to as SEL. This is defined as the level that, lasting for 1 s, has the same acoustic energy as a given noise event lasting for a chosen period of time T . As a measure of acoustic energy, the SEL can be used to compare unrelated noise events, since the time element in its definition is always normalized to one second. The value of T is often chosen as the time interval in which the sound level is no lower than 10 dB of the maximum value, in order to avoid an increase in the measurement time that is not justified by a significant SEL variation.

Some integrating sound-level meters measure the continuous equivalent level over a fixed period of 60 s, $L_{Aeq,60s}$. There is a certain similarity in use between the quantities $L_{Aeq,60s}$ and $L_{EA,T}$, since both are based on measurement performed at fixed time intervals. In addition, for any noise event lasting no more than 60 s, there is a fixed difference between its SEL and its $L_{Aeq,60s}$ (the former is greater by 17.8 dB).

A simplified block diagram of an integrating sound-level meter is shown in Fig. 6. It adds an integrating section fed by the signal already amplified and weighted to the already described features of a sound-level meter (see Fig. 6). This section performs the numeric evaluation of the equivalent sound level on the basis of the previously introduced formulas. It is important here to stress the differences among the average operations performed by this integrating section and those performed by sound-level meters. At first, traditional sound-level meters can average signals for prefixed and, in any case, short time intervals (maximum 1 s for slow operating), whereas integrating instruments can average for hours. Furthermore, integrating sound-level meters equally weight all the sounds occurring in the considered time window, while traditional sound-level meters give a greater weight to more recent sounds. In fact, the time weighting of traditional sound-level meters decreases exponentially; for example, having chosen a slow 1 s time constant, greater weights are given to sounds occurring less than 1 s previously while sound occurring 10 s previously has very little influence on the meter output. The integrating section has both a better capability in following rapid time evolution of the sound pressure and a higher crest factor (ratio between peak and rms values), which can reach 60 dB.

Noise Dose Meter

Noise dose meters are used to measure the continuous equivalent level of randomly fluctuating noise on an 8 h or longer time interval (14). Namely, a noise dose meter measures D , the percentage of daily noise dose that is allowed by the standards. This instrument is a miniature integrating sound-level meter that uses the A-weighting network and allows the desired exchange factor to be selected. The noise dose meter measures continuously and at the same time reads out the dose as a percentage of the maximum allowable (100%) over an exposure period (usually 8 h). When representative data can be obtained in less time, the reading can be converted easily to an equivalent 8-h exposure.

Two kinds of noise dose meters can be used: personal dose meters, carried by a person, and noise dose meters, installed in a fixed place. The former are usually provided with a small microphone located near the ear; the latter, used to control noisy areas, are usually exposed to a sound level lower than that of a microphone installed on a person immersed in the same sound field. This difference, usually contained within

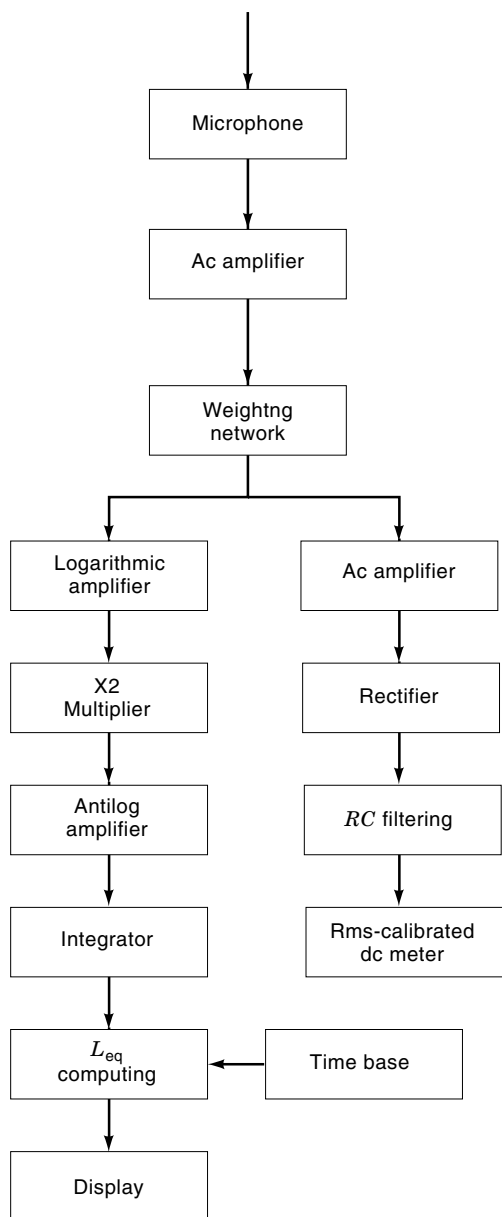


Figure 6. Simplified block diagram of an integrating sound-level meter.

2 dB, is due to the different sound pressure on a microphone located near a reflective surface: The actual increase depends on the noise spectrum, the noise direction, and the installation.

Sound-Level Statistical Analyzer

This kind of instrument allows the sound level to be analyzed statistically in a defined time interval. It was designed to deal with urban noise, especially traffic noise, but it can be used to measure noise in working areas as well. Its measurement principle is conceptually the following: Suppose we want to evaluate the evolution of the sound level in a 50 dB(A) range—for example, from 40 dB(A) to 90 dB(A)—and that this range is subdivided into 25 contiguous classes of 2 dB(A) each. By sampling the sound level at a constant rate (e.g., every 0.1 s) and

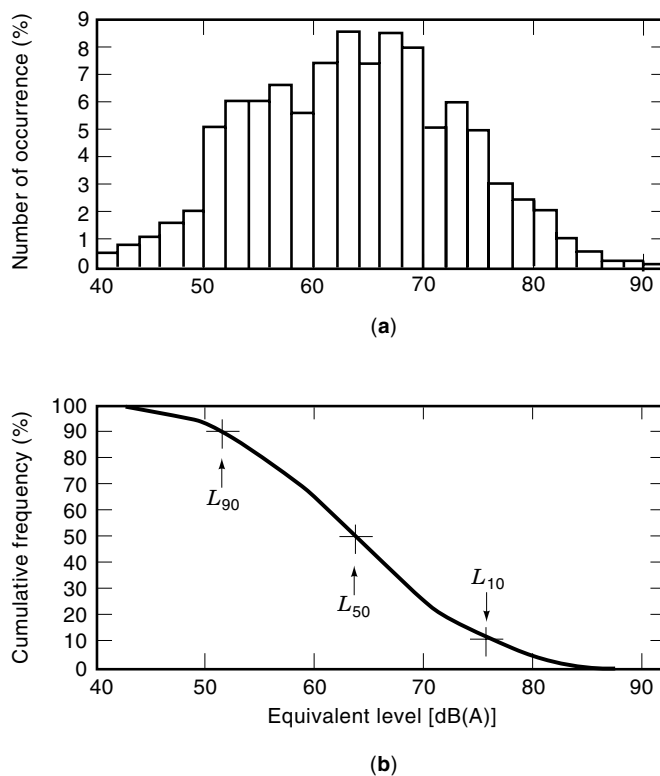


Figure 7. Example of sound-level statistical distribution obtained by using a sound-level statistical analyzer: (a) absolute frequency distribution; (b) cumulative frequency distribution [in this example, $L_{90} = 51.7$ dB(A); $L_{50} = 63.9$ dB(A); $L_{10} = 75.8$ dB(A)].

by assigning each sample to the corresponding class, the measurement result is obtained as a number of counts for each class or, since the number of counts can be linked to a time interval, as the time percentage of the total investigation time in which the sound level was contained in each dB(A) interval. For example, if we have a 600 s measurement time in which 6000 samples were acquired (0.1 s sampling time) and the class 58 dB(A) to 60 dB(A) was counted 1200 times, the sound level was contained within 58 and 60 dB(A) for 20% of the total measurement period. The result of the statistical distribution is shown in Fig. 7(a). A cumulative frequency distribution [see Fig. 7(b)], is almost always preferred, thus considering the time percentage during which a certain sound level was exceeded.

On the basis of the information gathered according to this criterion, the meter is able to produce, as a final result, together with the sound-level statistical distribution, the equivalent sound level L_{eq} and the so-called statistical levels, typically L_{90} (sound level exceeded for 90% of the measurement time and consequently representative of background noise), L_{50} (sound level exceeded for 50% of the measurement time and consequently representative of the average level), and L_{10} (sound level overcome for 10% of the measurement time and consequently representative of the maximum sound levels).

A statistical analyzer can be used instead of an integrating sound-level meter to evaluate L_{eq} , taking into consideration that this value is not estimated by means of a real integration but by summing the values obtained by sampling, at a constant rate, the sound level that varies according to the meter time

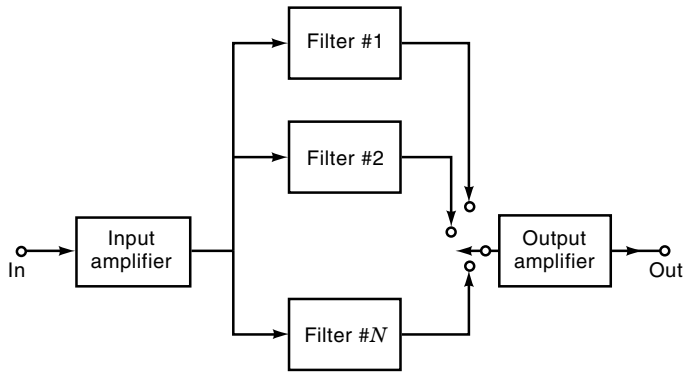


Figure 8. Filtering section of a sequential frequency analyzer.

constant (slow or fast). This can create problems when dealing with variable or impulsive noises.

Sound-Level Spectrum Analyzer

In some applications a frequency analysis of the acoustic signal is required, in terms of separation of the different frequency components present in the overall signal (15). In fact, the spectrum represents an additional element to the global equivalent sound level expressed in dB(A) to characterize noise correctly, by highlighting the presence of pure tones or high frequencies. Frequency analysis is indispensable for the design of noise control techniques since it allows the main noise sources to be identified and noise-control techniques, materials, and structure to be optimized. This task is carried out by instruments called *spectrum analyzers*.

As far as measurement principles, characteristics, and problems of usage of general-purpose spectrum analyzers as well as of FFT-based digital spectrum analyzers are concerned, the related articles of this Encyclopedia, together with other specialized texts, can be used. In the field of spectrum analysis of acoustic signals, a further possibility is available: performing a sequential frequency analysis, according to Fig. 8. In this case the signal is conceptually applied in parallel to a set of suitable filters. The output for the different bands is read sequentially by an output unit. It is evident that this approach is correct only when a stationary noise has to be measured, or at least a noise that is stationary during the overall measurement time (usually a few minutes). In any case this kind of approach, even if not rigorous, can be used in the absence of more sophisticated instruments to obtain useful qualitative information about the noise spectrum. The frequency range of each filter defines its band or bandwidth; the most commonly used bandwidth in acoustics is the octave (16). An octave is the interval between two frequencies having a ratio of 2, [e.g., from 707 Hz to 1414 Hz (central frequency 1000 Hz, determined as the geometric average between the two frequencies)]. An analyzer that uses this bandwidth is called an octave-band analyzer. Other analyzers use narrower bandwidths to allow a more detailed frequency analysis (e.g., third-octave-band analyzers). Other analyzers use filters with a constant bandwidth in hertz. An ideal filter has a uniform response within its passband and no response out of its passband. Of course this behavior is only approximated by real filters, since their response is not uniform within the passband and not zero outside the passband, giving a meaningful output when relevant frequency components are

present immediately out of the passband. The *effective bandwidth* of a filter is the bandwidth of an ideal filter that has the same maximum response and the same output of the real filter when a white noise is fed into both filters. Most analyzers are designed to have an effective bandwidth that is very similar, practically equal, to the filter bandwidth.

SOUND INTENSITY ANALYZING SYSTEM

A sound intensity analyzing system consists of a suitable probe and an analyzer. The probe measures the pressure at a pair of microphones, while the analyzer performs the integration and calculation necessary to evaluate the sound intensity.

Sound intensity is the time-averaged product of the pressure and particle velocity. While instantaneous pressure can easily be measured by using a microphone, a direct measurement of particle velocity is not common and requires the use of devices like hot wire anemometers or delicately suspended mica disks (17). However, as the particle velocity is related to the pressure gradient (the rate at which the instantaneous pressure changes with distance), a simpler measurement method can be set up using two identical pressure microphones. The method is based on Newton's second law:

$$\rho \frac{\partial \mathbf{u}}{\partial t} = -\nabla p \quad (12)$$

also called Euler's relation, where ρ is the density of air and $\mathbf{u}$ the particle velocity.

In one direction, r , we have

$$\rho \frac{\partial u_r}{\partial t} = -\frac{\partial p}{\partial r} \quad (13)$$

Since the pressure gradient is proportional to particle acceleration, particle velocity can be obtained by integrating the pressure gradient with respect to time.

$$u_r = -\frac{1}{\rho} \int \frac{\partial p}{\partial r} dt \quad (14)$$

In practice, the pressure gradient can be approximated by measuring the pressures, p_A and p_B , at two closely spaced points, A and B, and dividing the pressure difference $p_A - p_B$ by the transducer separation distance Δr , thus giving the following estimate for the particle velocity component u_r in the direction r :

$$u_r = -\frac{1}{\rho \Delta r} \int (p_A - p_B) dt \quad (15)$$

This approximation is valid as long as the separation is small compared with the wavelength ($\Delta r \ll \lambda$).

Practical sound intensity probes therefore consist of two closely spaced pressure microphones, allowing measurement of both pressure and the component of the pressure gradient along a line joining the centers of the microphones. Hence, the magnitude and the direction of the component of the intensity vector along this line is measured.

The probe arrangement of two microphones mounted face to face with a solid spacer in between has been found to have better frequency response and directivity characteristics than

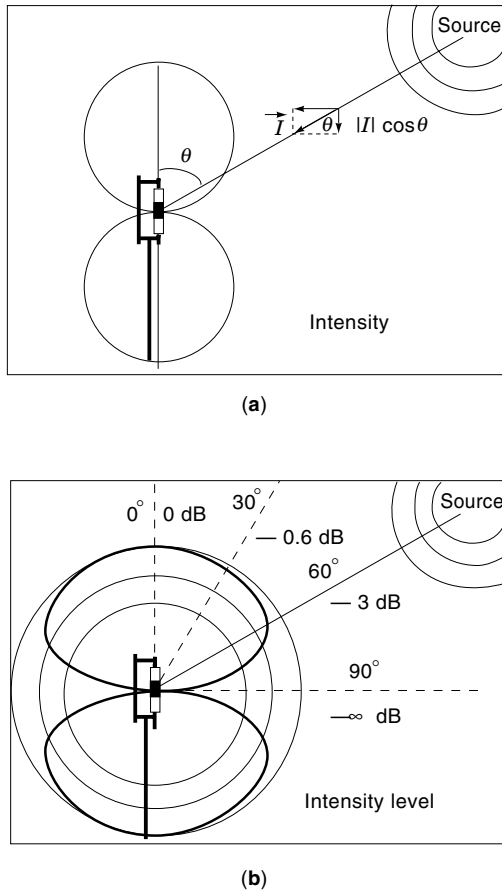


Figure 9. Intensity measurement by two-microphone sound-intensity probe. The difference between intensity (a) and intensity level (b) is shown. The directivity characteristic (cosine characteristic) is also highlighted in (b), where the differences between the measured intensity component and the intensity module are reported in decibels for different angles.

side-by-side, back-to-back, or face to face without solid spacer arrangements. As for directivity characteristics of the sound intensity analyzing system, as already mentioned, it is an intensity vector component and not the intensity vector that is measured by this technique. The consequence is that the theoretical directional characteristic of the sound intensity probe is a cosine function

$$|I_r| = |I| \cos \alpha \quad (16)$$

where α is the angle between the direction of the sound intensity vector and the orientation of the probe (Fig. 9).

For processing the signals from the two microphones, two approaches are in current use today. One approach is a direct method that can be implemented by analog as well as digital techniques. The second approach, the indirect method, can only be implemented by use of a digital technique.

Direct Method

The sound intensity vector component in the direction r is calculated from

$$I_r = -\frac{1}{2\rho\Delta r} (p_A + p_B) \overline{\int (p_A - p_B) dt} \quad (17)$$

where sound pressure is taken as being the mean pressure $(p_A + p_B)/2$ between the two microphones, and where the velocity is calculated from Eq. (15) (the superior line indicates an averaging).

Figure 10 shows a block diagram of a practical real-time sound intensity meter (including third-octave digital filters) that follows the equation step by step. Instruments like this have an analysis range from a few hertz to about 10 kHz.

Indirect Method

A dual channel FFT analyzer can be used for intensity calculations within the well-known FFT-limitations. It can be shown (8) that the intensity can be calculated from the imaginary part of the cross-spectrum G_{AB} between the two microphone signals.

$$I_r = -\frac{1}{\omega\rho\Delta r} \text{Im}(G_{AB}) \quad (18)$$

Today, this forms a commonly used method of calculating sound intensity. However, a computer is required to carry out the final calculations. Unfortunately, this method has certain disadvantages. One of these is that sound measurement is normally specified in octaves and third octaves, and the calculation of these from narrow-band spectra is a time-consuming procedure, usually requiring multipass analysis and synthesis, which cannot be performed easily in real time.

MICROPHONES

The microphone is the first element of each acoustic measurement chain, since it transduces sound pressure variations into corresponding variations of an electric signal (17). While microphone design is a specialized and complex field with a large

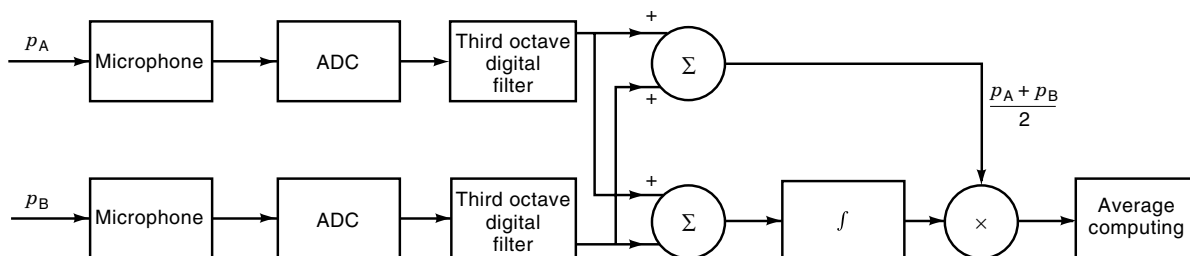


Figure 10. Block diagram of a real-time digital sound intensity meter.

amount of technical literature, some of the main considerations will be presented in the section. As for many other transducers, sensitivity, amplitude, and frequency response are characteristics of major interest. However, sound wavelength and the direction of propagation produce effects on transducer amplitude and frequency response that are aspects of dynamic behavior not regularly encountered in other measurements.

The *pressure response* of a microphone refers to the frequency response relating a uniform sound pressure applied at the microphone diaphragm to the output voltage of the microphone. The pressure response of a given microphone may be estimated theoretically or measured experimentally by one of a number of accepted methods (18).

What is usually desired is the *free-field response* of the microphone (namely, the relation between the microphone output voltage and the sound pressure that existed at the microphone location before the microphone was introduced into the sound field). As a matter of fact, the microphone distorts the pressure field because its acoustical impedance is radically different from that of the medium (air) in which it is immersed. For most purposes the microphone (including its diaphragm) may be considered a rigid body. Sound waves impinging on this body give rise to complex reflections that depend on the sound wavelength (frequency), the direction of propagation of the sound wave, and the microphone size and shape. When the wavelength of the sound wave is very large compared to the microphone dimensions (low frequencies), the effect of reflections is negligible for any angle of incidence between the diaphragm and the wave-propagation direction, and the pressure response equals the free-field response. At very high frequencies, where the wavelength is much smaller than the dimension of the microphone, it acts as an infinite wall, and the pressure at the microphone surface (for waves propagating perpendicular to the diaphragm [0° angle of incidence]) is twice what it would be if the microphone were not there. For waves propagating parallel to the diaphragm (90° incidence angle), the average pressure over the diaphragm surface is zero, giving no output voltage.

For simple geometric shapes, such as spheres and cylinders, theoretical results are available; otherwise an experimental characterization of the actual microphone has to be carried out. Note that for sufficiently low frequencies (below a few thousand hertz) there is little change in pressure because the presence of the microphone and the angle of incidence have little effect. This flat frequency range can be extended by reducing the size of the microphone; however, a smaller size tends to reduce sensitivity. The effect of size is directly related to the relative size of the microphone and the wavelength of the sound. The wavelength λ of sound waves in air is roughly $330/f$ m, where f is the frequency in hertz. When λ becomes comparable to the microphone diaphragm diameter, significant reflection effects can be expected.

Each microphone has a random incidence curve that refers to its response to a diffuse sound field (namely, where the sound is equally likely to come to the microphone from any direction, the waves from all directions are equally strong, and the phase of the waves is random at the microphone position). Such a field may be approximated by constructing a room with highly irregular walls and placing reflecting objects of various sizes and shapes in it. A source of sound placed in such a room gives rise to a diffuse sound field at any point in the room. Microphones

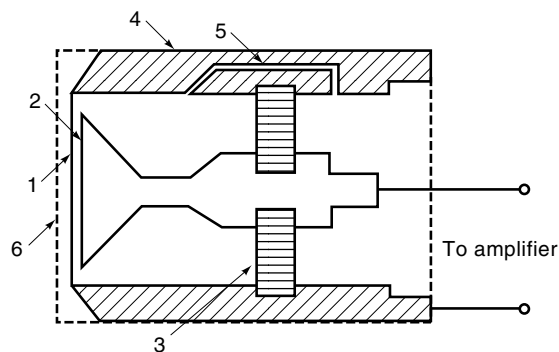


Figure 11. Simplified diagram of a capacitor microphone. (1) Diaphragm; (2) back plate; (3) insulator; (4) case; (5) capillary tube for average pressure equalization; (6) protection grid.

calibrated under such conditions are of interest because many sound measurements take place in enclosures that, while not giving perfect random incidence, certainly do not give pure plane waves. Microphone calibrations usually give the pressure response and the free-field response for selected incidence angles, usually 0° and 90° .

Microphones used for engineering measurements are usually capacitor, electret, or piezoelectric types.

A capacitor microphone is constituted by a thin metallic membrane (about $5\ \mu\text{m}$ thickness) that represents the sensing element, mounted in parallel (at about $25\ \mu\text{m}$) from a rigid posterior plate (back plate), thus forming a capacitor (Fig. 11). The capacitor charge is maintained constant by a constant polarization voltage. When the membrane is solicited by sound pressure, a capacitance and, consequently, a voltage variation occur proportional to the sound pressure. Microphones often have a slow leak (capillary tube) connecting the two sides of the diaphragm, to equalize the average pressure (atmospheric pressure) and prevent bursting of the diaphragm. This is necessary because the (slow) hour-to-hour and day-to-day changes in atmospheric pressure are much greater than the sound pressure fluctuations to which the microphone must respond. (Note that the Eustachian tube of the human ear serves a similar function.) The presence of this leak dictates that microphones will not respond to constant or slowly varying pressures. This is usually not a problem since many measurements involve a human response to the sound, and this is known to extend down to only about 20 Hz. Thus the microphone frequency response need only reach this value, not zero frequency. The typical sensitivity is of about 50 mV/Pa.

Electret-type microphones are related to the capacitor types; however, they require no polarizing voltage since their charge is permanently built into the polymer film that forms the diaphragm. Since the unsupported polymer film would sag and creep excessively, a backup plate with raised points is used. Such microphones are less expensive than the capacitor type, can be used in high-humidity conditions (where the capacitor type may arc over), and result in smaller instruments with lower power consumption. A version that preserves the desirable features of an all-metal diaphragm has also been developed.

Piezoelectric microphones use PZT (or lead zirconate titanate) as a bending beam coupled with the center of a conical

diaphragm of thin metal foil. This kind of microphones offers the following advantages compared to a capacitor microphone: less sensitivity to high humidity; greater robustness; response in a wide frequency range, even with a lower regularity; no polarization voltage required; and excellent stability. On the contrary, it has greater vibration sensitivity; greater sensitivity drift with temperature; and electrical characteristics that vary with temperature.

The selection and use of microphones for critical applications require some background in acoustics, which is beyond the scope of this article; fortunately, useful references are available (2,19).

METER CALIBRATION

Calibration of Level Meters

Traditional and integrating sound-level meters, as any other measurement instrument, must be calibrated periodically in order to verify the stability of their functionality (20). In particular, it is strongly recommended to calibrate a sound-level meter at the beginning and at the end of a set of measurements and at least before and after each day's measurement. If the calibration levels do not coincide, measured data can be corrected for differences contained within 1 dB, but data with greater differences will be discarded.

The apparatus used for the calibration is the calibrator (namely, a sound source that can operate on the basis of different principles but represents the transfer standard between national standards and the meter under calibration). Electromechanical and electroacoustic calibrators are normally used; in the former, sound pressure is generated by the oscillation in a phase opposition of two small pistons, driven by a disk cam; in the latter, a stable oscillator supplies a metallic membrane of a piezoelectric drive. The reference temperature (20 °C) and pressure (101.3 kPa) must be assured or the appropriate correction must be applied.

Calibration of Sound Intensity Meters

One of the advantages of using the two-pressure-microphone technique is the ease with which very accurate calibration can be carried out using a pistonphone, which provides a known sound pressure level at a known frequency. As already detailed, the reference values for sound pressure levels and for intensity levels are 20 μ Pa and 1 pW/m², respectively. These reference values have been chosen so that for a freely propagating plane wave, a 0 dB sound pressure level corresponds to a 0 dB sound intensity level.

A calibrated barometer is also necessary to determine the necessary correction for the ambient atmospheric pressure. In fact, both the sound pressure of the pistonphone and the air density are proportional to the ambient pressure. Keeping in mind Eq. (4), the correction term is

$$\begin{aligned}\Delta L_1(p_{\text{amb}}) &= 20 \cdot \log_{10}(p_{\text{amb}}/p_0) - 10 \cdot \log_{10}(p_{\text{amb}}/p_0) \\ &= 10 \cdot \log_{10}(p_{\text{amb}}/p_0)\end{aligned}\quad (19)$$

where p_0 equals 0.101 MPa (1 atm). Therefore, when calibrating the system for use in sound intensity mode, only half the

atmospheric pressure correction indicated on the barometer scale must be applied to each microphone.

The air density is inversely proportional to the absolute temperature T , which leads to the temperature correction term:

$$\Delta L_1(T) = 10 \cdot \log_{10}(T/T_0) \quad (20)$$

where $T_0 = 293$ K (20 °C).

In general, these correction terms can often be ignored since they are relatively small; for example, for a temperature of 40 °C and an ambient pressure of 75 kPa (Mexico City, 2300 m above sea level) the correction term is only 1.0 dB.

BIBLIOGRAPHY

1. L. L. Beranek, *Acoustics*, New York: American Institute of Physics, 1990.
2. P. G. Peterson and E. E. Gross, *Handbook of Noise Measurement*, West Concord, MA: General Radio, 1978.
3. C. M. Harris, *Handbook of Noise Control*, New York: McGraw-Hill, 1986.
4. J. R. Hassal and K. Zaveri, *Acoustic Noise Measurement*, Marlborough, MA: Brüel and Kjær Instruments, 1988.
5. L. L. Beranek, *Noise and Vibration Control Engineering*, New York: Wiley, 1992.
6. ANSI S1.1-1960—American Standard Acoustical Terminology.
7. ISO 3740-3746—Determination of sound power levels of noise sources.
8. S. Gade, Sound intensity, in *Brüel and Kjær Technical Review Vol. 3, 4*, Marlborough, MA: Brüel and Kjær Instruments, 1982.
9. ISO R226-1961—Normal equal-loudness contours for pure tones and normal threshold of hearing under free-field listening conditions.
10. L. L. Beranek, *Acoustical Measurements*, New York: American Institute of Physics, 1988.
11. IEC 651-1979—Sound-level meters.
12. ANSI S1.4-1983—American National Standard Specification for sound level meters.
13. IEC 804-1985—Integrating sound level meters.
14. ANSI S1.25-1978—American National Standard Specification for personal noise dosimeter.
15. R. B. Randall, *Frequency Analysis*, Marlborough, MA: Brüel and Kjær Instruments, 1977.
16. IEC 225-1966—Octave, half-octave and third octave band filters intended for the analysis of sound and vibrations.
17. E. O. Doebelin, *Measurement Systems: Applications and Design*, 4th ed., New York: McGraw-Hill, 1990.
18. P. V. Brüel and G. Rasmussen, Free field response of condenser microphones, *Brüel and Kjær Technical Review*, Vol. 1 and 2, Marlborough, MA: Brüel and Kjær Instruments, 1959.
19. W. R. Kundert, Everything you've wanted to know about measurement microphones, *Sound and Vibration*, pp. 10–26, March 1978.
20. ANSI S1.40-1984—American National Standard Specification for acoustical calibrators.

GIOVANNI BETTA
University of Cassino
ANTONIO PIETROSANTO
University of Salerno

FLOWMETERS

FLOW MEASUREMENTS

The measurement of flow is important in many industrial, scientific, and research applications. Flowmeters are used to measure the quantity of moving fluids or solids in open or closed conduits. A large number of different flowmeters are available, and the instrument selected for a particular measurement depends on the physical variables and the cost. Historically, flowmeters were first used for open channel flow measurement of the most common fluid, water. In many industrial applications, the ability to accurately measure flow is important because these measurements can be related directly to overall profitability of the plant. In recent years, the advent of microelectronics in flowmeters has improved their accuracy, ease of use, and maintainability significantly. Nowadays, with the widespread use of microprocessors, intelligent flowmeters are being produced. They allow versatile communication capabilities, automatic compensations for temperature and pressure, alarm facilities, selection of display capabilities, and so on.

Applications of flowmeters may include liquids, vapors, gases, solids, or any combination of these. In some cases, the measurement at a point may be necessary, whereas in the others, measurement of volume flow rates or mass flow rates may be needed. Often, measuring flow with high accuracy over a wide range of operating conditions is necessary. Flow measurements are based on many different physical principles. Some of those principles may include conductivity; capacitive, radiation, thermal, optic, and sonic characteristics; optics and lasers; and various mechanical methods.

Flowmeters are used most often for gas and liquid flow measurements. They can be categorized as either intrusive or nonintrusive. Intrusive flowmeters make contact with the flowing material with either moving parts or nonmoving parts. Positive displacement, variable area, and turbine flowmeters are examples of intrusive flowmeters with wetted moving components. Orifice plate, oscillatory, target, and thermal flowmeters are examples of intrusive types with wetted but no moving components. Nonintrusive flowmeters such as magnetic and coriolis flowmeters do not disturb the material flow, but they still have wetted parts. A subsection of nonintrusive flowmeters has no wetted components, such as clamp on ultrasonic flowmeters or some optical types.

Flowmeter selection depends on the physical requirements of the measurements. They could be employed to measure the volumetric flow, velocity, or mass. For example, a positive displacement flowmeter measures the volume flow. Whereas, magnetic, oscillatory, turbine, or ultrasonic flowmeters measure velocities from which the total flow can be determined. Coriolis and thermal flowmeters measure the mass directly. Apart from these, some flowmeters, such as differential pressure, target, and variable area flowmeters, may be classified as inferential types because they infer the flow through some physical phenomenon (e.g., pressure differences).

In many cases, the selection and application of flow-measuring devices require further information on variables such as pressure, density, viscosity, and temperature in order to obtain an accurate output from the instrument. The information about temperature and pressure is particularly important in gas flow measurements. Most commercial gas flowmeters specify flow ratings in volume at standard conditions of

1 atm and 20°C. However, information on viscosity may carry greater importance in liquid flow measurements, and density is important in solid flow applications. Accuracy of flowmeters is of utmost importance, especially when they are used for measuring expensive fluids like petroleum products. In order to effect correction for pressure and temperature, flowmeters are used in conjunction with flow computers when used for measurement of petroleum products in liquid and gaseous form.

Characteristics of Liquids and Gases

Although liquids and gases are different in many respects, they change shapes in a similar manner under the influence of a deforming force. In general, most liquids cannot be compressed. That is, the change in the volume liquid under pressure may be negligible. Because gases are easy to compress and their densities depend on both temperature and pressure, the compressibility of gases must be considered carefully in flow measurements.

During the flow measurements, the density of a gas may vary significantly depending on the absolute pressure. An increase in the pressure of a gas at constant temperature causes the gas to be compressed to a smaller volume of the same mass. Boyle's Law states that for ideal gases or mixtures of ideal gases at constant temperature, the volume V is inversely proportional to the absolute pressure, that is

$$V = \text{Constant}/P \quad (1)$$

Equation (1) may be rewritten to compare the volumes of an ideal gas at a constant temperature but different pressures P :

$$V/V_0 = P_0/P \quad (2)$$

Charles's Law states that the density of gas will vary significantly with absolute temperature T . Increasing the temperature of a gas at constant pressure causes the gas molecules to increase their activities and motions in relation to each other, thereby increasing the volume and decreasing the density of gas for the same mass. Charles' Law may be stated in the following form:

$$V/V_0 = T/T_0 \quad (3)$$

Charles' and Boyle's Laws can be combined to yield the Ideal Gas Law:

$$PV = nRT \quad (4)$$

where R is the universal gas constant in consistent units and n is the number of moles.

The Ideal Gas Law can also be expressed in the following form for different temperatures and pressures:

$$V/V_0 = (TP_0)/(T_0/P) \quad (5)$$

During gas flow measurements, if the variations in pressure and temperature are small, the temperature and pressure act almost independently of each other. Thus, estimates of reasonable flow accuracy can be obtained by adding percentage temperature and pressure deviations from a given set of conditions.

Corrections are necessary if the temperature and pressure of gas is changing during the process of flow measurements.

Often measuring the mass or volume flow of nonideal gases is necessary simply because they do not act like ideal gases at certain conditions, such as at high pressures or low temperatures or under saturation. Their nonideal behavior may be accounted for by modifying the Ideal Gas Law with a Z factor:

$$V/V_0 = (TP_0Z)/(T_0PZ_0) \quad (6)$$

The Z factor is numerically dependent on operating conditions and can be read from generalized compressibility charts with a reasonable degree of accuracy.

For these reasons, extra care and thought are necessary when measuring gas flow.

In the case of liquid and solid flow measurements, the volumetric effect of temperature on the density of liquid or solid may be expressed as

$$V = V_0(1 + \beta \Delta t) \quad (7)$$

where β is the coefficient of solid or liquid expansion, which is consistent with the temperature units used.

Because the mass is the same before and after the temperature rise, the change in density is inversely proportional to the change in volume and can be expressed as

$$\rho/\rho_0 = V/V_0 \quad (8)$$

Nevertheless, in many applications, the effect of volumetric changes caused by temperature are negligible.

An important factor in the liquid flow measurements is the determination of velocity, which indicates the speed and direction of the flow. When the average velocity is slow, the flow is considered to be laminar. This indicates that material flows in layers with the slower moving layers in the outer edges. As the velocity increases, the flow may become turbulent with the layers disappearing and the velocity across the flowing stream being more uniform. In this case, the term *velocity* refers to the average velocity of particular cross section of the stream.

The nature of the fluid flow can be described by the non-dimensional Reynold's number. The Reynold's number R is found by

$$R = vD\rho/\mu \quad (9)$$

where v is the velocity, D is the inside diameter of the pipe, ρ is fluid density, and μ is the kinematic viscosity.

The Reynold's number is significant to determine whether the flow is laminar or turbulent. Correctly predicting the Reynold's number leads to accurate measurements when appropriate correction factors are applied. Generally, if R is less than 2000, the flow is laminar; if R is greater than 4000, the flow is turbulent. Between these two values, the flow is in the transition region.

MASS FLOW MEASUREMENTS

Mass flowmeters are used extensively in oil and gas, iron and steel, power, water distribution, pharmaceutical, food and drink, and other industries. The techniques employed for mass

flow measurements can be classified to be indirect or direct methods. In indirect methods, the velocity or volumetric flow of fluid is measured by a suitable flowmeter, and the necessary calculations are made by considering such process conditions as density, volume, temperature, and pressure. Some of the devices that can be used in indirect measurements may be venturi, target, vibrating vane, ultrasonic, or electromagnetic flowmeters. The most commonly used direct methods using coriolis and thermal mass flowmeters and will be discussed next.

Coriolis Flowmeters

The coriolis flowmeter is a type of direct mass flowmeter where the output depends on the mass flow rate of fluid passing through. Coriolis flowmeters are used to measure harsh chemicals, low- and medium-viscosity fluids, foods, slurries, and the like. They find limited applications in gas technology because the density and pressure of gases are relatively low. Coriolis flowmeters can also be constructed and configured to measure the volumetric flows and the densities of liquids or gases.

The principle of the flowmeter depends on the conservation of angular momentum, such that if a mass m is moving with a velocity v along a rod, the rod itself is moving with an angular velocity ω , and the mass experiences a Coriolis force of magnitude

$$F = 2mv\omega \quad (10)$$

The direction of the coriolis force is perpendicular to both linear and angular velocities. The coriolis force is created as a result of coriolis acceleration acting on the fluid as the fluid changes position in relation to the movement of the tube. The coriolis flowmeter consists of a vibrating tube, as shown in Fig. 1, in which the coriolis force is created and measured. This force results in the a twist of the tube, and this twist is sensed optically to indicate the mass inside the tube.

Coriolis flowmeters are largely applied to measure fluid flow. These flowmeters can measure flow rates from 0.04 kg/s to 15 kg/s in pipes with diameters of 3 mm to 50 mm. Accuracy

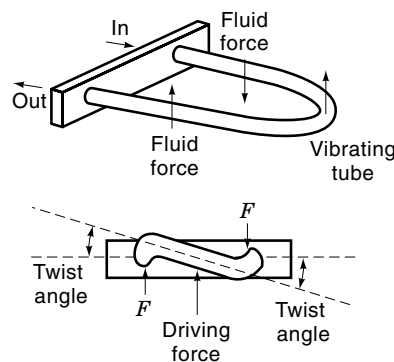


Figure 1. A coriolis flowmeter. The flowmeter consists of a vibrating tube carrying a portion of fluid flowing in the pipe. A force is created as a result of coriolis acceleration acting on the fluid as the fluid changes position in relation to movement of the tube. The direction of this force is perpendicular to both linear and angular velocities. Coriolis force results in the twist of the tube, which can be sensed magnetically or optically to indicate the mass inside the pipe.

better than 0.5% is possible between 5% and 100% of maximum flow rates.

The tube of coriolis flowmeters is constructed from stainless steel and vibrated by a drive mechanism. In a harsh environment, the inside of the tube is lined with Teflon to minimize corrosion and abrasion. Coriolis flowmeters can operate at high temperatures, but they are sensitive to pressure drops in the pipes. The accuracy of these flowmeters is in the region of 0.2% to 0.5% over the designed accurate measurement range. They are manufactured in different ranges and sizes. The selection of these flowmeters depends on the fluid characteristics, pressure drops in the piping systems, liquid friction losses, and so on. Manufacturers usually supply instructions, procedures, and graphs for particular applications. When installed, coriolis flowmeters must be oriented such that the meter is completely full of liquid that is free from air or gas bubbles. In gas applications, care must be exercised to not introduce liquid or condensed gas in the flowmeter.

Thermal Flowmeters

Thermal flowmeters use thermal properties of fluids or gases. Depending on the flowmeter design, they can measure velocity or mass flow. There are many different types of which the thermal dispersion, hot wire anemometer, and constant-heat infusion types are the most widely used flowmeter.

The thermal dispersion flowmeter is based on the heat loss from a heated element placed in the flow path, as illustrated in Fig. 2. The element is heated above the ambient temperature and maintained at constant power. The cooling of the element is proportional to the mass flow, which can be described by King's Law as

$$P = A(T_e - T_f)(C_1 + C_2 v^{1/2}) \quad (11)$$

where P is the heat transfer (W); A is the effective area of the heated element (m^2); T_e and T_f are temperatures of heated element and fluid, respectively; C_1 and C_2 are constants; and v is the local speed of the fluid.

The temperature difference is converted to a linearized mass flow rate signal within the electronics.

Constant-heat infusion flowmeters are used to measure the average mass flow by externally injecting a known amount of heat into the flowing liquid, as shown in Fig. 3. The average

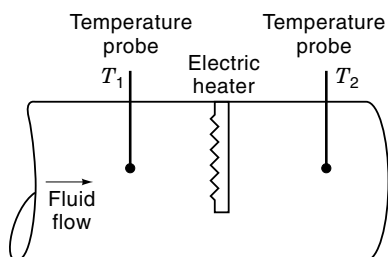


Figure 2. A thermal dispersion flowmeter is based on the heat loss from a heated element in contact with the flowing fluid. The element is heated above the ambient temperature, and the heating is maintained at constant power. Thus the cooling of the element is proportional to the mass flow inside the pipe. These flowmeters are suitable for gas and liquid flow measurements.

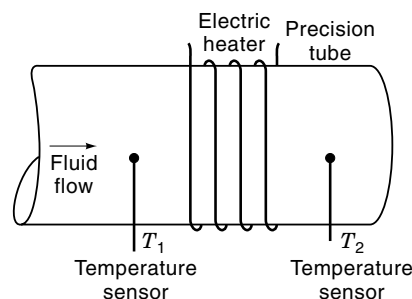


Figure 3. A constant-heat infusion flowmeter is based on the heat injected into the liquid, and the temperature is measured up and down the stream. In some cases, the temperature difference is maintained automatically, giving flow rate as a function of heat injected to the flow stream.

mass flow can be calculated by measuring the temperatures upstream and downstream:

$$F = q/c_f \Delta T \quad (12)$$

where F is the mass flow (kg/s), q is the rate of heat added (W), c_f is the specific heat of fluid (J/kg K), and ΔT is the temperature difference measured upstream and downstream in Kelvin.

In one version of the constant-heat infusion flowmeter, a control loop keeps the temperature difference constant. The resulting power input is a linear function of mass flow.

Hot wire anemometers use temperature rise in the flowing liquid. They have probes inserted into the flow stream, as shown in Fig. 4, to measure the temperature of the fluid. The velocity sensor is a fine metal wire made from tungsten, platinum, or nickel. In one version, the current through the sensor is kept constant, and the resistance change is then a measure of local fluid velocity. In the other version, resistance is kept constant, and the change in the input power is related to the local velocity. The heat from the heated probe is removed depending

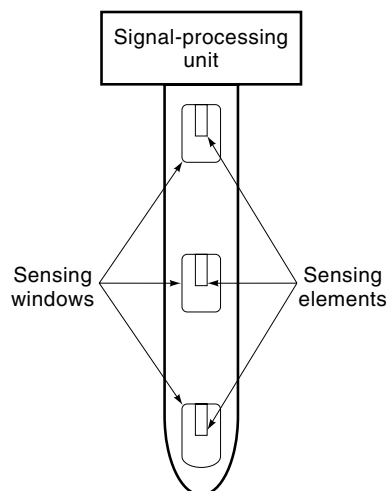


Figure 4. A hot wire anemometer uses probes inserted into the flowing stream. Current through the sensor is kept constant, and the resistance change is monitored; or resistance is kept constant, and the change in the input power is monitored. These devices are suitable for measuring local flow velocities of fluids.

on the flow rate of the fluid. In some cases, a capillary tube is uniformly heated, and the temperature dissipation is sensed at different points of the tube; thus with a constant power input, the difference in the temperature at the sensing points is a linear function of the mass flow of the fluid.

Thermal flowmeters are used in applications where other types of flowmeters are not suitable, such as where fluids are not dense enough for mechanical sensing. They are usually applied to clean pure gases or mixtures of gases with predictable compositions. Typical applications are air, nitrogen, oxygen, CO₂, and methane. The body of the sensing system is usually constructed from stainless steel, brass, monel, or aluminum. Probes that usually come in replaceable tips are manufactured by using stainless steel. The thermal sensors or heaters are located in the probes. The capillary tubes are usually constructed from constantan to maintain a good thermal conductivity. They can operate at high-temperature applications (e.g., 450°C). The accuracy is typically $\pm 2\%$.

Mass Flow Measurement of Solids

Often mass flow rate measurements of solids are necessary where processing of granular or powdered materials are involved. In some systems, solid materials are allowed to fall from a fixed height from suitable slots on to sensors that can indicate the flow rate. In other systems, the capacitive properties are used as the solids pass through a number of electrodes suitably arranged on the flow path. However, weighing systems are most widely used in solid mass flow measurements. Load cells are often mounted under the conveyor systems to monitor continuously the solids on the conveyor, as illustrated in Fig. 5. The belt speed is sensed by a tachogenerator giving

$$Q = mv/l \quad (13)$$

where Q is the mass flow rate, m is the load cell output, v is the belt velocity, and l is the length of the belt being weighed.

An alternative technique uses isotopes emitting gamma rays. The transmitter and the receiver are placed on the opposite sides of the belt, and the gamma rays are allowed to penetrate through the solid materials. The rate of absorption of the gamma rays is directly proportional to the solid mass on the belt. The flow rate is then calculated by taking into account the belt speed, density, and temperature.

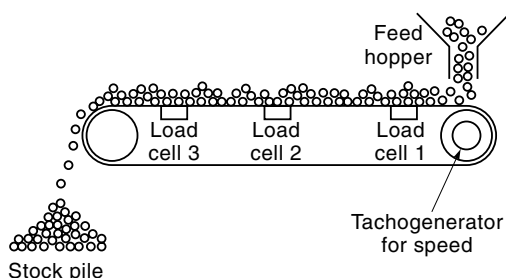


Figure 5. A solid flow-measuring arrangement. A number of load cells are located under the conveyor belt carrying solids. The output of the cells and the conveyor speed indicate the flow rate of solids. In some cases, instead of load cells, gamma ray devices are used, and attenuation through the solids is monitored.

VELOCITY AND VOLUMETRIC FLOW MEASUREMENTS

In many applications, measuring velocity may be necessary, whereas in others volumetric flow is more important. Some flowmeters will give a direct indication of the volume flow (e.g., positive displacement), and others will measure the velocity (e.g., ultrasonic, magnetic, turbine) from which volume flows can be determined. Inferential flowmeters (e.g., differential pressure, target, and variable area) are also used for velocity and volume measurements. These flowmeters can be classified as mechanical including differential, electrical, sonic, or optical types. Here, operational principles and applications of commonly used mechanical flowmeters together with differential pressure types will be discussed first, followed by optical, electrical, and sonic types.

Turbine Flowmeters

Turbine flowmeters, sometimes called fixed-vane flowmeters, are typical mechanical flowmeters. They are known to be highly accurate with good repeatability, particularly for turbulent flow applications. There are many different types such as axial, propeller, and paddle wheel.

The most frequently used flowmeter is the axial type shown in Fig. 6. A multiblade rotor is positioned in the flow stream with the axis of rotation parallel to the direction of the flow. The fluid impinges on the blades causing a rotation with an angular velocity approximately proportional to flow rate. The number of machine cycles per second can be related directly to the volume flow rate by using a constant as the meter factor. Generally, turbine flowmeters are highly nonlinear at low flow rates, and they are not suitable to measure flows that reverse in direction.

In some cases, dual rotors are used to eliminate the errors inherent in the single-rotor designs. The rotors are designed to be as light as possible so that the momentum of fluid is relatively larger. In this way, the rotor can respond rapidly to the changes in the velocity of fluid. The rotation of the rotor is generally sensed magnetically or by tachogenerators. Other techniques such as radiofrequency (RF) methods are also used for sensing rotation.

Turbine flowmeters are manufactured in different ranges and sizes to measure liquid flows from a few centiliters to

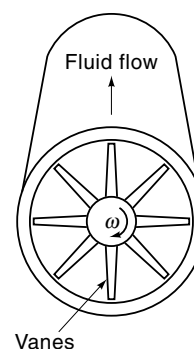


Figure 6. A turbine flowmeter consists of a multiblade rotor positioned suitably in the flowing stream of fluid. The fluid impinges on the blades and causes the rotor to rotate with an angular velocity proportional to the flow rate. The rotation of the rotor is sensed mechanically or electrically for further signal processing.

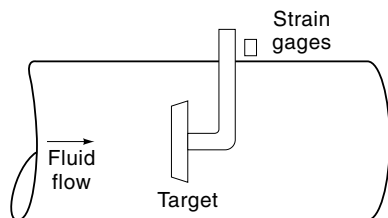


Figure 7. Target flowmeters are used for gas and liquid flow measurements especially in large pipes. A target suspended in the flow stream experiences a force proportional to the square of the velocity of the fluid. The force on the target is usually picked up by strain gauges.

200,000 L/min. The accuracy of these meters is typically about $\pm 0.5\%$. The pressure drop range in gases is from 0.2 psi to 90 psi and in liquids from 1 psi to 20 psi. They are to be installed as per standards in order to achieve the levels of accuracy claimed. Orifice flowmeters are being replaced with these meters in many petroleum and natural gas metering stations.

Target Flowmeters

Target flowmeters, also named as drag flowmeters, are convenient to use for liquid and gas flow measurements in large pipes. They operate on the principle of measurement of force exerted on body. A target is suspended, as shown in Fig. 7, in the flow stream, and the force on that target is the difference between the upstream and downstream pressures, which can be expressed as

$$F = k\rho Av^2 \quad (14)$$

where k is a constant, ρ is the density of fluid, A is the target area, and v is the velocity of the fluid.

The force on the target is sensed by a force balance system or strain gages. The accuracy of these devices is relatively inferior, in the region of $\pm 1\%$ to 2% full scale, mainly as a result of the square root relationship between the flow rate and drag force. These devices are usually applied in turbulent flows with sufficient momentum to exert force on the target. In laminar flow applications, the fluid velocity becomes a function of fluid viscosity, which may necessitate the use of calibration curves. They are sensitive to installation effects requiring straight runs of piping.

Positive Displacement Flowmeters

Positive displacement flowmeters are based on the principle of entrapping a known amount fluid as it passes through the flowmeter. The entrapped fluid rotates the lobes whose rotation in turn generates electrical pulses. Many types of flowmeters such as helical gear, nutating disc, oscillating piston, oval gear, piston, and rotary vane positive displacement flowmeters are available. A typical example of this flowmeter is shown in Fig. 8. The flow of fluid through a volume of fixed size causes rotation of an output shaft. This rotation can be sensed by many methods including optical techniques.

Positive displacement flowmeters are sensitive to viscosity of the fluid. Also, maximum allowable pressures and flow rates are limited. Although they are accurate over a wide flow ranges

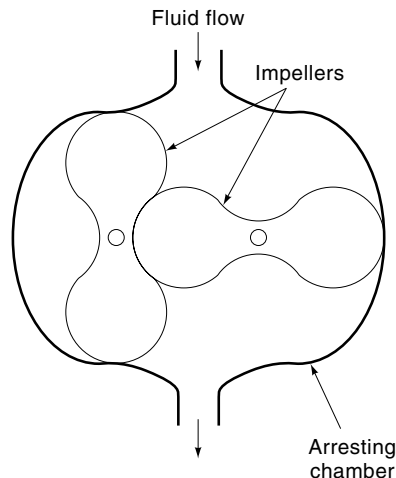


Figure 8. A positive displacement flowmeter is based on the principle of entrapping a known amount of flowing fluid. There are many different types available, and they are often used in water-metering applications.

(e.g., 1000:1), friction, inertia, and fluctuating flow rates can cause serious errors. Also minimum measurable flow may also be limited as a result of leakage between the lobe and casing. They have accuracy of the order of $\pm 0.1\%$. These are extensively used in the petroleum industry, especially for tanker loading.

Variable Area Flowmeters

In these flowmeters, a float, which is dynamically balanced by the flowing liquid, is positioned in the flow line, as illustrated in Fig. 9. The upward force on the float as a result of liquid velocity is equal to the weight of the float less the weight of the liquid that it displaces. An increase in the flow makes the metering tube rise. As the float rises, the annular area between the float and the metering tube increases until the upward and downward forces are dynamically equalized. Thus the level of the float is an indication of the flow through the flowmeter.

The metering tubes are available in different sizes and shapes to suit specific application requirements. Often the motion of the float is magnetically sensed. These devices are suitable for pipes that are less than 10 cm in diameter. Pressure of operation is limited by the glass material of the tube. Their accuracy is $\pm 1\%$. Variable area flowmeters are suitable for all kinds of gas and clean liquid applications.

Insertion Flowmeters

In insertion flowmeters, transducers such as differential pressure, magnetic, oscillatory, target, thermal, turbine or ultrasonic are inserted in the flow stream, as shown in Fig. 10. A representative velocity at a critical point is measured to represent the average velocity in the flow stream. After the average velocity is determined, the volumetric flow in the pipe can be calculated. In the applications, piping effects must carefully be considered. These devices are most suitable in determining local velocities of the flowing liquids or gases. Errors in the positioning of the transducers can cause serious errors in measurements.

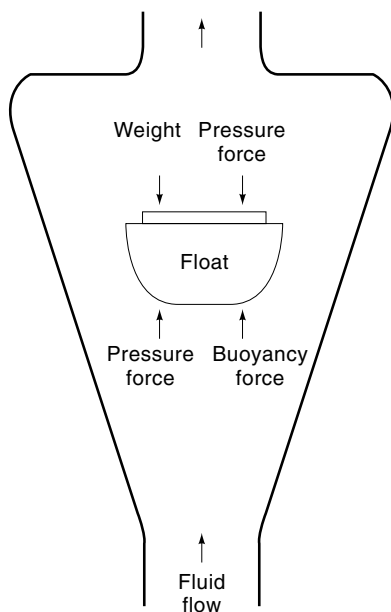


Figure 9. A variable area flowmeter uses a float dynamically balanced by the flowing fluid. An increase in the flow rate increases the level of the float. The level of the float is sensed mechanically, electrically, or optically. They are suitable in gas and clean liquid applications because contaminated liquids may block the flow path between the float and the container.

Differential Pressure Flowmeters

Differential flowmeters are used extensively for liquid and gas flow measurement applications. It is based on empirical correlation of the relationship between the differential pressure and volumetric flow through a restriction pipe. True flow rate is generally determined by weighing or volumetric collection of fluid over a measured time interval. Theoretical flow is calculated from flow equations using the measured differential pressure and known properties of the fluid. There are many types of flowmeters such as pitot tube, orifice plate,

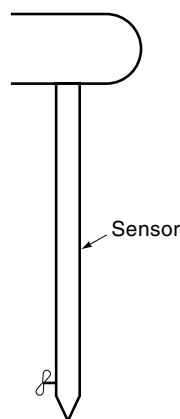


Figure 10. An insertion flowmeter contains various types of sensors, such as magnetic or differential pressure, that are inserted in the flow stream and that sense the local velocities of fluids. The representative velocity is then averaged to find the volumetric or mass flow rates.

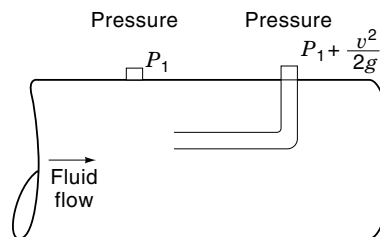


Figure 11. A pitot tube flowmeter inserted in the flow stream arrests fluid flowing in the pipe. The head of the fluid represents the potential energy stored in the tube, which can be related to the velocity. Pitot tube devices are well-established flowmeters, and there are many different versions.

elbow, flow-nozzle, segmental wedge, V-cone, venturi, and bypass flowmeters. Here only a selected few will be discussed.

Pitot tube devices are typical examples of differential pressure flowmeters. This is used for point velocity measurement. A simple example is shown in Fig. 11. The liquid moving with a velocity v is arrested in the tube inserted into the flowing liquid. By using the conservation of energy, the velocity of the fluid may be expressed as a function of head as $v = \sqrt{2gh}$. There are many versions of pitot tubes including multiple openings to the fluid and complex differential pressure transmitters and other complex gauge arrangements. For measurement of flow rate, the Pitot tube is traversed through the pipe diameter and the point velocities are measured. The velocity is integrated and multiplied by the flow area to get the flow rate.

Orifice plate flowmeters operate on Bernoulli's Principle, which states that the sum of the static energy, the potential energy and kinetic energy is conserved in a restriction pipe that is carrying fluid as shown in Fig. 12. Here the fluid flow rate of an incompressible fluid may be written as

$$Q = A_1 v_1 = A_2 v_2 \quad (15)$$

The velocity of fluid increases as the cross-sectional area of the pipe is reduced. Applying Bernoulli's equation to the upstream and downstream location of the orifice gives the pressure difference as

$$P_1 - P_2 = \frac{1}{2} \rho [(D/d)^4 - 1]^2 Q^2 / A_1^2 \quad (16)$$

Equation (16) indicates that the differential pressure generated across an orifice is proportional to the square of the flow through the orifice plate.

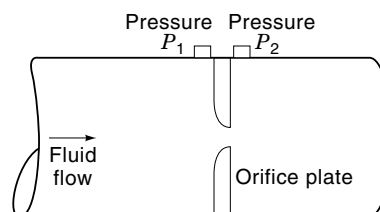


Figure 12. Orifice-type flowmeters are most widely used, and they operate on Bernoulli's Principle. The potential energy and kinetic energy indicate the flow rates in a restricted section of the pipe. These devices have accuracy levels of 0.5%.

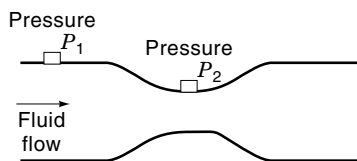


Figure 13. A venturi flowmeter has a restriction throat. The pressure drop before and in the restriction throat is measured to be related to the flow rate inside the pipe. Venturi flowmeters are accurate and widely used in industry for liquid and gas flow measurements.

The accuracy of orifice plate flowmeters is in the range of 0.5% to 3%. They are suitable in applications with Reynold's numbers from 5000 to 2×10^6 depending upon the size; however, they are highly sensitive to installation effects. Valves, pipe fittings, that can distort the velocity profile leading to inaccurate readings. When installed as per International Standard Organization (ISO) standards, higher accuracies can be obtained. They were used until recently in conjunction with flow computers for custody transfer flow measurement of natural gas.

Venturi flowmeters are a type of obstruction device, as shown in Fig. 13. This classical Herschel venturi has a very long flow element with tapered inlet and diverging outlet. Inlet pressure is measured at the entrance section. The inlet diameter is reduced to a throat section, and the static pressure is measured at this section. Venturi flowmeters find applications in fluid flows having a Reynold's number as high as 100,000. These flowmeters give relatively small pressure drops and are highly accurate, about 0.5%. They are not affected by velocity profiles. The construction and applications of venturi devices have been standardized by international standards (e.g., ISO 5167 and DIN 1952). These flowmeters find many applications in power, water, and water treatment industries.

Bypass flowmeters are obtained by employing a bypass stream created by differential pressure in the main stream. A flowmeter in the bypass stream measures the flow as illustrated in Fig. 14. The flow in the bypass stream is then inferred to the main stream. The accuracy of the system is limited by the accuracy of the bypass flowmeter. In many critical applications, the bypass flowmeter is used to increase the accuracy of the primary flowmeter, which is located in the main stream.

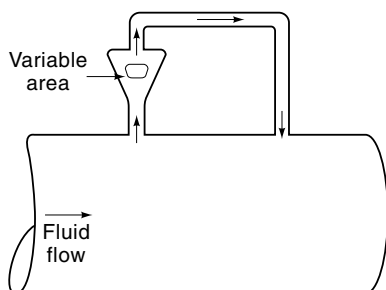


Figure 14. A bypass flowmeter is obtained by providing a bypass line to the main flow stream. The pressure drop of the bypass stream is inferred to the flow characteristics of the main stream. These flowmeters are often used as backup measurements in conjunction with main flowmeters in some critical and sensitive applications.

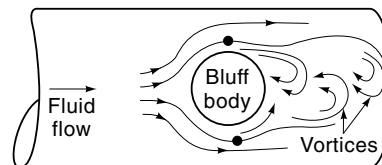


Figure 15. A vortex-shedding flowmeter is based on the creation of vortices in the flow stream. A suitably shaped bluff body is placed in the flow stream, and the frequency of vortices is measured by various techniques such as pressure or ultrasonic methods. They are suitable for applications using low-viscosity fluids.

Vortex-Shedding Flowmeters

Mathematically, the phenomenon of vortex shedding is described by the von Karman Effect. When a bluff body is placed in a flow stream, as the fluid passes the body, vortices are shed alternately from the back side of the body. The vortex shedding of the device is illustrated in Fig. 15. The frequency of the vortex shedding is directly proportional to the velocity of liquid. The frequencies of vortex shedding are monitored by appropriate sensors. The fluid parameter that describes the operation of this device is a nondimensional Strouhal number S

$$S = f_s d / v \quad (17)$$

where f_s is the vortex shedding frequency, d is the diameter of the bluff body, and v is the velocity.

Writing

$$Q = Av \quad (18)$$

where A is the flow area, and substituting for v gives

$$Q = A f_s d / S \quad (19)$$

The Strouhal number is constant for particular shapes of bluff body. For the body in Fig. 15 the number is about 0.88 for Reynold's numbers from 10^4 to 10^6 .

To measure the frequencies of the vortices, various forms of sensors are used. These sensors include diaphragms, magnetic sensors, pressure sensors, suitable torque tubes, and ultrasonic methods. Measurements by vortex shedding can indicate instantaneous flow rates or totalized flow rates over a known time interval. The meters are calibrated by the manufacturers for specific pipe sizes. They give the best results for low-viscosity fluids and unsuitable for most high-viscosity liquids. Manufacturers specify special installation requirements of their devices.

These flowmeters are available in different sizes, and they can handle from a few liters of flow to 15,000 L. They are used in liquid, gas, and vapor applications with relatively steady flow rates. The overall accuracy for gases is in the region of 1% to 2%. Accuracy in liquid applications is about 0.5%.

There are many different versions of devices based on vortex-shedding principle, including some devices using two bluff bodies to make stronger vortices. An improved version of the vortex-shedding transducers is the velocity probes. In this case, sensors are mounted together with the bluff body in a section of a short tube. This assembly can be inserted in various sections of the flow field to measure the local velocity.

OPTICAL METHODS

Optical methods are extensively used in flow measurements involving analytical applications. The properties of light are used to measure the flow of liquids and gases. Some of these properties may include shadowing, relative illuminations, deflections, interference, and Doppler shift. There are many different flowmeters based on optical principles. Flowmeters based on the laser Doppler effect will be discussed next.

Laser Doppler Anemometers

Like Pitot tubes, laser Doppler anemometers are also used for point velocity measurements. In laser Doppler anemometers, the properties of scattering of a laser beam are utilized. For the correct operation of the device, the fluid must contain particles to scatter laser beam. A typical example of such devices is shown in Fig. 16. The laser beam is focused on a small volume element, and the scattered beams are received by an appropriate arrangement of lenses. The scattered light experiences a frequency shift directly proportional to the velocity. The scattered and unscattered beams are combined through a beam splitter. The resultant beam is put through a photomultiplier tube to give an output proportional to flow velocity. The data from the photomultiplier is processed carefully by employing appropriate signal processors.

He-Ne gas lasers operate at a frequency of 5×10^{14} Hz (wavelength of 632.4 nm) and are preferred over argon lasers. There are laser flowmeters that measure more than one velocity component simultaneously. Laser techniques are useful to determine turbulence and other flow phenomena. Focused laser beams can measure flows of samples as small as tens of cubic micrometers. The use of this is generally limited to laboratory studies related to point velocity measurements. Moreover, it requires the pipe carrying fluid to be transparent to the laser beam.

ULTRASONIC FLOWMETERS

In ultrasonic flowmeters, acoustic waves detect the flow of liquids and gases in pipes. There are a few different types, such as the Doppler effect, transit time, correlation, and ultrasonic vortex types.

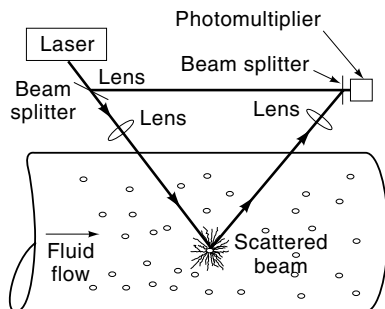


Figure 16. A laser Doppler anemometer is based on scattered laser beams. As the laser light encounters particles in the flowing fluid, it gets scattered. Some of the scattered light is picked up by suitably located lenses. The frequency shift of the returned light depends on the velocity of particles in the fluid.

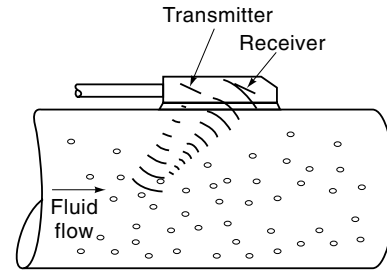


Figure 17. A Doppler effect ultrasonic flowmeter is based on the reflection of ultrasonic beams from particles or air bubbles carried in the flowing fluid. For signal processing, proper amplification, filtering, modulation, and demodulation of returned signals is necessary.

In *Doppler effect flowmeters*, ultrasonic signals of a fixed frequency are transmitted through the flowing fluid, and the frequency shift of the returned signals is measured. The transmitted signals can be continuous or in pulse-modulated form. Such impurities as solids, bubbles, or any other discontinuity in the liquid reflect the signals back to the transmitter. A simplified example of this flowmeter is depicted in Fig. 17. The frequency of the reflected signal shifts from the transmitted frequency in proportion to the velocity. The frequency f_r of reflected signals can be expressed as

$$f_r = f(c + v \cos \theta)/c \quad (20)$$

where f is the frequency of transmitted ultrasonic wave, c is the velocity of sound in the still fluid, v is the relative velocity of fluid with respect to transmitter, and θ is the angle of transmission with respect to direction of fluid flow.

Particles reflecting the ultrasonic signals act as a source moving with a velocity v relative to the receiver. Now the frequency of reflected signals to the receiver may be expressed as

$$f_{rr} = f_r c / (c - v \cos \theta) \quad (21)$$

It can further be proved that the frequency shift between the transmitted and received signal is

$$\Delta f = f_{rr} - f = (2fv \cos \theta)/c \quad (22)$$

The particles scatter the transmitted signal in all directions; therefore, the received signals by the receiver are attenuated severely. During the signal-processing stages, the receiver signals are amplified to the level of transmitted signal and added to the transmitted signals. This gives an amplitude modulated signal with a carrier frequency of $(f + f_{rr})/2$ and a modulating frequency of $\Delta f/2$. Rectification of the output of the adder and low-pass filtering gives a demodulated signal proportional to fluid velocity v .

The *transit time* ultrasonic flowmeters are based on the measurement of the difference in travel time between pulses transmitted along or against the fluid flow in a pipe. A typical arrangement is shown in Fig. 18. Transducers 1 and 2 act as transmitters and receivers. As transmitter 1 sends a pulse, the corresponding transit time may be expressed as

$$t_d = L/(c + v \cos \theta) \quad (23)$$

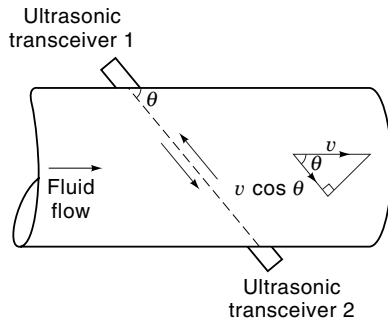


Figure 18. A transit time ultrasonic flowmeter uses the time of flight between two transducers. The two transducers are used as transmitters and receivers alternately. The transit time difference can be directly related to the flow rate of the fluid. Because the time difference is small, careful signal processing is necessary for accurate measurements.

where t_d is time taken for pulse to travel from transducer 1 to 2, L is the distance between transducers, c is the velocity of sound in still fluid, v is the velocity of the fluid, and θ is the angle between transducers with the pipe.

When transducer 2 sends a pulse, the corresponding transit time becomes

$$t_u = L/(c - v \cos \theta) \quad (24)$$

From these two equations, the differential time ΔT can be found as

$$\Delta T = t_d - t_u = (2Lv \cos \theta)/c^2 \quad (25)$$

The differential time ΔT is very small, of the order of a few hundred nanoseconds. Therefore, carefully designed electronic circuits are necessary for a reasonable accuracy. In some systems (e.g., sing around systems) transmitters continuously switch between transmitter and receiver modes, which leads to better accuracy. There are two different types—wet and clamp-on. In the wet type, the transducers are inserted in the stream making direct contact with the fluid. In the clamp-on type, the transducers are attached to the pipes externally. The clamp-on types are used for rough estimate of flow in pipes of large diameters.

Ultrasonic *cross-correlation flowmeters* consist of two transducers sensing fluctuations in the properties of the fluid. For example, the properties of the fluid may be variations in the acoustic impedance resulting from the presence of particles and air bubbles. The variations in the sensing of the fluid properties are randomly changing but present in each sensor. Therefore, the delayed version of the output signal $x(t)$ of sensor 1 is correlated with the output $y(t)$ of the sensor 2 through a correlator to determine the cross-correlation function. The output of the correlator has a maximum value proportional to L/v against delayed versions of $x(t)$. The received signals are amplified, demodulated, and filtered to eliminate the high-frequency carrier signal. The remaining demodulated signals are amplified and input to the correlator to determine the flow rates. There are many other cross-correlation flowmeters based on different techniques and sensors such as capacitance transducers, electrodes, and infrared or optical sensors. The fluid properties that

may be used for measurements can be densities, temperatures, conductivity, acoustic impedance, and the like.

Ultrasonic transducers are available for gas and liquid flow measurements. The accuracy of these devices can vary between $\pm 0.5\%$ and 10% of full-scale (FS). The liquid must contain solid particles or air bubbles in the stream for Doppler-type flowmeters to operate effectively. The amount of solid particles must not exceed a certain limit (e.g., 30%) so that ultrasonic energy can penetrate into the stream for the measurement of average velocity. Also, for accurate results, the particles must also flow with the same speed as the liquid. The time-of-flight type of ultrasonic transducers are more suitable for gas applications.

ELECTROMAGNETIC FLOWMETERS

Magnetic flowmeters have been widely used in industry for many years. Unlike many other types of flowmeters, they offer true noninvasive measurements. They can measure reverse flows and are insensitive to viscosity, density, and flow disturbances. Electromagnetic flowmeters can rapidly respond to flow changes and are linear devices for a wide range of measurements. Recently, the technological refinements have resulted in much more economical, more accurate, and smaller instruments than the previous versions.

The underlying principle of the electromagnetic flowmeter is Faraday's Law of Electromagnetic Induction. This law states that if a conductor of length l (m) is moving with a velocity v (m/s), perpendicular to a magnetic field of flux density B (T), then the induced voltage e across the ends of conductor may be found by

$$e = Blv \quad (26)$$

The application of Faraday's Law to electromagnetic flowmeters is shown in Fig. 19. The magnetic field, the direction of the movement of the conductor, and the induced electromotive force (emf) are all perpendicular to each other. The induced voltages are linearly proportional to the mean velocity of liquids or to the volumetric flow rates.

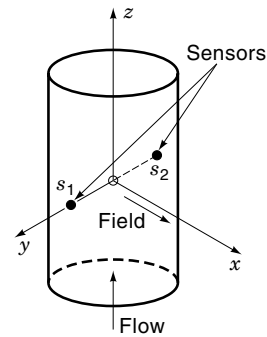


Figure 19. Operational principle of an electromagnetic flowmeter is based on Faraday's Law, which states that a voltage is induced in a conductor moving in a magnetic field. In electromagnetic flowmeters, the direction of movement of conductor, the magnetic field, and the induced emf are perpendicular to each other in x , y , and z axes. Electrodes S_1 and S_2 experience a virtual conductor resulting from liquid in the pipe.

In the case of electromagnetic flowmeters, the conductor is the liquid flowing through the pipe, and the length of conductor is the distance between the two electrodes, which is equal to the tube diameter. The velocity of the conductor is proportional to the mean flow velocity of the liquid. Hence, the induced voltage becomes

$$e = BDv \quad (27)$$

where D (m) is the diameter of pipe. If the magnetic field is constant and the diameter of the pipe is fixed, the magnitude of the induced voltage will only be proportional to the velocity of the liquid. If the ends of the conductor, in this case the sensors, are connected to an external circuit, the induced voltage causes a current i to flow. The current can be processed suitably as a measure of the flow rate. The resistance of the moving conductor may be represented by R to give the terminal voltage v_T of the moving conductor as $v_T = e - iR$. Using modern amplifiers with very high input impedance, the current i is minimized to such an extent that v_T is almost equal to e .

Electromagnetic flowmeters are often calibrated to determine the volumetric flow of the liquid. The volume of liquid flow Q (L/s) may be related to the average fluid velocity as

$$Q = Av \quad (28)$$

Writing the area A (m²) of the pipe as

$$A = \pi D^2/4 \quad (29)$$

gives the induced voltage as a function of the flow rate

$$e = \frac{(4BQ)}{(\pi D)} \quad (30)$$

This equation indicates that, in a carefully designed flowmeter, if all the other parameters are kept constant, then the induced voltage is linearly proportional only to the liquid flow.

Many different types of electromagnetic flowmeters are available, and all are based on Faraday's Law of Induction. The two most commonly used electromagnetic flowmeters are the ac and the dc types.

Alternating Current Flowmeters

In many commercial electromagnetic flowmeters, an alternating current of 50 Hz to 60 Hz in coils creates a magnetic field in the liquid flowing in the pipe. A voltage is induced in the liquid as described by Faraday's Law of Induction. A typical value of the induced emf in an ac flowmeter fixed on a 50 mm internal diameter pipe carrying 500 L/min is about 2.5 mV.

Historically, ac magnetic flowmeters were the most commonly used types because they reduced polarization effects at the electrodes. In general, they are less affected by the flow profiles of the liquid inside the pipes. They allow the use of high-power amplifiers with low drifts and high-pass filters to eliminate slow and spurious voltage drifts that emanate mainly from thermocouple and galvanic actions. These flowmeters find many applications as diverse as measuring blood flow in living specimens. Miniaturized sensors allow measurements on pipes

and vessels as small as 2 mm in diameter. In these applications, the excitation frequency is higher than industrial types, varying between 200 Hz and 1000 Hz.

A major disadvantage of the ac flowmeters is that the powerful ac field induced spurious ac signals in the measurement circuits. This necessitates periodical adjustment of zero output at zero velocity conditions. Also, in some harsh industrial applications, currents in the magnetic field may vary as a result of voltage fluctuations and frequency variations in the mains. The effect of fluctuations in the magnetic field may be minimized by the use of a reference voltage proportional to the strength of the magnetic field to compensate for these variations. To avoid the effects of noise and fluctuations, special cabling and calibration practices recommended by the manufacturers must be used to ensure accurate operations.

Readily available are ac flowmeters operating at 50 Hz, 60 Hz, or 400 Hz. In general, ac flowmeters can operate at 10 Hz to 5000 Hz. High frequencies are preferred in determining the instantaneous behavior of transients and pulsating flows. Nevertheless, in applications where extremely good conducting fluids and liquid metals are used, the frequency must be kept low to avoid skin effects. On the other hand, if the fluid is a poor conductor, the frequency must not be so high that dielectric relaxation is not instantaneous.

Direct Current Flowmeters

Direct current or pulsed magnetic flowmeters excite the flowing liquid with a field operating at 3 Hz to 6 Hz. As the current to the magnet is turned on, a dc voltage is induced at the electrodes. The signals observed at the electrodes represent the sum of the induced voltage and the noise, as illustrated in Fig. 20. When the current in the magnetic coils is turned off, the signal represents only the noise. Subtracting the measurement of the flowmeter when no current flows through the magnet from the measurement when current flows through the magnet effectively cancels out the effect of noise.

If the magnetic field coils are energized by normal direct current, then there occur several problems such as polarization, which is the formation of a layer of gas around the measured electrodes, as well as electrochemical and electromechanical effects. Some of these problems may be overcome by energizing

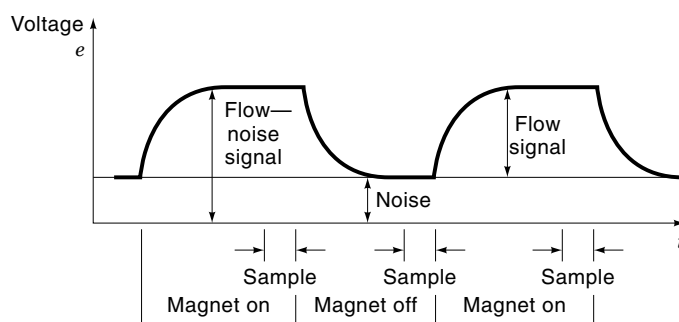


Figure 20. The signals observed at the electrodes represent the sum of the induced voltage and the noise. When the current in the magnetic coils is turned off, the signal across the electrodes represents only the noise. Subtracting the measurement of the flowmeter when no current flows through the magnet from the measurement when current flows through the magnet effectively cancels out the effect of noise.

the field coils at higher frequencies or ac. However, higher frequencies and ac generate transformer action in the signal leads and fluid path. Therefore, the coils are excited by dc pulses at low repetition rates to eliminate the transformer action. In some flowmeters, by appropriate sampling and digital signal-processing techniques, the zero errors and the noise can be rejected easily.

The zero compensation inherent in the dc magnetic flowmeters eliminates the necessity of zero adjustment. This allows the extraction of flow signals regardless of zero shifts due to superious noise or electrode coating. Unlike ac flowmeters, larger insulating electrode coating that may shift the effective conductivity significantly without affecting performance can be tolerated. It can be said that as long as effective conductivity remains high enough, a dc flowmeter will operate satisfactorily. Therefore, dc flowmeters are less susceptible to drifts, electrode coatings, and changes in the process conditions in comparison to conventional ac flowmeters.

Because of the slow pulsed nature of operations, dc magnetic flowmeters do not have good response times. However, as long as there are not rapid variations in the flow patterns, zero to full-scale response times of a few seconds do not create problems in the majority of applications. Power requirements are also much less because the magnet is energized part of the time. This gives an advantage in size reduction and in power saving of up to 75%.

Application of all types of magnetic flowmeters can be realized only on conductive liquids such as acids, bases, slurries, foods, dyes, polymers emulsions, and suitable mixtures that have conductivities greater than minimum conductivity requirements. Generally, magnetic flowmeters are not suitable for liquids containing organic materials and hydrocarbons. As a rule of thumb, the magnetic flowmeters can be applied if the process liquids that constitute a minimum of about 10% conductive liquid in the mixture.

For electromagnetic flowmeters to operate accurately, the process liquid must have minimum conductivity of $1 \mu\text{S}/\text{cm}$

to $5 \mu\text{S}/\text{cm}$. Most common applications involve liquids whose conductivity is greater than $5 \mu\text{S}/\text{cm}$. Nevertheless, for accurate operations, the requirement for the minimum conductivity of liquid can be affected by length of leads from sensors to transmitter electronics.

The wetted parts of a magnetic flowmeter include the liners, electrodes, and electrode holders. Many different materials such as rubber, Teflon, polyurethane, and polyethylene are used in the construction to suit process corrosivity, abrasiveness, and temperature constraints. The main body of a flowmeter and electrodes can be manufactured from stainless steel, tantalum, titanium, and various other alloys. Liners are selected mainly to withstand the abrasive and corrosive properties of the liquid. The electrodes must be selected such that they cannot be coated with insulating deposits of the process liquid during long period of operations. The main problem in electromagnetic flowmeter is due to the electrodes getting coated over a period of time, thereby leading to malfunctions. Present day flowmeters have means to overcome this by sensing high frequency pulses to the electrodes, thereby the coating gets removed automatically.

During the selection of electromagnetic flowmeters, the velocity constraints should be evaluated carefully in order to secure accurate performance over the expected range. The full-scale velocity of the flowmeter is typically 0.3 m/s to 10 m/s. Some flowmeters can measure lower velocities with somewhat deteriorated accuracy. Generally, employment of electromagnetic flowmeters over a velocity of 5 m/s should be considered carefully because erosion of the pipe and damages to liners can be significant.

A good electrical grounding of magnetic flowmeters, as illustrated in Fig. 21, is required to isolate relatively high common-mode potential. The sources of ground potential may be in the liquid or in the pipes. In practice, if the pipe is conductive and makes contact with the liquid, the flowmeter should be grounded to the pipe. If the pipe is made from nonconductive materials, the ground rings should be installed to maintain contact with the process liquid. Improper grounding results in excessive common-mode voltages that can severely limit the accuracy and damage the processing electronics.

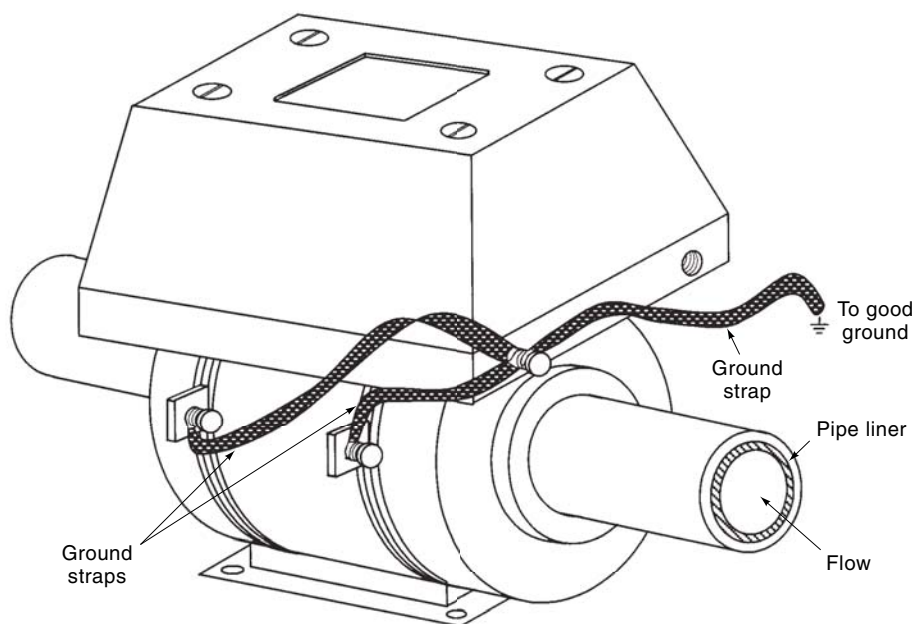


Figure 21. Good grounding of electromagnetic flowmeters is absolutely essential to isolate noise and high common-mode potential. If the pipe is conductive and makes contact with the liquid, the flowmeter should be grounded to the pipe. If the pipe is made from nonconductive materials, the ground rings should be installed to maintain contact with the process liquid. Improper grounding results in excessive common-mode voltages that can severely limit the accuracy and damage the processing electronics.

Table 1. List of Manufacturers

ABB K-Flow, Inc. P.O. Box 849 45 Reese Road Millville, NJ 08332 Tel: 1-800-294-8116 Fax: 609-825-1678	Flowmetrics, Inc. 7959 Alabama Avenue Conoga Park, CA 91304 Tel: 818-346-4492 Fax: 818-346-8991	Lambda Square, Inc. P.O. Box 1119M Bay Shore, NY 11706 Tel: 516-587-1000 Fax: 516-587-1011	Signet George Fischer, Inc. 2882 Dow Avenue Tustin, CA 92680 Tel: 800-280-5544 Fax: 714-731-6201
Brooks Instrument 407 West Vine Street Hatfield, Pennsylvania 19440 USA Tel: 215-362-3500 Fax: 215-362-374	Hastings P.O. Box 1436 Hampton, VA 23661 Tel: 800-950-2468 Fax: 804-723-3925	Marsh-McBirney, Inc. 4539 Metropolitan Court Frederick, MD 21704 Tel: 301-879-5599 Fax: 301-874-2172	Smith Meter, Inc. Smith Building 1602 Wagner Avenue Erie, PA 16510 Tel: 814-898-5000 Fax: 814-899-8927
Danfoss Inc. 4971 Mercantile Road Baltimore, Maryland 21236, USA Tel: 410-931-8250 Fax: 410-931-8256	Hoffer Flow Controls 107 Kitty Hawk Lane P.O. Box 2145 Elizabeth City, NC 27909 Tel: 800-628-6586 Fax: 252-331-2886	Micromotion Inc., 7070 Winchester Circle, Boulder, CO 80301 800-522-MASS, 522-6277 303-530-8100 USA	Sparling Instruments Company, Inc. 4097 Temple City Boulevard P.O. Box 5988 El Monte, CA 91734-1988 Tel: 800-423-4539
Daniel Industries, Inc. P.O. Box 19097 Houston, Texas 77224 USA Tel: 713-467-6000 Fax: 713-827-3880	Johnson Yokogawa Dept. P, Dart Road Newman, GA 30265 Tel: 800-394-9134 Fax: 770-251-6427	Nice Instrumentation, Inc. 1122 Campus Drive West Morganville, NJ 07751 Tel: 908-591-8899 Fax: 908-591-8909	Universal Flow Monitors, Inc. 1751 E. Nine Mile Road Hazel Park, MI 48030 Tel: 313-542-9635 Fax: 313-398-4274
Davis Instruments 4701 Mount Hope Drive Baltimore, MD 21215 Tel: 410-358-3900 Fax: 410-358-0252	Key Instruments 250 Andrews Road Trevosa, PA 19053 Tel: 215-357-0893 Fax: 215-357-9239	Rosemount Inc. Dept. MCA 15 12001 Technology Drive Eden Prairie, MN 55344 Tel: 612-828-3006 Fax: 612-828-3088	Xolox Corporation 6932 Gettysburg Pike Ft. Wayne, IN 46804 Tel: 800-348-0744 Fax: 219-432-0828
FCI Fluid Components International 1755 La Costa Meadows Drive San Marcos, CA 92069-5187 Tel: 619-744-6950 Fax: 619-736-6250	King Instrument Company 16792 Burke Lane Huntington Beach, CA 92647-4559 Tel: 714-841-3663 Fax: 714-847-2062	Schlumberger Fluid Power Div. 8635 Washington Avenue Racine, Washington 53406, USA Tel: 414-884-7400 Fax: 414-884-7440	
Fischer Porter 50 Northwestern Drive P.O. Box 1167T Salem, NH 03079-1137 Tel: 603-893-9181 Fax: 603-893-7690	Krohne America, Inc. 7 Dearborn Rd. Peabody, MA 01960 Tel: 978-535-6060 Fax: 978-535-1760		

Table 2. Accuracy and Applications of Flowmeters

Flowmeter	Liquid			Accuracy (%)	Reynold's Number	Example Standards
	Clean	Slurry	Gas			
Coriolis	Yes	Yes	Yes	±0.2–1.0	No limit	
Magnetic	Yes	Yes	No	±0.5–1.0	No limit	ISO 6817, ASME 16Mt
Orifice plate	Yes	Yes	Yes	±0.5–3.0	>1,000	ANSI/ASME MFC 3M, ISO 5167
Oscillating vane	Yes	Yes	Yes	±0.5	Depends on size	
Pitot tube	Yes	No	Yes	±5.0	No limit	ASME/MFC SC16
Positive displacement	Yes	No	Yes	±0.5–1.0	<8,000	OIML R31
Target	Yes	Yes	Yes	±0.5–2.0	>1,000	
Thermal mass	Yes	Yes	Yes	±1.0	No limit	ISO TC30
Turbine	Yes	No	Yes	±0.5–1.0	Viscous	ASME/ANSI 4M
Ultrasonic	Yes	Yes	Yes	±1.0–5.0	No limit	ANSI/ASME MFC 5M
Variable area	Yes	No	Yes	±0.5–1.0	Viscous	
Venturi	Yes	No	Yes	±0.5–2.0	<75,000	ASME/ANSI ISO 5167
Vortex	Yes	No	Yes	±0.5–1.5	<10,000	ASME/ANSI MFC 6M

The accuracy of conventional magnetic flowmeters is usually expressed as a function of full scale, typically 0.5% to 1% FS. However, dc flowmeters have a well-defined zero because of the automatic zeroing nature; therefore, they have percentage rate of accuracy better than ac types, typically at 0.1% to 2% rate.

CALIBRATION, ACCURACY, PRECISION, AND STANDARDS

There are many manufacturers of flowmeters, as listed in Table 1. They are used in diverse industries such as oil and gas, iron and steel, power, food and drink, water distribution, pharmaceutical, agricultural, mineral processing, and manufacturing.

In the past, the flowmeter repeatability rather than accuracy was the main concern. However, nowadays, with the use of microprocessors together with improved transducers, better accuracies are possible. For example, it is possible to obtain mass flowmeters having accuracies of 0.2% to 0.5% after having compensated for temperature and pressure effects. A list of flowmeters together with their accuracy and suitability for various applications are given in Table 2.

It is necessary to calibrate flowmeters because of the possible changes in fluid properties. In field applications where characteristics of fluids change often or mixtures of fluids are used, frequent rechecking and recalibration of flowmeters may be necessary for continual and accurate operations. There are a number of national and international codes of practice for calibrating flowmeters. A number of organizations such as ASME, ISO, and British Standards codify standards about flowmeters. The standards are codified for accuracy of components, calculating precision, bias errors, in situ testing, uncertainty analysis, and the like. The American Petroleum Institute (API) standards pertain to measurement of liquid petroleum products, and American Gas Association (AGA) standards pertain to measurement of natural gas. Also, a number of committees are looking into flow measurement standards such as ISO/TC28/SC2/WG6 of Israel and the United States and BSI/PCL/2/9 of the United Kingdom. Calibrations of flowmeters are done in a standards laboratory. Flow being a derived quantity, its calibration should be traceable to the fundamental units of Length, Mass, and Time (L, M, and T). The gravimetric methods of calibration as per ISO 4185 is the method for calibration of flowmeters used for liquid application. The Bell power is used for primary calibration of gas flowmeters.

Reading List

- J. P. Bentley, *Principles of Measurement Systems*, 2nd ed., Burnt Mills, UK: Longman Scientific and Technical, 1988.
- E. O. Doebelin, *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990.
- D. Ginesi, Choices abound in flow measurements, *Chem. Eng.*, New York, **98** (4): 88–100, 1991.
- J. P. Holman, *Experimental Methods for Engineers*, 5th ed., New York: McGraw-Hill, 1989.
- R. S. Medlock and R. A. Furness, Mass flow measurements—A state of art review, *Measurement Control*, **23** (4): 100–113, 1990.
- R. W. Miller, *Flow Measurement Engineering Handbook*, 3rd ed., New York: McGraw-Hill, 1996.
- J. A. Shercliff, *The Theory of Electromagnetic Flow—Measurements*, Cambridge, UK: Cambridge University Press, 1987.
- D. W. Spitzer, *Industrial Flow Measurements*, 2nd ed., Research Triangle Park, NC: Instrument Society of America, 1990.

HALIT EREN
Curtin University of Technology

DYNAMOMETERS

The dictionary defines a dynamometer as (1) a device for measuring mechanical force, as a balance, and (2) a device for measuring mechanical power, especially one that measures the output or drive torque of a rotating machine. The prefix “dynamo-” itself is a variation of “dyna-,” which is a learned borrowing from Greek meaning “power,” and is used in the formulation of compound words, such as dynamometer. According to the US Patent Office, however, dynamometers include a very wide range of measuring instruments. Some manufacturers refer to only “complex” devices, which incorporate load sensors or transducers in their structure as dynamometers.

The classical usage of the word dynamometer is primarily for power measurement. In such a dynamometer, power is determined as the product of the measured torque and angular velocity. Currently, and more generally, the term dynamometer and load transducer are used synonymously, and they cover load measurement instruments ranging from the simple to the more complex. For example, a simple device measuring grip strength of the human hand is known as a grip dynamometer. Also, terms such as load sensor, load cell, force transducer, and force platform (plate) are being used to cover the same range of instruments. For completeness, a brief description of specialized dynamometers as power measurement instruments is given first, followed by a more detailed discussion of dynamometers as load transducers in general.

POWER MEASUREMENT

Two types of mechanical power measuring devices exist: (1) those absorbing power and dissipating it as heat and (2) those transmitting the measured power. These are called, respectively, absorption dynamometers and transmission dynamometers. Common absorption dynamometers are the Prony brake, the electric generator, the water brake, and the magnetic or eddy-current brake. Figure 1 schematically depicts a Prony brake. A Prony brake is typically used for wheels smaller than 0.6 m (24 in.) in diameter, where the power to be absorbed is relatively small. As the brake is applied, the force, f , perpendicular to the cantilever beam is measured. Simultaneously, the angular velocity of the wheel is also monitored. The product of the angular velocity, the measured torque (i.e., force, f , times moment arm, L), and the appropriate unit conversion constants gives the instantaneous power. The other absorption dynamometers work in a similar way, where a load is applied to impede the rotating shaft while both the load applied and the angular velocity of the shaft are measured.

Many types of transmission dynamometers have been used in the past, but currently only the torsion dynamometer and the cradle-mounted electric motor (i.e., chassis dynamometer)

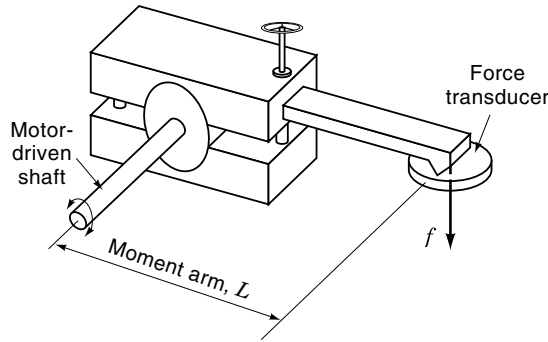


Figure 1. Schematic of a Prony brake.

are in use. Torsion dynamometers directly measure the torque transmitted through a shaft using a torque transducer placed in series with that shaft. When the power output of an electric motor is in question, this can also be determined by measuring the reaction forces transmitted to the floor at the motor supports. Naturally, in both systems the angular velocity of the shaft also needs to be measured.

FORCE AND TORQUE MEASUREMENT

The force applied to a body is measured by measuring its physical effect on that body. This can take the form of a balance, where the unknown force is balanced against the known gravitational force of a standard mass, or against a magnetic force developed by interaction of a current-carrying coil and a magnet. Alternatively, the physical effect measured can be the elastic deformation of a body. Relating force to an acceleration via a known mass, or relating it to fluid pressure, is a variation of the latter method and will not be discussed separately. A simple example of an instrument where force is measured by relating it to a deformation is a spring scale, where the spring deforms proportional to the applied force. All of the above methods can be successfully used if the force to be measured is static or slowly varying. However, if the force is dynamic, the method of measuring deformations is used. These measured deformations are extremely small, and such a measuring device is generally referred to as a force transducer or a dynamometer.

A simple mathematical model is used to establish the input-output relationship of a force transducer. Figure 2 shows a physical model of a force transducer, where the system is idealized as a known mass, m , carried by a spring of known stiffness, k , and a damper with a known constant, b . This relationship, along with the resulting transfer function in the Laplace domain, is seen as a simple second-order relationship as

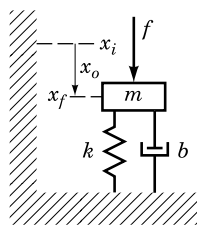


Figure 2. Idealized model of an elastic force transducer.

follows:

$$m\ddot{x}_o = -kx_o - b\dot{x}_o + f$$

$$\frac{x_o}{f}(s) = \frac{K}{s^2 + \frac{2\zeta s}{\omega_n} + 1} \quad (1)$$

The dynamic response characteristic parameters of natural frequency, damping ratio, and sensitivity are found to be the following:

$$\omega_n = \sqrt{\frac{k}{m}} \quad (\text{natural frequency})$$

$$\zeta = \frac{b}{2\sqrt{km}} \quad (\text{damping ratio}) \quad (2)$$

$$K = \frac{1}{k} \quad (\text{sensitivity})$$

As can be seen in Eq. (2), the stiffness, k , is used to determine each of the above parameters, because it is inversely proportional to the damping ratio and sensitivity and is directly proportional to the natural frequency. Therefore, the stiffness is very important in transducer behavior and design.

So far, only a force acting in a known direction has been discussed. More generally, all six components of a load acting on a body can be considered, where the force vector and the moment vector are each expressed with their three orthogonal components. A transducer which measures more than one component of the load is called a multicomponent load transducer or is referred to by the number of its components (i.e., a six-component load transducer).

Torque is a special case of a moment acting about the long axis of an elongated member. It causes shear deformation of this member, and thus it requires a somewhat different measurement method. Because torque measurement is usually carried out as an independent measurement on a shaft, torque transducers are commonly looked upon as a separate class of single-component load transducers.

Different sensing technologies are used to measure the elastic deformations of a transducer element. These include strain gauges, piezoelectric crystals, piezoresistive elements, and capacitive transducers. Of these, the former two are the most commonly used methods in load measurement technology, and they will be discussed here.

Strain-Gauged Transducers

Strain-gauged transducers are used to measure both static and dynamic loads. These devices use either metal-foil gauges or semiconductor gauges. Metal-foil strain gauges are produced by etching the gauge pattern into a thin sheet of metal foil deposited onto a plastic backing. Typical sizes vary from a few millimeters to several centimeters. Semiconductor strain gauges are generally smaller in size and have a higher sensitivity. However, they are fragile and relatively brittle, and they are more temperature-sensitive than metal-foil gauges. Semiconductor gauges have mainly been used in miniature transducers, the availability of which are increasing as technology advances.

Strain gauges are applied to the body of an elastic member of a transducer using adhesive so that when the elastic member deforms, the length of the strain gauge, and therefore its

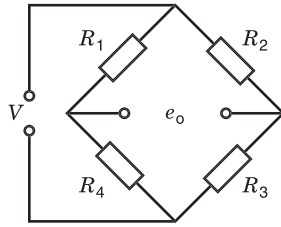


Figure 3. Wheatstone bridge.

electrical resistance, changes proportionally. When a constant current is applied across the gauge, the change in length is seen as a change in voltage proportional to the strain induced. Strain gauges are made to be sensitive to changes in length along a principal direction, but insensitive in the transverse direction. Whether a transducer is designed to measure a single load component or multiple load components, strain gauges used for each load channel are wired in a Wheatstone bridge configuration. This not only allows measurement isolation of the desired load component, but also compensates for deformations resulting from temperature changes. A Wheatstone bridge is depicted in Fig. 3, where each resistance represents one of the strain gauges used in the transducer. Also shown are the input and output voltages. When the bridge is balanced (i.e., $R_1 = R_2$ and $R_3 = R_4$) the output voltage is zero. In a typical transducer the nominal gauge resistances are equal, so $R_1 = R_2 = R_3 = R_4 = R$, where standard values for R are 120 Ω , 350 Ω and 1000 Ω ; the most commonly used value is 350 Ω . The bridge output voltage, as a function of individual gauge strains, is given as

$$e_o = \frac{V \cdot G_f}{4} [(\epsilon_1 - \epsilon_2) + (\epsilon_3 - \epsilon_4)] \quad (3)$$

where, the ϵ 's are the individual strains, V is the excitation or supply voltage, and G_f is a material property of the gauge called the gauge factor. The gauge factor is the relationship of the strain to resistance change (i.e., the sensitivity of the gauge) and is expressed as

$$G_f = \frac{\frac{\Delta R}{R}}{\epsilon} \quad (4)$$

Typical foil gauges would have a gauge factor of approximately 2, while the gauge factor for semiconductor strain gauges can be as high as 100.

When selecting a material to be used for a transducer elastic element, we need to consider mechanical properties, thermal properties, and machinability of the material. Linear elastic load response with minimal hysteresis is one of the most desirable mechanical properties for a transducer material. As for thermal properties, the thermal conductivity contributes to the transducer accuracy and stability. The most commonly used materials include SAE 4340 steel, 631 (17-4 PH) stainless steel, and 2024-T4 and 7075-T6 aluminum. A comprehensive list can be found in Ref. 1.

Strain-gauged transducers involve a considerable design trade-off. From a natural frequency point of view, which influences the capability of measuring dynamic forces, the transducer element needs to have as high a degree of stiffness as

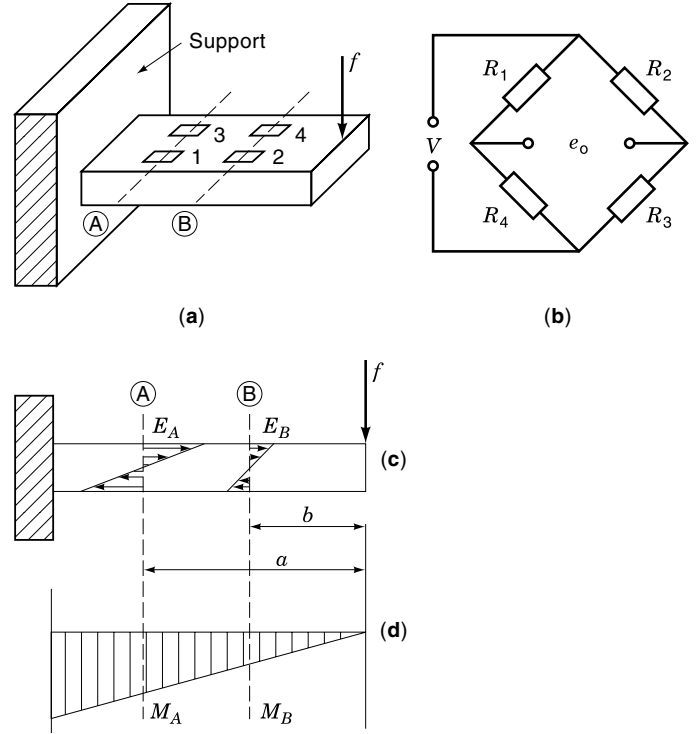


Figure 4. Single-component force transducer. (a) Strain gauge location and orientation, (b) Wheatstone bridge, (c) stress (strain) distribution in beam, and (d) bending moment diagram.

possible. On the other hand, higher sensitivity requires greater deformations and thus a lower degree of stiffness. Bridging the gap in this design trade-off are high-end electronics. Today, electronics providing a high signal-to-noise ratio (SNR) allow for sensitive transducers which have relatively stiff elastic elements.

Considering single-component force transducers, a relatively low capacity transducer would typically utilize a bending beam member as the spring element. Figure 4(a) depicts such a member with gauges located at A and B; this member is loaded by the force to be measured, f . In Fig. 4(b) the Wheatstone bridge is shown, where the numbers correspond to the gauge designations in Fig. 4(a). Note that the bridge output is proportional to the difference between strains measured at levels A and B. Figure 4(c) shows the stress (or strain) distribution in the beam at the two levels, and Fig. 4(d) shows the bending moment diagram for the beam. Strain at A is proportional to the moment at the same location (i.e., $f \cdot a$), and similarly the strain at B is proportional to $f \cdot b$. The difference between the strains at the two levels, therefore, is proportional to $f \cdot (a - b)$. Thus, the output of the bridge is independent of the location of the force. This makes it possible to design a transducer as depicted in Fig. 5, where the force can act anywhere on the surface provided. To increase sensitivity without decreasing stiffness, very often the bending beam is weakened at the locations where the strain gauges are attached. Two examples of such a design are shown in Fig. 6.

Using a shear beam, as shown in Fig. 7(a), would be an appropriate solution when the force to be measured is relatively high and the above design is not satisfactory. The gauges are positioned at 45° angles to the long axis of the beam (on

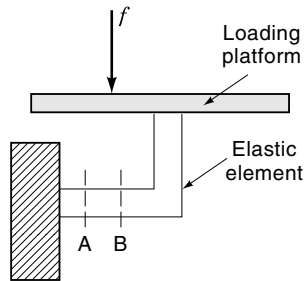


Figure 5. Adapted single-component force transducer.

both sides). Figure 7(b) shows the shear stress distribution on the gauged surface, and Fig. 7(c) is the same shear stress acting on a small area where the gauges are located. If the element of Fig. 7(c) is rotated by 45° , the compressive and tensile stresses of Fig. 7(d) are obtained. Therefore, gauge 1 would be subjected to compression, while gauge 2 would be subjected to tension, and the difference between the two gauge readings would still be proportional to the applied force. When we combine all four gauges (both sides), a full Wheatstone bridge with an output proportional to the force is obtained. It is again possible to modify the cross-sectional geometry by reducing the beam thickness at the gauge location, thus increasing the sensitivity without significantly affecting the stiffness of the beam. References 1 and 2 provide design variations based on the above principles.

The beam of either Fig. 4 or Fig. 7 can be instrumented on the currently blank adjacent sides to measure the force perpendicular to the one depicted. Furthermore, another set of gauges, as shown in Fig. 8 (gauges on the opposite side of the beam are not shown), can be used to measure the force along the long axis of the beam. It should be noted, however, that the configuration of Fig. 8 usually provides the highest stiffness and the least sensitivity. One way of increasing sensitivity in this direction would naturally be reducing the cross-sectional area. This could be achieved without significantly lowering the stiffness in the transverse direction by using a hollow geometry such as

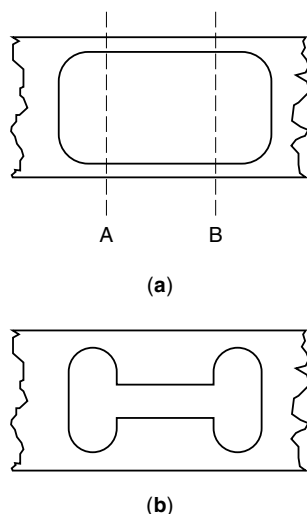


Figure 6. Example cross-sections to increase sensitivity without decreasing stiffness.

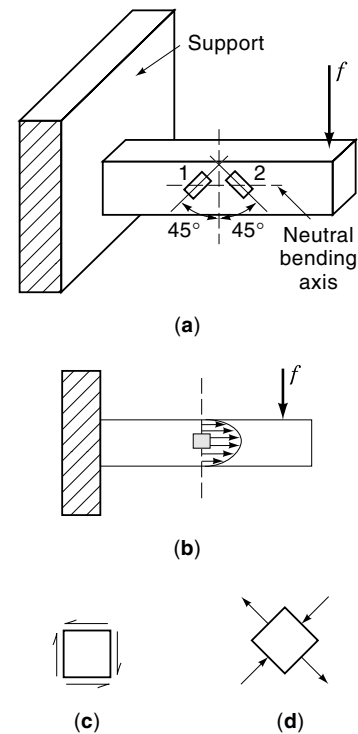


Figure 7. Shear beam single-component force transducer. (a) Strain gauge location and orientation, (b) shear stress (strain) distribution in beam, (c) unit area shear stress, and (d) rotation of unit area to obtain normal stresses.

a tube. By selecting a suitable ratio between the wall thickness and the diameter of the tube, it is usually possible to find a satisfactory solution to measure all three force components with the desired resolution.

A tubular cross section is also very suitable for torque measurement. Referring to Fig. 9, note that the gauges are oriented to measure the shear due to the torque only, and any shear due to a transverse force (similar to the one in Fig. 7) is eliminated. Actually, a detailed analysis of the stresses caused by all possible loads applied to the end of the tube will show that the gauge configuration of Fig. 9 is only sensitive to the torque.

Piezoelectric Load Transducers

Certain materials respond to an applied electrical charge by mechanically deforming. Alternatively, an electrical charge is

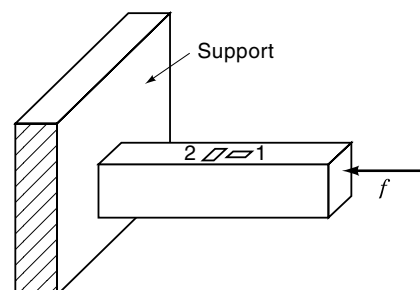


Figure 8. Gauge location and orientation for normal stress measurement.

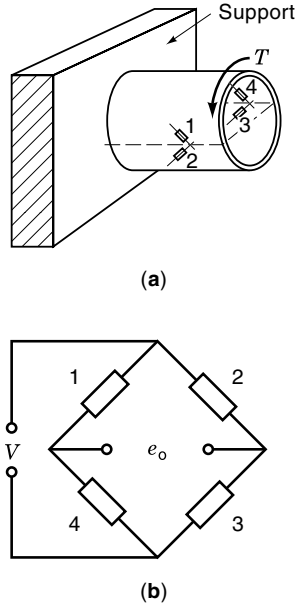


Figure 9. Torque transducer. (a) Gauge location and orientation. (b) Wheatstone bridge.

developed if the material is mechanically stressed. These actions are referred to as piezoelectric effects, and materials which exhibit this behavior are called piezoelectric materials. Quartz is a piezoelectric material with high stability and is normally used in building piezoelectric load transducers.

The piezoelectric properties of quartz crystal are directional, and axial and shear force components can be isolated by orienting the material in the transducer. Figure 10 shows a quartz crystal with disks cut out to be sensitive to compression (disk a) and shear (disk b). When a disk is placed between metal electrodes and subjected to a force, a charge, q , is produced. This charge is proportional to the deformation of the crystal, and thus it is also proportional to the force applied. Simultaneously, a voltage is developed between the electrodes. Therefore, the voltage output can be amplified and calibrated to measure the force applied.

Normally, a pair of disks is used to measure each load component. Using the disks in pairs doubles the sensitivity and permits simple electrical contact by a central electrode. By stacking three pairs of disks, one pair for the axial load and one pair for each of the two orthogonal shears, a three-component force transducer can be obtained. This type of transducer must be assembled with a compressive preload, which prevents the disks from separating from the metal plate electrodes during tensile loading.

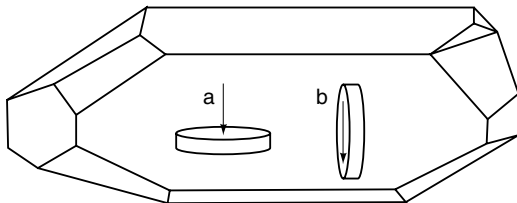


Figure 10. Quartz crystal. (a) Normal force orientation and (b) shear force orientation.

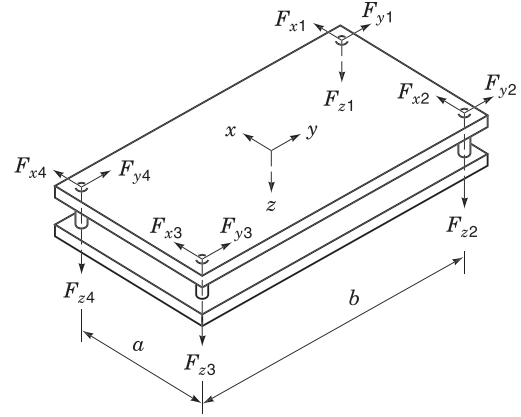


Figure 11. Six-component piezoelectric dynamometer made out of four three-component load transducers.

Six-component transducers can be built by placing four such three-component load transducers at four corners of a platform. Figure 11 (commonly called a piezoelectric dynamometer) shows such a configuration, where each load cell is numbered from 1 to 4, and the load direction is indicated by the direction of each coordinate axis. Thus, the two subscripts of each load component uniquely identify the force. The 12 outputs of the transducer are combined inside the transducer such that eight signals are actually output. These eight channels represent the four individual vertical force components (F_{z1}, F_{z2}, F_{z3} , and F_{z4}); two shears in the x -direction ($F_{x1} + F_{x2}$ and $F_{x3} + F_{x4}$), and two shears in the y -direction ($F_{y1} + F_{y4}$ and $F_{y2} + F_{y3}$). In order to get the three force and the three moment components, the output needs to be further reduced as follows:

$$\begin{aligned}
 F_x &= (F_{x1} + F_{x2}) + (F_{x3} + F_{x4}) \\
 F_y &= (F_{y1} + F_{y4}) + (F_{y2} + F_{y3}) \\
 F_z &= (F_{z1} + F_{z2}) + (F_{z3} + F_{z4}) \\
 M_x &= (F_{z1} + F_{z2}) - (F_{z3} - F_{z4})(b/2) \\
 M_y &= (-F_{z1} + F_{z2}) + (F_{z3} - F_{z4})(a/2) \\
 M_z &= [(F_{x1} + F_{x2}) - (F_{x3} + F_{x4})(b/2) \\
 &\quad + [(F_{y2} + F_{y3}) - (F_{y1} + F_{y4})(a/2)
 \end{aligned} \tag{5}$$

Piezoelectric transducers have the advantage of having high sensitivity and high rigidity. Typical stiffness properties of these crystals result in natural frequencies as high as 30,000 Hz. However, the governing natural frequency is normally that of the members supporting the transducer, which is usually much lower than that of the quartz alone. A disadvantage of piezoelectric transducers is that they are not inherently suited for static measurements. This is due to the charge developed in the quartz under load gradually leaking, causing signal decay eventually to zero. Then, if the applied load were released, the transducer will erroneously indicate a load in the opposite direction equal to the amount of the decay.

As stated previously, the applied force, f , to a piezoelectric transducer causes a deflection, x . A potential voltage, E , is produced from the resulting charge buildup. This can be

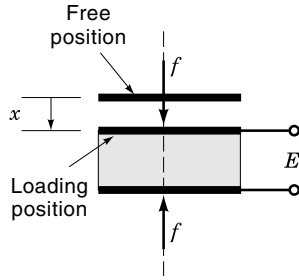


Figure 12. Piezoelectric transducer.

expressed as

$$E = \frac{q}{C_{cr}} \quad (6)$$

where C_{cr} is the crystal capacitance of the piezoelectric material and is defined as

$$C_{cr} = \frac{\phi A}{t} \quad (7)$$

In Eq. (7), ϕ is the dielectric constant of the crystal, A is the area of the plates, and t is the distance between the plates. Figure 12 shows such a piezoelectric transducer. A piezoelectric constant, k , relates the applied force, f , to the resulting charge, q as shown:

$$q = kf \quad (8)$$

Using Eq. (8) in Eqs. (6) and (7) results in the following relation for the potential voltage, E :

$$E = \frac{kft}{\phi A} \quad (9)$$

In order to quantify the applied force the induced, or potential, voltage, E , must be measured. A charge amplifier is typically implemented due to the induced voltage's small magnitude. This leads to another complication in that the charge may leak back through the crystal. Figure 13 depicts the circuits required to determine the behavior of a piezoelectric transducer. Figure 13(a) shows the crystal itself, while Fig. 13(b) is a schematic of the cabling and amplifier. In Fig. 13 the charge

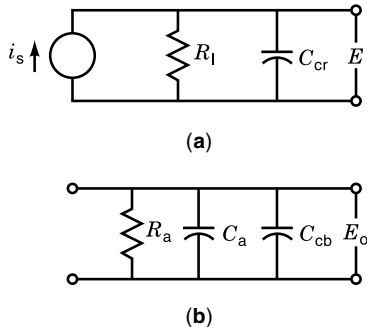


Figure 13. Equivalent circuits for (a) piezoelectric transducer and (b) amplifier and cables.

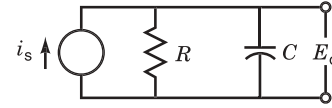


Figure 14. Equivalent circuit for piezoelectric transducer, amplifier, and cables.

generator is represented as a current generator, and the current source i_s is the time derivative of the charge generated, C_{cr} is again the crystal capacitance, C_{cb} is the capacitance of the connecting cables, C_a is the capacitance of the charge amplifier, R_a is the internal amplifier resistance, R_l is the sensor leakage resistance, and E_o the resulting output voltage to be measured.

Figure 14 is the equivalent circuit of the transducer, cabling and amplifier. From this circuit:

$$i_s = \frac{dq}{dt} = k \frac{df}{dt} \quad (10)$$

$$R = \frac{R_a R_l}{R_a + R_l} \quad (11)$$

In a piezoelectric force transducer, R_l is usually on the order of $10^{11} \Omega$; however, there are commercially available systems with higher R_a and R_l values resulting in up to $10^{14} \Omega$ total resistance, R . Systems with this high a total resistance have a very slow leakage rate and therefore enable quasi-static force measurement.

Analyzing the equivalent circuit of Fig. 14 results in the following current loop equation:

$$i_s = i_c + i_r \quad (12)$$

The output voltage is then expressed as

$$E_o = E_c = \int \frac{i_c}{C} dt = \frac{\int (i_s - i_r) dt}{C} \quad (13)$$

$$CE_o = \int (i_s - i_r) dt$$

After differentiation and substitution we obtain

$$C \frac{dE_o}{dt} = i_s - i_r = k \frac{df}{dt} - \frac{E_o}{R} \quad (14)$$

This differential equation can be put in standard form and then expressed in the Laplace domain as

$$(RCs + 1)E_o = (Rks) f \quad (15)$$

or as a first-order system input-output relationship, or transfer function:

$$\frac{E_o}{f}(s) = \frac{K\tau_s}{\tau_s + 1} \quad (16)$$

where

$$\tau = RC \text{ (time constant)}$$

$$K = k/C \text{ (static sensitivity)}$$

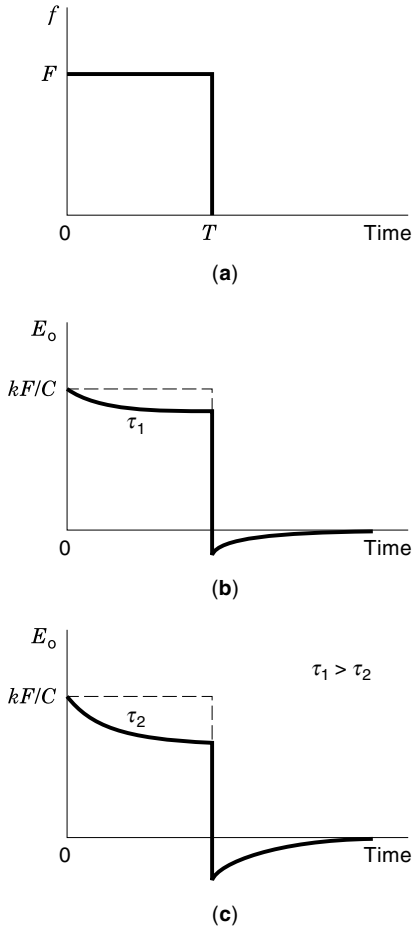


Figure 15. Response of a piezoelectric transducer to constant force input. (a) Input, (b) response for relatively high time constant, and (c) response for a relatively low time constant.

Figure 15(a) shows a constant force input, f , for a time of T . With these initial conditions for the time periods from 0 to T and from T to infinity, Eq. (16) can be solved:

$$E_o = \frac{kf}{C} e^{-t/\tau} \quad \text{and} \quad E_o = \frac{kf}{C} (e^{-T/\tau} - 1) e^{-(t-T)/\tau} \quad (17)$$

The response of Eq. (17) is shown in Fig. 15(b) and 15(c). These figures show undershoot of the response and also show that as the time constant decreases, the signal response decay increases. This is the reason for the need of a high time constant in piezoelectric force transducers.

The amplifier properties, along with the dimensions and physical properties of the piezoelectric material, determine the actual time constant of the transducer system. However, there are a limited number of ways to increase the time constant. By the relation above, an increase in either R or C will increase the time constant. A shunt capacitor or a series resistor at the amplifier can achieve this; however, the sensitivity of the transducer would decrease if a shunt capacitor were used. Therefore, much like strain-gauged transducers, there is a design trade-off with piezoelectric transducers as well. This trade-off is between the high time constant required for static measurement and the desired quasi-static sensitivity.

Cross-talk

Generally, in the measuring range of a transducer, relationships between the loads and signals are linear, and calibration parameters are given by constant values. Ideally, each output channel of a transducer would only be sensitive to the load component that it is designed to measure. However, manufacturing tolerances as well as some design considerations cause a transverse sensitivity or a cross-talk between the channels. For example, in a six-component load transducer, the force in the x -direction might be sensitive to some or all of the other five outputs of the transducer. Thus, F_x would be expressed as a linear combination of these signals:

$$F_x = c_{11} \cdot s_1 + c_{12} \cdot s_2 + c_{13} \cdot s_3 + c_{14} \cdot s_4 + c_{15} \cdot s_5 + c_{16} \cdot s_6 \quad (18a)$$

or

$$F_x = c_{ij} \cdot s_j \quad (18b)$$

where c_{1j} ($j = 1$ to 6) are the calibration constants, and s_j ($j = 1$ to 6) are the six output signals. The relationship between the six signals and the six load components can be expressed as

$$\{F_x F_y F_z M_x M_y M_z\}^T = [c_{ij}] \{s_1 s_2 s_3 s_4 s_5 s_6\}^T \quad (19)$$

where each load and signal vector is a column vector, and the calibration coefficients, c_{ij} , are given in a 6×6 matrix. Typically, the nondiagonal terms of the calibration matrix represent the cross-talk between the channels.

In most commercially available multicomponent transducers the cross-talk terms are less than 1% or 2% of the main calibration constants. This was common practice when data were primarily recorded by chart recorders and correcting for cross-talk usually involved a tedious process. Current means of data acquisition and processing has removed this requirement, thus providing designers with more possibilities of creative load transducer configurations.

Users may incorrectly assume that the load components which are not measured by a transducer do not affect the measured components. However, this is usually not guaranteed. For example, a single axis tension or compression transducer might give erroneous results in the presence of shear forces.

Calibration

Transducer calibration normally involves applying a known force, and relating the output of the transducer to the magnitude of this force. Applying methods of good experimental procedure, the applied load is incrementally increased up to the working limit of the transducer, then again incrementally decreased to zero. The linearity as well as any hysteresis present in the transducer should then be verified. If the transducer is to measure a dynamic load, the frequency response, or the first natural frequency of the transducer also needs to be determined. It is desirable to have the natural frequency of the transducer at least four or five times higher than the maximum frequency content of the applied load.

In calibrating a multicomponent transducer, at least as many independent load sets as the number of channels involved

must be applied. If higher number of independent load sets are applied, then the calibration matrix can be obtained using a least-squares fit. If $\{F_k\}$ is the k th independent load set and $\{s_k\}$ is the corresponding output signal set, then

$$[F_1 F_2 \dots F_k] = [c_{ij}][s_1 s_2 \dots s_k] \quad (20)$$

where c_{ij} are the elements of the unknown calibration matrix. For example, for a six-component transducer where the number of output channels is also six, the calibration matrix would be a 6×6 matrix. In the case of a transducer having eight outputs representing the six loads (i.e., the three forces and three moments), like most piezoelectric six-component transducers, a 6×8 calibration matrix would be used.

Representing $[F_1 F_2 \dots F_k]$ by $[F]$, and similarly representing the signals by $[s]$, Eq. (20) can be written as

$$[F] = [c_{ij}][s] \quad (21)$$

Post-multiplying both sides by the transpose of $[s]$, $[s]^T$, and then post-multiplying by the inverse of $\{[s][s]^T\}$, the calibration matrix is obtained:

$$[c_{ij}] = [F][s]^T \{[s][s]^T\}^{-1} \quad (22)$$

In order to determine whether there is any advantage to using more calibration points than the minimum required, the authors used up to 50 independent load sets in calibrating six-component load transducers used in gait analysis (i.e., force plates). For these particular transducers, any number of load sets larger than 15 have not been found to improve the accuracy of the calibration matrix. Although this number of load sets does not necessarily apply to all six-component transducers, it provides a guideline for establishing a successful calibration procedure.

BIBLIOGRAPHY

1. The Technical Staff of Measurements Group, Inc., *Strain Gage Based Transducers: Their Design and Construction*, Raleigh, NC: Measurements Group, 1988.
2. C. P. Wright, *Applied Measurement Engineering: How to Design Effective Mechanical Measurement Systems*, Upper Saddle River, NJ: Prentice-Hall, 1995.

Reading List

- A. S. Z. Barnes and N. Berme, Measurement of kinetic parameters technology, in R. L. Craik and C. A. Oatis (eds.), *Gait Analysis: Theory and Application*, St. Louis: Mosby-Year Book, 1995, pp. 239–251.
- N. Berme, Load transducers, in N. Berme and A. Cappozzo (eds.), *Biomechanics of Human Movement Applications in Rehabilitation, Sports and Ergonomics*, Worthington, OH: Bertec Corp., 1990, pp. 140–149.
- E. O. Doebelin, Force, torque, and shaft power measurement, in *Measurement Systems: Application and Design*, 4th ed., New York: McGraw-Hill, 1990, pp. 390–433.

NECIP BERME
The Ohio State University
Douglas R. Morr
Bertec Corporation

TORQUEMETERS

Torque, or moment about an axis, is measured in several industrial situations and, increasingly, in consumer products. A common experience of torque measurement is the use of a torque spanner for tightening a wheel-nut on an automobile. The torque being applied is indicated on a scale, and the device usually operates through the bending of a rod. The meaning of torque as the product of a force and lever length is clear in this situation, and it is measured in Nm (newton meters). The axis about which the torque is applied is that of the stud on which the nut is tightened. Usually, the term *torquemeter* is taken to mean a device for the measurement of the torque in a rotating shaft, but this is not necessarily so, and other situations need to be instrumented. Measurement may be required in testing small motors or in aircraft or tanks, and in many other situations. Thus the range of requirements is very large in terms of torque range, sensitivity, accuracy, physical size, and, increasingly, of cost. Because of this, several different techniques have been developed successfully for different applications. Continuing work is mainly in the search for low-cost high performance devices for use in industrial control systems and consumer products. The overall picture is complicated because it is not always appropriate to purchase a torquemeter but rather to design it as an integral part of the experimental rig or commercial product, and there are well-established principles for doing this. In some cases, the torquemeter will also be required to act as a mechanical load, absorbing energy. In other cases, it will be important that the device does not change the operation of the system significantly through loading or through its own mechanical flexibility. There are torquemeters in use which were installed a long time ago and which continue to give good service because of their ruggedness and reliability. The user needs to be aware of these, although generally, many of these would not be installed today on grounds of cost and lack of operational flexibility. This is particularly true for the measurement of shaft torque, and this is the type of measurement which is considered first.

THE MEASUREMENT OF SHAFT TORQUE

A typical application of a shaft torquemeter incorporating a mechanical load is for the measurement of the torque/speed characteristics of a motor or engine in performance testing. In older, mechanical instruments, the integration of the two functions, the measurement and the loading, often forms an essential part of the design. The simplest and most direct device is the Prony Brake (Fig. 1), in which the rotating shaft carries a narrow drum. Braking is applied to this by means of a brake band, or possibly by friction blocks, and a screw mechanism, A, tightens this to increase the braking effect. The torque applied by the brake mechanism is measured in a number of ways, the traditional way being to use a spring balance as shown in the Fig. 1. The torque acting on the drum is $(F_1 - F_2)r$ Nm and is transferred to the lever which has a fulcrum midway between the points where the belt is attached. The belt is tightened by the screw mechanism indicated in order to increase the braking torque. The torque is then balanced by the force F which is shown on the balance and which acts on a lever of length R . Thus, the torque is FR Nm. Normally, this would be plotted against the shaft speed as the brake is tightened to give

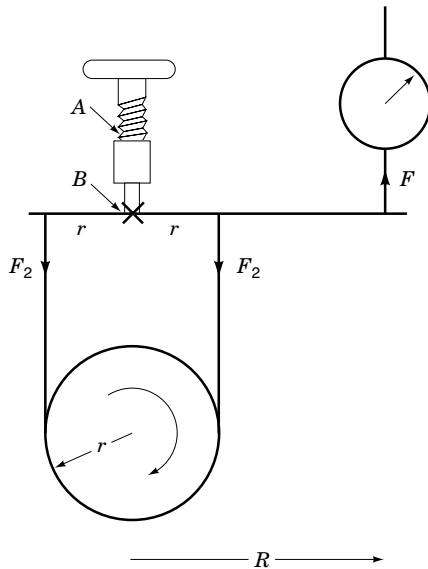


Figure 1. The Prony Brake. In this type the screw arrangement “A” is used to tighten the belt. “B” is the fulcrum of the lever. The force F shown on the spring balance is a measure of the torque.

a characteristic of the motor or engine driving the shaft for a particular supply voltage or fuel supply. The Prony Brake is generally used for small power levels at lower speeds because of the problems of conducting the heat away and sustaining the large forces involved at high power levels.

In essence, the Prony Brake transfers the torque in the rotating member to a static member where it is more easily measured. Other methods are also utilized to carry out this transfer. These include a hydraulic device, essentially an inefficient centrifugal pump, an air fan, various electromagnetic techniques, depending, for example, on eddy current losses, and the use of a generator with a resistive or other load to absorb the energy. Again, in each case, the reaction torque is measured. In these cases, a spring balance is not normally used but rather strain gauge load cells in the supports with the advantage that the torque can then be recorded automatically. Some of the methods depend on nonlinear properties in carrying out the power transfer and can be difficult to use over wide ranges of torque. Currently, the more basic devices such as the Prony Brake have applications, but it is now simpler to use a commercially available torque meter which is inserted into the shaft, the energy being absorbed in an appropriate way which for higher powers would be an electrical generator. This also has the advantage of being easily incorporated into an automated test system. It should not be forgotten that the simplest method may be to measure the reaction torque on the motor. This has the drawback that the inertia of the motor spoils the response of the measurement to fast changes in the torque.

Transmission Torque Measurement

In applications where the energy is transmitted rather than being absorbed in the measuring instrument, a commercial device inserted in the shaft is again the simplest solution, but in many laboratory situations and consumer products, it is better to integrate the device into the overall design. In the latter case, this would usually be on the grounds of convenience,

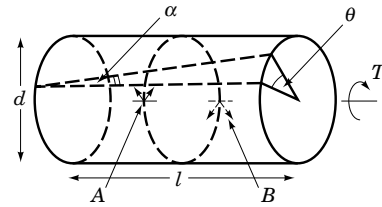


Figure 2. Shaft torsion. The torque T causes the torsion θ and shear α . The principal axes are shown at “A” and at the concealed point “B”. Strain gauges would be placed at these positions.

physical size, or cost. The techniques used depend on measuring the small distortions caused by the stresses related to the applied torque. Figure 2 shows the torsion, which is evaluated either by measuring the distortion of the surface or the twist between the ends. The former can be measured using strain gauges attached to the surface, this being a well-developed technology, and suitable gauges exist tailored to this situation. When a shaft is twisted, the directions of greatest stretch and compression on the surface are at 45° to the axis of the shaft, these being known as the principal axes of the surface strain. Figure 2 shows these, and strain gauges would be applied to take account of these. Combination gauges are available to measure strains in more than one direction at a point and are called rosettes. It is important that the gauges are attached to the surface in such a way that a signal can be derived which represents the torque but which does not respond to shaft bending, compression, or stretching. Attaching gauges on opposite sides of the shaft as indicated and differencing the signals eliminates the spurious responses, and standard electronics is available to process the signals. The problem is passing power and signals to and from the rotating shaft. In the past, sliprings have been used satisfactorily, but noncontacting methods are preferred for avoiding electrical noise and for long-term reliability. This has been achieved electromagnetically and optically. Usually, it is best to mount some of the electronics on the shaft so that the overall performance can be independent of the efficiency of the transmission across the gap. For example, the signal can be coded in a digital form. Alternatively, the magnetostrictive effect is exploited to measure the surface strain. This effect is the compression (or possibly extension) experienced when a ferromagnetic material is magnetized. The application of a stress to the material modifies this effect, and this can be measured. If the material is isotropic, that is, uniform in all directions, then there is no preferred direction to the effect, but for materials in a crystalline form, there are preferred directions. If the shaft, or a ribbon wrapped round it, is of suitable material, the effect can be exploited to measure the strain. Considering the principal axes of the strain, the magnetic permeability of isotropic materials is usually increased along the direction of greatest compression and decreased along the direction of greatest stretch. This can be detected by a coil system in proximity to the shaft, placed and oriented to take account of the directions of principal strain shown in Fig. 2.

If the shaft is known to be rotating, then direct measurement of the twist between the ends of the shaft section is most simply carried out by detecting the time difference for corresponding points on each end to pass fixed points. The methods used are electromagnetic, optical, and capacitive. It would be simplest to place small magnets on the shaft and pick up signals

in coils as they passed. However, it has been found to be better to use gear teeth on the shaft surrounded by static internal teeth. For ferromagnetic materials, the rotation changes the magnetic properties of the system through the ease with which the magnetic flux passes between the outer and the inner parts. This is detected by a coil. Optically, the path of a light beam is interrupted by lines on a disk mounted on the shaft. If electrical capacity is used, then some protrusion on the shaft can be detected as it passes an electrode. In the applications where the shaft may sometimes be stationary, it is still possible to arrange for signals to be generated in similar ways, but the interpretation is more difficult, and greater care is needed to obtain good performance. It should not be forgotten that for some electric motors the measurement of the input current, possibly allowing for the phase in the case of alternating current may give a sufficiently good measure of the torque supplied.

STATIC TORQUE MEASUREMENT

There are applications which are essentially static and are exemplified by a robot wrist. To apply feedback in the control of a robot arm, it is useful to know the forces and torques in the wrist connecting the arm to the gripper. Tasks which require feedback include the insertion of a peg in a hole, force-controlled assembly of mechanical parts, and screwing parts together. It also allows the detection of jamming. In general, the measurement of torque is more important than that of force, and this is discussed here; Fig. 3 shows the three torques about the three axes. The methods used depend on the small distortion of the wrist caused by the forces and torques. Because the movement is limited, the connections are by flexible cables, and there is no problem of power and signal transfer. Strain gauges are normally used, and the design skill lies in applying these to a suitable mechanical structure in such a way that the various forces and torques can be distinguished, possibly using a microcomputer. Again, the use of rosettes helps in this. There are many works on strain gauge technology, and the reader is referred to Ross (1) for a helpful introduction. Workers have also used magnetic inductance to detect the small distortions, magnetic circuits being sensitive to small changes in air gaps. Other workers have explored the use of electrical capacity in order to make compact devices at low cost. In all these cases, the possibility of integrating the electronics and incorporating it in the wrist is attractive. Falkner (2) has established that this is possible for driving a capacitor type of wrist sensor. In this, a remote host computer interrogates the sensor and receives

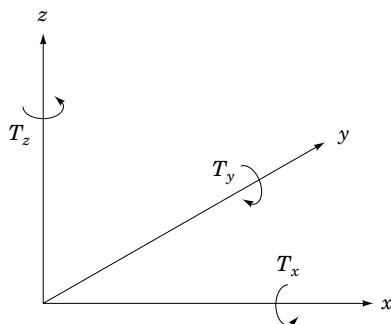


Figure 3. Torques about the three axes.

a digital signal in return. Workers in this field usually design and instrument their own devices. Static torque may also be measured on a large scale, including heavy machinery, industrial processes and construction equipment, and in these situations strain gauge technology is well-tried and usually most appropriate.

PRINCIPLES OF OPERATION

In this section, torque meters are discussed in fuller detail with more detailed analysis of the principles. As pointed out earlier, it is necessary to understand instruments which may no longer be current but which continue in use and which are perfectly satisfactory. This is certainly the case for many of those which absorb the power, and these are the first to be considered.

Power Absorbing Torquemeters

The Prony Brake has been discussed in the introduction and demonstrates the principle very clearly, the torque being transferred to a stator where it can be measured more easily. The hydraulic and electrical types are discussed more fully here, the former transferring the torque in one of two ways. There is the friction type in which disks on the rotor rotate between disks on the stator. The gaps are designed to be small to increase the friction, the velocity gradient in the fluid then being large. The friction is controllable by the quantity of fluid, usually water, in the device, although it will depend on the speed. The water will also be used for cooling to remove the heat generated. The nonlinearity of the effect can make operation difficult as speed instabilities are possible for some power sources. The second type works through agitation of the fluid and is essentially an inefficient centrifugal pump, the rotor comprising a disk with vanes on each side. This operates as a centrifugal pump for the fluid which is impelled into the stator vanes from where it is recirculated. The fluid gains and loses kinetic energy, thus generating heat. Control of the braking effect can be through varying the quantity of fluid and valves for controlling the fluid flow. Again, water is usually used and is circulated for cooling. The rotor revolves between fixed stator blades to which the torque is transferred. Again, the device is nonlinear and may not be easy to control. But in each case, relatively small devices can absorb large powers because of the cooling facility. The outer casing is mounted in bearings so that it would be free to rotate but for the constraints of the mounting which is instrumented to measure the forces. In Fig. 4, the torque is $(F_1 r_1 - F_2 r_2)$ Nm.

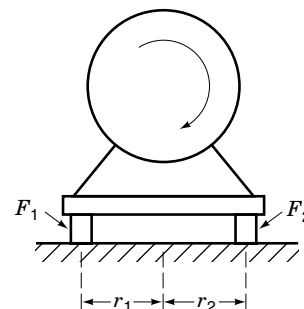


Figure 4. The reaction torque on the motor mounting. The torque is $(F_1 r_1 - F_2 r_2)$.

For simplicity, the principle is demonstrated by showing only two of the mounts. Devices have also been used which employ a fan working with air in a situation in which heat is generated, and the energy is absorbed.

The use of electrical braking for the torque transfer offers greater flexibility and the simplest method is that in which the shaft drives a generator which has a variable resistive load. Reducing the resistance increases the output current and increases the load and torque transmitted. It is also possible to feed the power into the mains. The casing of the generator is mounted in bearings, and the torque is measured as above. An alternative method of transferring the torque is through eddy current loading and, in devices using this principle, the rotor comprises a metallic disk of magnetic material. There are large teeth on this rotor which sweep past the inner surface of the field magnet during rotation. The relative motion causes rapid fluctuations in the magnetic flux passing between the rotor and the stator depending on whether they are opposite one another. This, in turn, causes eddy currents which flow near the surface in the stator. This process absorbs energy, and the eddy currents flow in such a way as to retard the rotor. The torque transferred is controlled by varying the magnetization of the stator through variation of the current in the stator coils. The torque will also depend on the speed of the rotor at lower speeds, but a useful feature is that the torque saturates above a critical speed and is then mainly dependent on the field current. This makes for smooth, controllable operation. The dissipation of the heat generated is usually achieved through water cooling, adding a complication in use. As in other devices, the reaction torque on the stator is measured using a strain gauge load cell or a linear voltage differential transformer to give the output. In the latter case, there is a spring to resist motion, and the small compression or extension is measured. Care must be exercised to ensure that the electrical leads and the cooling pipes do not contribute to the torque and cause an error. The reader is referred to Ambrosius (3) for further details of these instruments.

Direct Shaft Torque Measurement

These methods all depend on the torsion, or twist, of a length of shaft when the torque is applied. No power is absorbed. In Fig. 2, the torsion θ is given by

$$\theta = (32/\pi)(l/Gd^4)T \text{ rad} \quad (1)$$

with the shear angle α given by

$$\alpha = (16/\pi Gd^3)T \text{ rad} \quad (2)$$

where

d = the shaft diameter (m)

l = the shaft length (m)

G = the modulus of rigidity (N/m^2)

and T = the applied torque (Nm)

As an example, for steel, $G = 7.7 \cdot 10^{10} \text{ N/m}^2$ so that a length 100 mm of a shaft of diameter 25 mm would twist approximately 3.3 mrad when a torque of 100 Nm was applied. This is equivalent to a movement of approximately 0.33 mm at the surface. A sensitivity of 1% of this corresponds to a surface movement of $3.3 \mu\text{m}$. Clearly, if the shaft is made thinner at this

section or a longer section is used, then the twist is greater, and the measurement is easier. The problem is that excessive flexibility causes distortion of the system being measured. In some cases, depending on the characteristics of the power source and the load, flexibility may lead to resonances building up. In other cases, the flexibility might reduce the response to higher frequencies in the measurement. For the use of strain gauges or magnetostriction, which depend on the local surface distortion, the angle α is critical but not the length l . Strain gauges are considered first.

Strain Gauges. This is a well established technology, and the correct use in this application takes account of the strain pattern in a twisted shaft. There is a shear on the surface which is perpendicular to the axis and if there is no bending of the shaft or any axial strain, then there is a strain system which has maximum and minimum values along the principal axes at 45° to the axis of the shaft. These are shown in Fig. 2 and are the most sensitive directions for detecting the surface distortion with strain gauges. The strain theory is discussed in Ross (1), and for the strain angle α , the principal strains σ_1 and σ_2 are $\pm(\alpha/2)$. In the example given above, these are ± 163 microstrain. Details of the operation of strain gauges are given in the appropriate section of this encyclopedia, and the gauges are placed to detect the principal strains, probably using a rosette which comprises more than one element. The gauges on the opposite side of the shaft are utilized to avoid spurious operation when the four gauges are connected into a bridge as shown in Fig. 5. The bridge output is

$$V = [(R_2 R_3 - R_1 R_4) / 4R^2] E \quad (3)$$

where the resistors all have nominal value R , and the small size of the increments due to the strain are taken into account. For pure torsion, the strains $\pm(\alpha/2)$ cause proportional changes in the resistors of $\pm K(\alpha/2)$ where K is the gauge factor of approximate value 2 for most materials used. Thus, R_1 becomes $R(1 + K\alpha/2)$ etc., and

$$V = (K\alpha/2)E \quad (4)$$

The torque is given by

$$T = (\pi G d^3 / 8K)(V/E) \quad (5)$$

If the shaft is under longitudinal tension or compression, the output is unaffected because all the gauges are affected in the same proportion. Similarly, bending of the shaft affects the resistances in such a way that there is no contribution to the output. The required gauges are readily available but require

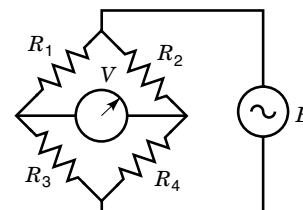


Figure 5. The strain gauge bridge. Unbalance between the resistances $R_1 - R_4$ of the strain gauge elements causes an output voltage V .

Table 1. Manufacturers of Torquemeters

Torquemeters Ltd., Ravensthorpe, Northampton, NN6 8ET, UK, +1604 770232
Vibro-Meter SA, Rte de Moncor 4, CH-1701, Fribourg, Switzerland, +37 87 11 11
Industrial Measurements Limited, Melbourne, Derby, DE73 1DY, UK, +1332 864 0000
6-Axis Torque Sensor: Lord Industrial Automation Division, Lord Corporation, 407 Gregson Drive, Cary, NC 27511, +919-469-2500

care in mounting, and the electronics units are also available. The problem is how to connect the power and signals to the rotating shaft. The most direct method is through sliprings, but these can cause electrical noise and would not be considered reliable in the long term. IML (Table 1) manufactures a range of noncontacting strain gauge torquemeters. Electromagnetic coupling has proven successful, and Zabler, Dukart, Heintz and Krott (4) describe a system for detecting the torque in the steering column of an automobile as part of a power assisted steering system. The particular requirements are low cost associated with high reliability and high performance. The last mainly concerns the stability of the null position because the output forms the error signal for the servo system. The non-contacting power and signal transfers are achieved through annular transformers. The inner annulus rotates, but the outer one is static, and the transformer action is unaffected by the rotation. The researchers decided against the simplest configuration in which the ac power supply was transmitted by one transformer and then fed the bridge. The other would take the output of the bridge, and the electronics would all be off the shaft. This arrangement was considered sensitive to noise for the low output signal and to the variations in sensitivity during rotation when the gap would vary using normal manufacturing tolerances. Rather, the electronics was mounted on the shaft supplied by one of the transformers. The bridge output was amplified and converted to a pulse width modulated form for minimum noise sensitivity when transmitted across the gap. In this application, special strain gauge elements were used to reduce cost. This approach resulted in a robust and high performance system.

Magnetostrictive Devices. In these devices the change in the magnetic permeability of certain materials when stressed is exploited. In principle, the effect is considerably more sensitive to strain than the strain gauge effect. This is discussed by Beihoff (5). Exploiting this depends on successfully subjecting the material to the strain. With strain gauges, the technology for attaching the gauges is well developed, but attaching a thicker, more rigid, material is more difficult. The alloys which are useful in this context are nickel-iron, nickel-cobalt, and high nickel steels. One solution is to manufacture a section of shaft from this material and give it the required thermal mechanical pretreatment. This appears to be expensive, and there may be a compromise over the mechanical properties. More usually, a layer is attached in the form of a ribbon. As explained earlier, the directions of maximum positive and negative stress are in directions at 45° to the axis, and the physical arrangement needs to take account of this. The simplest approach is the coil arrangement shown in Fig. 6. If there is no torsion, then the inductances of the two arms B and C are balanced, and there is no output from the electrical bridge. If

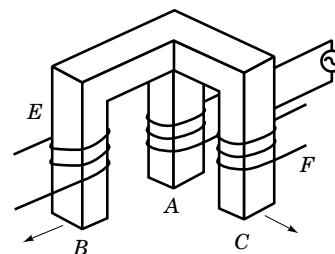


Figure 6. Magnetostrictive torque measurement. This device compares the properties of the surface along AB and along AC. Unbalance is detected by the difference in the signals in coils E and F. This figure is reproduced from *A Survey of Torque Transduction Methodologies for Industrial Applications* by B. Beihoff in the Proceedings of the 1996 Annual Pulp and Paper Industrial Technology Conference. © 1996 IEEE.

there is torsion present, one inductance is increased, and the other decreased, unbalancing the bridge and resulting in an output signal that depends linearly on the torsion over a useful range. Koga and Sasada (6) report an instrument using a more advanced coil system. Figure 7 shows the figure of eight coils and their orientation with respect to the principal axes on the adjacent shaft. This ensures that the overall measurement of the effect is localized. Sections A and B with sections C and D of the magnetic core form one inductor with coil E. Sections A and C with sections B and D with coil F form the other, and the lack of balance again gives the required signal. The compact design is mainly to reduce the effects of variations in the gap between the sensor and the shaft and variations in the magnetic properties of the shaft. The effective sensor for each of the principal axes uses the same region of the shaft surface. The inductance is particularly sensitive to very small variations in the gap because of the high relative permeability of the core material. In the work described, two sensors are used on opposite sides of the shaft which is of carburized steel. They use a carrier frequency of 60 kHz, and for a torque range of ± 400 Nm with a 25 mm shaft, the linearity is $\pm 0.6\%$. In this first work, there are fluctuations in the performance, including those which are caused by rotation of the shaft. An alternative way of measuring the magnetostrictive effect is to exploit the

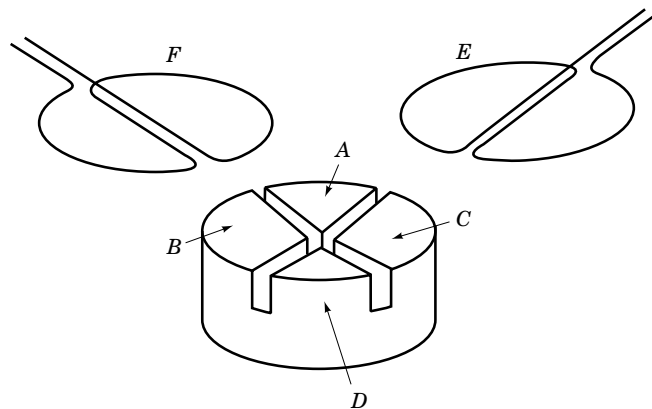


Figure 7. A more compact arrangement for magnetostrictive torque measurement. This figure is reproduced from *Static Characteristics of Torque Detection from Carburized Steel Shafts* by F. Koga and I. Sasada in the *IEEE Trans. Magn.* 31 (6): 1, 1995. © 1995 IEEE.

coupling between two adjacent coils as suggested by Sasada and Koga (7). One sets a magnetizing field parallel to the axis, and the other detects any magnetization of the shaft at right angles to the axis. When there is no torsion, there is no signal, but the torsion causes anisotropy in the permeability and magnetisation at right angles to the axis. This has been found to be proportional to the torsion over a useful range.

Phase Measurement. Devices which depend on the time difference in signals picked up from the two ends of the shaft section when it is rotating use electromagnetic, optical, or capacitive pick-ups. This may also be thought of as the measurement of the phase difference between the two signals. In general, such devices will not operate while the shaft is stationary or rotating very slowly. An example of an electromagnetic system is the range of Torquetronic torque meters manufactured by Torquemeters Ltd. (Table 1). The essential mechanism comprises internal and external gears with teeth. The outer contains a coil, and the magnetic circuit around the coil changes as the teeth rotate and pass through positions in which they are opposite to each other. If, for example, there is a constant current fed into the coil, then a voltage is generated as the magnetic circuit changes. The manufacturers point out that the use of several teeth in this way compensates for problems associated with any radial movements. The possible torque range of this device depends on the shaft dimensions, and meters with ranges as high as 75,000 Nm are available. These particular devices also record the speed. The problem of lack of signal if the shaft is stationary will not matter in many applications, but it is overcome in a range of devices in which the outer teeth are rotated, the coil still being stationary.

The simplest optical system suitable in a laboratory situation is the use of a stroboscope enabling the twist to be seen visually. A commercial instrument by Amsler is described in Ambrosius (3) with particulars of its source. The power is transmitted through the torsion rod which, at one end, has a disc attached with gradations around the edge. Concentric with this rod and outside it is a tube which is not subject to any torque. This carries the information about the angular position of the other end to another smaller flat ring which also has a graduated edge. The two scales are adjacent, and the torsion causes relative motion of the two scales, a vernier scale being used for accuracy. To inspect the scales during rotation, a synchronized stroboscope is used. Clearly, it would not be feasible to inspect the scales for a general static position, and the device depends on rotation. Measurements are best made for steady torques, and it is not possible to automate the operation of the device.

Other Optical Devices. Another type of optical torque meter uses mirrors. Again, rotation is required for successful operation. The principle is that if a light beam is reflected off two parallel mirrors in sequence, then the direction of the ray is unaltered; although, of course, it will be shifted sideways. The meter is designed so that one mirror is attached to one end of the flexible shaft and the other to an outer cylinder which is connected to the other end in a similar manner to the optical instrument described above. The two mirrors are at the same position at the end of the shaft. A collimated light ray, possibly from a laser, strikes the mirrors in turn for a small part of the time of rotation. During this time, the change in direction depends on the lack of parallelism, equivalent to the angle of

twist of the shaft. The transmitted ray is focused onto a scale, and the position of the light spot corresponds to the twist in a linear manner. Further details are given in Adams (8).

Beyond this, the simplest system is one that depends on the rotation of the shaft and uses a disk mounted axially on the shaft at each of the two positions. Each contains a pattern of radial lines and is otherwise transparent. Units comprising light-emitting diodes and photo cells detect the motion of a disk, and the twist in the shaft gives a phase difference at the two positions. This is measured as described above. A different approach is required if the shaft is stationary or slow moving, and to overcome this problem, Hazelden (9) has developed an optical sensor for use in an electrical power assisted steering system. In this, the light passes through both the disks which are close to each other, the patterns being designed so that there are two outputs whose ratio determines the torsion, taking account of sign. This arrangement makes the signal independent of the light intensity. An optical sensor is suited to the electrically noisy environment. The performance is not claimed to be suitable for general instrumentation but it is adequate for this application and capable of being manufactured at the low cost which is demanded.

Inductive Devices. The principle of these instruments is the variation of the transformer coupling between two coils through the variation in the magnetic circuit caused by the distortion of the shaft. There are concentric cylinders attached to each end of the flexible section of the shaft. These contain slots arranged so that when there is no twist, they just do not overlap so that there is no coupling between the coils. The deformation caused by twisting causes overlap and creates the magnetic circuit allowing coupling. With careful design, the coupling is proportional to the applied torque, and it is also possible to make the system detect the direction of the torque so that the electronics generates a dc output signal of sign depending on the direction. This device has no lower speed limit. Vibrometer (Table 1) manufactures a range of instruments using this principle with measurement ranges of ± 1 to ± 500 Nm. The bandwidth is 1 kHz. The linearity is of the order of $\pm 0.2\%$ over a wide range of environmental conditions.

Capacitive Devices. Capacity has been exploited to detect the shaft torsion, and the methods to be described are ones in which the variable capacitor or capacitors form part of the rotating shaft. An older design developed by the Royal Aircraft Establishment, Farnborough, is described in (10). In this, there are two concentric cylinders, one connected to each end of the shaft so that one is inside the other with a small gap. The torsion causes a small relative twist which is detected by teeth, or serrations, on the outside of the inner element and on the inside of the outer element. When these coincide, the capacity between the inner and outer is a maximum and reduces as the relative angular position moves from this state. By using square teeth and a very small gap, there is a region in which the capacity varies linearly with the angle, and this is used. The capacity is measured from outside the device through sliprings and a bandwidth up to 80 kHz is reported. On those applications where vibrations are being looked at, long-term drift will not matter. In other cases, it is better to use a bridge technique to remove temperature and other effects. The technique is also sensitive to stray capacities. Today, sliprings would be considered a possible source of unreliability.

An alternative use of capacity is that developed by Falkner (11). The torsion between the ends of the flexible section of shaft is detected through two capacitors. The plates of these are connected to the ends in such a way that the torsion increases the gap in one and decreases it in the other. Capacitors are more sensitive to gap changes than to relative lateral motion of the plates. The two capacitors form a bridge and the three connections to the bridge are connected capacitively to the stator through rings with small gaps, thus maintaining the requirement for no contact. Such bridges have been found to be very stable and sensitive at the submicrometer level. In this situation, the performance is affected by these coupling capacitors, but it is the stability of their values which is critical. They are made as large as practical, which can be shown to reduce these effects. The prototype is extremely sensitive and has a range of ± 0.9 Nm with nonlinearity better than $\pm 1\%$ of full scale. There is some residual variation in the readings as the shaft is rotated, the sensitivity and zero reading each varying $\pm 1\%$ of full scale. There would be no difficulty in designing a stiffer torsion device to operate at higher torque levels.

Static Torque Measurement

The situation has already been described above in the introduction and represented diagrammatically in Fig. 3, and again the robot wrist example is considered first. The most obvious approach is to use strain gauges on an appropriate mechanical configuration. For example, gauges could be arranged around a metal wrist block. In this case, the signals obtained would each contain information about more than one of the torques and forces present, but the use of the appropriate rosettes and bridges as described above would help the situation. There is also the possibility of using a microprocessor program to distinguish the individual outputs. More satisfactory is the use of a metal cross, in which case the arm and the gripper are joined to the center and the outside of the unit which is designed to give enough flexibility for useful signals. In this case, the arms are instrumented with the gauges on all the faces, and inevitably, there is some mixing of the signals so that it becomes essential to use a microprocessor. Such a device has been made commercially by Lord Automation (Table 1) with an accuracy of approximately $\pm 2\%$ and some residual cross-coupling. Other workers have used inductance and capacity to detect strain in the wrist. An example of the former is by Piller (12), the compliant element being essentially three helical springs between two plates. Any force or torque applied to this causes a small distortion, and the signals from a set of inductive pick-offs are interpreted to distinguish the forces and torques. Falkner (2) has used capacity in a wrist in which the compact mechanical design allows complete separation of the three forces and the three torque signals. Small distortions occur, each of which depends solely on one of these parameters. Each measurement is made capacitively. Figures 8 and 9 show the two types of basic elements used for the torque measurement. That, in Fig. 8, works on a torsion bar which the applied torque twists, unbalancing the capacitor bridge comprising the two capacitors shown. The antiphase drive signals are applied to the two fixed plates, and the signal pick-off is the moving element A. In Fig. 9, the device responds to twist about the axis to unbalance the capacitor bridge. Again, the

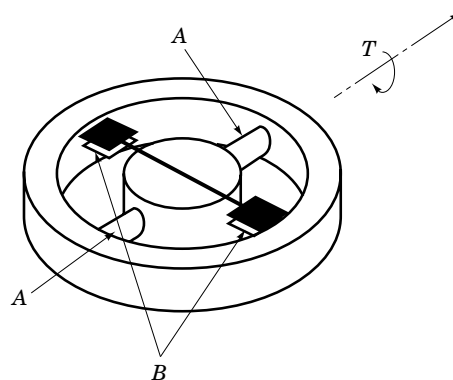


Figure 8. A capacitive torque sensing element. "A" is a torsion bar and torque unbalances the capacitor bridge "B." This material has been reproduced from the Proceedings of the Institution of Mechanical Engineers, Part C, *J. Mech. Eng.* **206**: 1992, by A. H. Falkner by permission of the Council of the Institution of Mechanical Engineers.

two moving plates are connected together to form the pick-off. In each case, torsion increases one capacitor and reduces the other, and the signal measures the amplitude of the torque and, through the phase, the sign. A compact three-axis torque sensor has been built using two elements of the type shown in Fig. 8 and one of the type shown in Fig. 9. To simplify the electronics, it is possible to multiplex the drive signals to the bridges and make the outputs common. The use of square waves makes the multiplexing simple using integrated electronic circuits of the complementary metal-oxide-semiconductor family. A prototype application specific integrated circuit has been made to drive the circuit and interface to a computer with the objective of fitting the electronics in the device and having the simplest possible electrical connections. Otherwise, it is possible to use long leads with no electronics in the device. It has been found

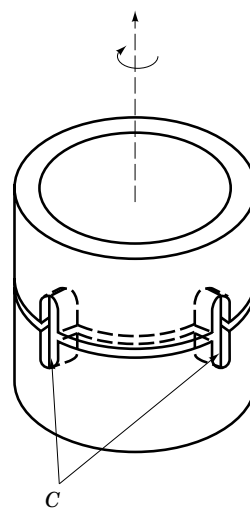


Figure 9. In this configuration, axial torque causes torsion through the flexing of the leaves "C." This material has been reproduced from the Proceedings of the Institution of Mechanical Engineers, Part C, *J. Mech. Eng. Sci.* **206**: 1992, by A. H. Falkner by permission of the Council of the Institution of Mechanical Engineers.

that the capacitor bridges are extremely stable in this situation, and overall performance is at the 1% level. The correct mechanical design results in complete isolation of the axes. For larger scale applications, including heavy equipment, the robust, well-tried techniques based on strain gauges are most appropriate.

CURRENT DEVELOPMENTS

The above shows that there are a number of well-tried methods available for measuring shaft torque and for measuring torque in static situations, such as the robotic wrist. The requirements vary, but the present emphasis in research is towards low-cost devices with high performance. The latter includes accuracy, sensitivity, and linearity as well as noise immunity and wide bandwidth. Environmental conditions vary, but there may be a requirement for operation over a wide temperature range and for ruggedness. In the case of an automobile power assisted steering system, for example, the device is expected to operate with little maintenance over many years in a hostile environment. The system must be designed so that any failure is safe in the sense that no assistance would be supplied, rather than a false turning which the driver could not control. To meet these challenges, there is work in the development of existing techniques and in the investigation of new ideas. It is difficult to predict which will be most successful, but it is always likely to be the case that a range of techniques will be used. The availability of application specific integrated circuits in electronics, which are inexpensive even in small quantities, means that as much complication as possible should be in the electronics with the simplest possible electrical and mechanical components. It is possible to incorporate a microprocessor to measure nonlinearities and temperature and compensate for these. This can remove the need for tight mechanical tolerances and materials specifications. These developments have led to falling costs across instrumentation in general and have made possible the use of new techniques. There is also interest in fully integrated sensors for many purposes. In these, micromachining of mechanical parts is undertaken on the silicon chip which contains the electronic circuitry. The wide range of adhesives now available may make an important contribution. Below is a survey of the current research which appears to be of most promise.

Magnetostrictive Devices

There is considerable activity in the field of the magnetostrictive sensors. This centers around the selection of the material itself and the use of ribbons as well as treatment of an existing shaft. Otherwise, compact pick-ups are being developed. For ribbons, the work is directed to the choice of material and the problem of attachment to the shaft. An example is work by Hase, Shoji and Wakamiya (13). The gluing process requires raising the temperature, and during the cooling, the differential coefficient of expansion results in a compressive stress in the ribbon. This compression is biased in direction, and this direction will coincide with a principal axis. The magnetic properties are affected in a way which can be exploited to arrive at a device with uniform characteristics over a wide temperature range. Hirose, Takagi, Mori and Imura (14) describe the following process for the fabrication of an amorphous magnetic alloy

layer. A crystalline layer of the alloy is attached to the surface by silver brazing and then melt quenched by laser irradiation. It is found that the easy direction of magnetization is along the direction of the laser scanning. This feature is exploited by making this direction along one of the principal axes at 45° to the shaft axis (Fig. 2). A further refinement is to make the treated area in stripes in this direction. The coil structures required are described above in the section on magnetostrictive devices, and research is directed to making these more compact. In particular, Rombach, Steiger and Langheinrich (15) have developed a micromachined torque sensor in which a magnetic layer and coils are deposited on the slice of silicon containing the electronic circuitry. The magnetic layer becomes the core of the coil and this is the equivalent of the Torquemaster (Table 1) devices which are described above. The manufacture could be highly automated. One objective is the compensation for variations in the gap between the shaft and the sensor which occur with normal machining tolerances. This is achieved by measuring the field in different positions around the sensor head to allow compensation.

Surface Acoustic Waves

Devices are being developed based on the effect on the propagation of surface acoustic waves of strain in the material used. The purpose is to develop instruments which require the minimum part on the rotating shaft and which can be treated as a part of an electrical circuit on the stator. In this case, the signal is connected by the electrical capacity of the gap, this being easy at the very high frequencies used. The corresponding impedances are relatively low, and the overall system is robust. It has been known for a long time that sound waves propagate in the surface of isotropic, elastic materials. It is possible to interface from electrical signals to the acoustic signals using a piezoelectric substrate, and devices which include resonators and filters can be fabricated in a small space because the wavelengths are small. A strain sensor depends on the effect of strain in the substrate on the wave velocity and is mounted on the shaft along one of the principal axes. For the case of a resonator, this affects the frequency, independently of the gap spacing. Typical frequencies are 200 MHz to 400 MHz corresponding to convenient physical dimensions of 1 or 2 mm. Hall, Wang and Lonsdale (16) describe how they have mounted the resonant element on the shaft. The coupling is capacitive as described above so that the resonator forms part of the circuit. Inductive coupling would be an alternative. The overall resonant frequency is measured using well established electronic techniques to give a measure of the strain. In the work reported, the emphasis has been on overcoming the sensitivity of the sensor to the temperature by using various configurations and cements for attaching the sensor. The current situation is that successful development would result in a very useful instrument, but further work is needed.

Other Developments

In the devices in which the sensor is mounted on a rotating shaft, the use of light to transfer the signal, and even the operating power, is attractive. The system can be designed so that variations in the light intensity are not important. In the case of the power, a simple regulated power supply is needed on the shaft. Electronic circuitry is available that consumes an

extremely small amount of power of the order of a few mW or less, and the main user of power will be the light emitting diode which transmits the signal back. The whole system becomes feasible with small standard optical devices. To ensure the noise immunity and tolerance to light power variations, a pulse code or digital code is used. The practical problem is arranging for light to be transmitted at all angular positions of the shaft. An example is given by Mu and He (17) in which they use a standard strain gauge bridge and report that the signal is transmitted in Hamming code using infrared light. The overall accuracy for measuring the torque is $\pm 0.1\%$. Dezhong and Zesheng give another example (18). They use light in the visible spectrum, and the power is picked up by a ring of photocells which allows operation in daylight or through an optical fiber. The modulation technique is to use a multivibrator with two strain elements switched alternately into the timing circuit leading to a form of pulse width modulation. The signal is transmitted by a light emitting diode back to a ring of photo diodes and the demodulator compares the width of alternate pulses. The electronics uses standard low power integrated circuits, and the linearity is of the order of $\pm 1\%$.

There are current developments in the capacitor techniques for measuring the torsion. Cooper (19) describes an instrument in which a variable capacitor on the shaft forms part of a tuned circuit, all the other elements being on the stator, and the coupling being capacitive. The capacity variation is achieved by using the torsion to move a dielectric material in the gap of the capacitor, thus varying its value. The units have been designed to fit in a small, hand-held electric screwdriver, the signal being required for feedback in the control of the operation. To achieve a linear relationship between the torsion and the overall output, the capacitance forms part of a multivibrator circuit so that simple digital counting leads to the torque value. The sensitivity is 0.005 Nm and the linearity is $\pm 0.6\%$ but the reported results show that further work is required to remove sensitivity to environmental conditions and to the angular position. It is suggested that a bridge technique might be helpful in this respect. Falkner (20) has investigated how capacity can be exploited with the minimum changes to the shaft. From the point of view of ease of manufacture, splines which already exist, in many cases, are suggested, but if the shaft may sometimes be stationary, then a timing method will not work. The only signals available are capacities to the grounded shaft, these varying with angular position. These can be measured at the two ends from internal splines on the stator. Although these readings will, in general, be different, there is ambiguity in the interpretation, and readings at or near the peak values are not very dependent on the position. To overcome this, a system of four external elements at each end is proposed which generates a signal whose phase depends on the angular position passing through 360° for each spline pitch. Then, the phases are compared at each end to give the torsion without ambiguity. Assessment of this proposal awaits the making of a prototype. Other techniques which have been investigated include the Faraday effect, the piezoelectric effect, and the use of the Hall effect to measure unbalance in magnetic bridges.

BIBLIOGRAPHY

1. C. T. F. Ross, Ch. 7, *Applied Stress Analysis*, Chichester: Ellis Horwood, 1987.
2. A. H. Falkner, The measurement of force and torque using capacitor devices, *J. Mech. Eng. Sci.*, C, **206**: 385–390, 1992.
3. E. A. Ambrosius, *Mechanical Measurement and Instrumentation*, New York: Ronald Press, 1966, p. 280–303.
4. E. Zabler et al., A noncontact strain-gage torque sensor for automotive servo-driven systems, *Sensors Actuators, A*, **A41-A42**: 39–46, 1994.
5. B. Beihoff, A survey of torque transduction methodologies for industrial applications, *Proc. IEEE Annu. Pulp Paper Ind. Technol. Conf.*, 1996, Birmingham, AL, 220–229.
6. F. Koga and I. Sasada, Static characteristics of torque detection from carburized steel shafts, *IEEE Trans. Magn.*, **31**: 1, 3143–3145, 1995.
7. I. Sasada and F. Koga, A new structure of torque sensors using thin pick-up head—use of mutual coupling modulation, *J. Appl. Phys.*, **75** (2A): 5916–5918.
8. L. F. Adams, *Engineering Measurements and Instrumentation*, London: Hodder and Stoughton, 1981, p. 128–129.
9. R. J. Hezelden, Application of an optical torque sensor to a vehicle power steering system. In J. Giber et al. (ed.), *Proc. Euroensors VII*, **2**: 39–46, Lausanne: Elsevier, 1994.
10. M. Hetényi, *Handbook of Experimental Stress Analysis*, New York: Wiley, 1950, p. 287–293.
11. A. H. Falkner, A capacitor-based device for the measurement of shaft torque, *IEEE Trans. Instrum. and Meas.*, **45**: 835–838, 1996.
12. G. Piller, A compact six-degree-of-freedom sensor for assembly robots, *12th Int. Symp. Ind. Robots*, 1982.
13. H. Hase, R. Shoji, and M. Wakamiya, Torque sensor using amorphous magnetostrictive ribbons, *Mater. Sci. Eng., A*, **A181-A182**: 1378–1382, 1994.
14. F. Hirose et al., Characteristics of an amorphous magnetic alloy layer prepared by laser quenching, *Mater. Sci. Eng., A*, **A181-A182**: 1359–1362, 1994.
15. P. Rombach, H. Steiger, and W. Langheinrich, Planar coils with ferromagnetic yoke for a micromachined torque sensor, *J. Micromech. Microeng.*, **5** (2): 136–138, 1995.
16. D. A. Hall, H. W. Wang, and A. Lonsdale, Torque Measurement by Means of a SAW Resonator, *Sensors VI. Technology Systems and Applications*, 201–206, 1993. Manchester, UK.
17. N. X. Mu and S. S. He, The infrared transmitting torque meter, *Conf. Proc. IEEE Conf. Advanced Technol. Instrum. Meas.*, **1**: 384–5, 1994. Hamamatsu.
18. Z. Dezhong and D. Zesheng, A new torque instrument driven by light power with two rings, *Proc. IEEE Int. Conf. Industrial Technol.*, New York, 510–513, 1994.
19. J. D. Cooper, Novel torque sensor for mechatronic applications, *4th IAESTED Int. Conf. Robotics Manuf.*, Honolulu, August 1996.
20. A. H. Falkner, Non-contact shaft torque measurement—A new technique, *Proc. 12th Int. Conf. Syst. Eng.*, Coventry, UK, 1997.

Reading List

References 4 and 5 in the Bibliography contain useful surveys which would be helpful in selecting the appropriate technique for a particular application. Reference 3 surveys older equipment which may still be encountered. Reference 1 gives the mathematical background to the stresses generated when torque is applied. Most technical libraries hold several books on instrumentation and measurement and many contain work on torque sensing. It is also suggested that information from the first three manufacturers listed in Table 1 is very helpful.

ANTHONY H. FALKNER
Coventry University

WEIGHING

The term “weighing” usually implies a measurement of mass. The mass of a body is a measure of the amount of substance it contains. A direct measurement of mass is not simple, but we are accustomed to sensing the mass of a body in terms of its weight, that is, how “heavy” it is. The weight of a body is dependent on its mass and also on the strength of the gravitational field in which it is located. Comparison of the weight of one body with the weight of another in the same gravitational field is a relatively simple task, and because much of our experience occurs within a common gravitational field, that of the earth, we can use the process of comparison of weights as a means of measuring mass. But in doing so we must be careful not to confuse mass and weight, remembering that they are essentially different quantities.

There are ways of measuring mass that do not rely on the presence of a gravitational field. One can for example measure the *inertial mass* of a body by measuring a change in momentum. But these methods are only appropriate in special circumstances and will not be described here.

As we shall see below, the measurement of weight is essentially the measurement of a force—specifically, the force exerted by a body in a gravitational field on its support. The second part of this article concentrates on the measurement of force, with particular reference to force measurements for the purpose of weighing.

MASS AND WEIGHT

We should first establish the meaning of the terms mass and weight. Mass is a measure of the amount of substance in an item. Everything has mass, and the mass of a body is independent of its other properties. Unfortunately, human beings cannot detect mass. However we are sensitive to the force that results from a body’s mass. This force is revealed in two ways, either by the inertia of the body (inertia is the tendency of a body to resist a change in motion), or by the attraction between bodies. The attraction between bodies provides the most common means of measuring mass, through weighing.

All bodies are attracted to one another. The strength of the gravitational attraction between two bodies results in an acceleration of each towards the center of mass of the other. In the case of a body whose mass we wish to measure in the earth’s gravitational field, the acceleration of the earth toward the body will not be apparent (and will be very small), so we need only consider the acceleration of the body toward the center of mass of the earth.

A body supported on the surface of the earth senses the acceleration of gravity but does not move. Instead it exerts a force on its support. This force is what we call weight. In the absence of an atmosphere the weight of a body is equal to the product of its mass and the gravitational acceleration it is subjected to. (The earth’s atmosphere complicates matters by creating an upward force proportional to the volume of the body, effectively reducing its weight. We shall discuss this issue in detail below).

So we have seen that while mass is a fundamental quantity of a body, weight is dependent on mass and on the presence of a gravitational field. In common usage the term “weight” is often used incorrectly to mean “mass,” a practice that causes much confusion when discussing the measurement of mass by

weighing. For the remainder of this article we shall carefully distinguish these two quantities and use the correct term for each.

However, we must also be aware that the word “weight” can have another meaning. We have already seen that it can mean the attracting force between a body and the earth. It can also mean an artifact used in weighing as a reference standard of mass. This can lead to confusion, so care must be taken to ensure that the meaning is clear (1).

A BRIEF HISTORY OF WEIGHING AND MASS STANDARDS

A History of Weighing

Weighing, by which we mean using a balance to measure the weight of a body or to compare the weight of two bodies, has been done for thousands of years. Images from the earliest civilizations in the Middle East show weighing using a beam balance. The process has continued almost unchanged until the present day. As recently as the medieval times, monetary value was measured by weighing. The coin in use in Britain was the silver penny, chiseled in two for a halfpenny and in four for a quarter-penny or farthing (fourthing). The pennyweight was 1/240 of the “pound of Troyes” (named after a city in France which was the center of mercantile trade in the twelfth century). This pound was 5760 grains—barleycorn grains selected in a precise manner. The *troy* weight system (20 pennyweights to the troy ounce and 12 ounces to the pound) was used for all precious materials and for bread, the price of which was controlled by law from the twelfth until the nineteenth century. (In 1280, a “farthing loaf” weighed about $3\frac{1}{2}$ pounds.)

A later system of weights introduced the “haber de peyse” or *avoirdupois* pound, which was eventually fixed at 7000 grains, and divided into 16 ounces each of 16 drams. The only direct connection between the values of the troy and avoirdupois systems is the grain of barleycorn.

Another important measure of wealth in Britain was wool, which by the fifteenth century accounted for over half of the country’s export trade. For this, larger measures of weight were needed, and the lowest value in common use was the stone, which varied in value for many years but eventually settled at 14 avoirdupois pounds. Two stones made a quarter, and four quarters a hundredweight. Three hundredweights made a sack.

Fine weighings were needed for medicines, and for that purpose the *apothecaries’* measure was based on the Troy ounce—20 grains to the scruple, three scruples to the drachm, and eight drachms to the ounce (2).

When dealing with all these different systems of weight measurement, the benefit of a single, universally accepted system soon becomes apparent. The metric system, based on the kilogram, is now in use in most of the world.

The Birth of the Metric System

In 1790 King Louis XVI of France commissioned his country’s leading scientists to recommend a consistent system for weights and measures. The report that the French scientists (Lagrange, Lalande, Laplace, Borda, Monge, and Condorcet) presented to the Academy of Sciences on March 19, 1791 recommended a system based on a unit of length, the meter, equal to one ten-millionth part of the arc of the Earth’s quadrant, pole to Equator. The unit of mass would be equal to the mass of a

defined volume of water *in vacuo* at its freezing point. Legislation authorizing construction of this new system of units was passed on March 26, 1791, and the metric system of measurement was born (2).

After the 1791 report, measurements were made to decide an appropriate volume of water for the standard of mass. In 1775 it was agreed that the unit would be the mass of one cubic decimeter of water at a temperature of 4°C (which is approximately the temperature at which the density of water has a maximum value). This unit would be called a kilogram, and the mass of one cubic centimeter of water would be called a gram.

Brass weights were made whose mass was equal to the new unit, the kilogram, then later a weight of platinum was made, and adjusted to the value for the new unit (18,827.15 French grains). This platinum weight became known as the *Kilogramme des Archives*, and was effectively the standard of mass for most of Europe. It was soon realized that the mass of this kilogram artifact was not exactly that of a cubic decimetre of water; thus the link between mass and volume measurements was effectively broken.

International interest in measurement standards grew quickly, and in 1870 and again in 1872 the French government called meetings to discuss the construction and distribution of new metric standards. At a third meeting, in 1875, eighteen countries subscribed to a treaty, called the *Convention du Metre*, by which the *Comité International des Poids et Mesures* (CIPM) and the *Bureau International des Poids et Mesures* (BIPM) were set up, to be responsible for the custody and verification of metric standards. The convention agreed that a new kilogram weight should be made using an alloy of platinum and iridium (90%:10%). After many attempts in France, a successful casting of the alloy was made by George Matthéy of Johnson, Matthéy and Co. of London, and in 1879 three cylindrical pieces of the alloy were delivered to a French metallurgist, St-Claire Deville. The cylinders were hammered in a press, and then polished and adjusted and finally compared with the *Kilogramme des Archives* by M. Collot, a maker of weights and balances. By 1883 the CIPM were convinced that one of the cylinders “was indistinguishable in mass from that of the *Kilogramme des Archives*,” and this weight was chosen as the international prototype of the kilogram (3).

A further 40 one-kilogram weights had been ordered from Johnson, Matthéy & Co. in 1882 and were delivered in 1884. After reheating and hammering to increase their density, these were adjusted to be close in mass to the newly selected international prototype. In 1889 each signatory of the *Convention du Metre* (by now 20 countries) was allocated one of these weights. The allocation was made by lot, the United States being given Copy No. 20 and the United Kingdom receiving Copy No. 18. The certificate that accompanied Copy No. 18 on its first journey to the United Kingdom gives its mass as 1 kg + 0.070 mg, with an uncertainty of ± 0.002 mg, and its volume as 46.414 ml at 0°C. To this day Kilogram 18 is the cornerstone of measurements of mass in the United Kingdom. At its last calibration at the BIPM in 1997 its mass was found to be 1 kg + 0.061 mg ± 0.004 mg, and its volume at 0°C (calculated from the original value in milliliters) as 46.4149 cm³.

The Unit of Mass in the SI System

The unit of mass in the International System of Units (SI) is the kilogram, which is abbreviated kg. The value of the kilogram

is defined as being “equal to the mass of the international prototype of the kilogram” (4). The kilogram is unique in the SI system of units in being the only unit whose definition is based on a physical artifact.

GRAVITY AND DENSITY

Gravity

Sir Isaac Newton described gravity in terms of the attracting force between two bodies, which is proportional to their masses and inversely proportional to the square of the distance between them. However, in considering the gravitational field of the earth and its effects on bodies that we are weighing, we need only consider the attraction of the body to the earth, which depends on the mass of the body and its distance from the center of the earth.

The earth is not quite spherical, having its greatest diameter at the equator. In addition, the earth is rotating on its axis, which creates a “centrifugal” force acting at 90° to the direction of spin and proportional to the speed of rotation. It therefore follows that the value of gravitational acceleration will be smaller at the equator than at the poles. The value of gravitational acceleration at any latitude, at sea level, may be calculated using the following formula:

$$g = g_e(1 + 0.005\,302\,4 \sin^2 \Phi - 0.000\,005\,8 \sin^2 2\Phi) \text{ m} \cdot \text{s}^{-2} \quad (1)$$

where

g = gravitational acceleration at desired location ($\text{m} \cdot \text{s}^{-2}$)

g_e = gravitational acceleration at the equator = $9.780\,327 \text{ m} \cdot \text{s}^{-2}$

Φ = latitude of location (deg)

The correction for altitude is $-3.088 \times 10^{-6} \text{ m} \cdot \text{s}^{-2}$ per meter above sea level. The above formulae give the best simple method of calculating the gravitational acceleration at a location where it has not been measured, and will almost always give results within $1 \times 10^{-3} \text{ m} \cdot \text{s}^{-2}$ and usually within $5 \times 10^{-4} \text{ m} \cdot \text{s}^{-2}$ (5).

The actual value of gravitational acceleration at a particular point will differ from that given by the equation above, due principally to the variability of the density of the earth from place to place. If an accurate determination at a particular location is required, a measurement can be made using appropriate apparatus. Also, in many countries the value of gravitational acceleration has been mapped for the entire country (by the British Geological Survey for the British Isles), and a value can be obtained on request.

Density

Density is the mass per unit volume of a substance, expressed in kilograms per cubic meter ($\text{kg} \cdot \text{m}^{-3}$). All substances, whether solid, liquid, or gas, have density, and an accurate knowledge of the density of the body being weighed and the density of the air at the time of weighing is essential to the accurate measurement of mass.

The density of a solid object can be determined by weighing it suspended in a liquid of known density and noting the apparent

loss in weight by comparison with its weight in air. The density of a liquid or gas can most easily be measured by reference to the apparent loss in weight of an artifact of known mass and volume immersed in it (often the change in apparent mass difference between two bodies of similar mass but different volume is used to give a smaller uncertainty of measurement).

The density of air can be measured as for other gases, but is commonly derived from a knowledge of the composition of the air and measurements of physical parameters. Air is composed of oxygen, nitrogen, water vapor, and a host of minor constituent gases, the most abundant of which are argon and carbon dioxide. A formula for the calculation of air density has been recommended by the CIPM and is internationally recognized (6).

Use of this formula requires measurement of temperature, pressure, and humidity, and for the highest accuracy of carbon dioxide content. The uncertainty of the calculated value of air density will depend on the uncertainty of measurement of the contributing parameters and on the uncertainty of the formula itself.

WEIGHING IN AIR—THE BUOYANCY EFFECT AND THE CONVENTIONAL VALUE

Archimedes' Principle and the Buoyancy Effect

The *weight* of a body is equal to the attractive force between it and the earth. We could measure this force and then divide it by the gravitational acceleration to determine the body's mass. We can measure the force by putting the body on the pan of a balance. But there is a complication.

Many centuries ago, the Greek philosopher Archimedes realized (while sitting in his bath, it is said) that when a body is immersed in a fluid, the fluid exerts an upward force on it that is equal to the weight of the fluid displaced by the body. When the upward force from the fluid is greater than the downward force due to gravity, the body will float on the surface of the fluid; otherwise it will sink, but part of its weight will be supported by the fluid. This discovery has since become known as Archimedes' principle. The magnitude of the upward force experienced by the immersed body is proportional to its volume and to the density of the fluid. We call this upward force the buoyancy force, or simply buoyancy.

When we think of fluids, we usually imagine liquids. But gases are also fluids and behave in the same way. We live immersed in a gas, which we call air. Air has a density, which, although only about one-thousandth of that of water, causes a significant buoyancy on most materials, the effect being greater on materials of lesser density. For any body in the earth's atmosphere, the force exerted on its support (= mass $\times$ gravitational acceleration) is reduced by the upward buoyancy force [= volume $\times$ density of air $\times$ gravitational acceleration, or (mass $\times$ density of air/density of body) $\times$ gravitational acceleration].

The result of this buoyancy effect is that a given mass of a more dense material exerts a greater force on its support than the same mass of a less dense material. Thus the two identical masses have different "weights." Hence one may deduce that to use the principle of weighing to determine mass, one must have a knowledge of the density of the materials being weighed and of the density of the air at the time of weighing. But it is unusual to have any accurate knowledge of the density of

most items that are to be weighed, and the measurement of air density requires sophisticated instrumentation or calculation.

Conventional Value of Mass

To solve this problem a *conventional value of mass* has been defined. In doing so it is supposed that all items have the same density, and the air density in which the weighing is made is also assumed to have a fixed value. The conventional value of mass will not always be the same as the true mass, but for most practical purposes it will be close enough. The chosen density values on which the conventional mass value is based are 8000 kg·m⁻³ for the item's density (because this is close to the density of stainless steel, iron, and brass, materials that are commonly used to manufacture weights) and 1.2 kg·m⁻³ for the air density (because this is typical of the air density at sea level in most parts of the world). Put simply, "For a weight taken at 20°C the conventional value of mass is the mass of a hypothetical reference standard weight of density 8 000 kg m⁻³ which it balances in air of density 1.2 kg m⁻³" (7).

For many practical purposes, weighings whose result is stated in conventional-value terms are quite satisfactory, although the true mass of the object being weighed may differ from the reported conventional mass. In particular, for the purposes of trade, commodities may be bought and sold using conventional mass with no net gain to the manufacturer, wholesaler, or buyer, provided the convention is consistently applied.

It is sometimes necessary to make a correction to the measured conventional value of mass, to allow for the difference in the density of the object being weighed or the difference in the density of the air at the time of weighing from the conventionally agreed values. Then we have

$$M_c = M_i \left[1 + \left(\frac{1}{\rho_w} - \frac{1}{8000} \right) (\rho_a - 1.2) \right] \quad (2)$$

where

M_c = corrected conventional mass value (kg)

M_i = indicated conventional mass value (kg)

ρ_w = density of artefact being weighed (kg·m⁻³)

ρ_a = density of air at the time of weighing (kg·m⁻³)

From Eq. (2) it can be seen that the correction is zero when either the density of the object or the air density at the time of weighing is equal to the conventionally agreed value. Knowing the normal range of air density (from meteorological data) and the range of density of materials to be weighed, it is possible to determine the limit of uncertainty for which there is no need to apply the correction.

It is also possible to deduce the true mass of an object from its conventional mass provided the density of the object is known:

$$M = M_c \left[1 + 1.2 \left(\frac{1}{\rho_w} - \frac{1}{8000} \right) \right] \quad (3)$$

where

M = true mass of object (kg)

M_c = conventional value of mass (kg)

ρ_w = density of object (kg·m⁻³)

From Eq. (3) it can be verified that the conventional mass value of an object whose density is $8000 \text{ kg}\cdot\text{m}^{-3}$ is identical to its true mass.

WEIGHING METHODS AND EQUIPMENT

Weighing is performed for a variety of purposes, ranging from retail trade activities such as selling food to the public, through food preparation and packaging, chemical analysis, and wholesale trading to industrial product preparation and process control. For simple trade and analytical weighing the requirement is usually to have a suitable balance or weighing machine available for a variety of weighing tasks, while for industrial applications the weighing system is more likely to be a purpose-built, permanently installed part of the processing plant.

In many cases the design of balances and simple weighing systems does not differ greatly from the design of process weighing plants, but from the user's viewpoint the purchase of a balance is a very different issue from the specification of an industrial processing plant. A balance can be purchased directly, and is usually selected on the basis of a manufacturer's specification, whereas the design of processing plant requires specialist engineering skills.

Force measurement systems and industrial weighing applications are discussed later. Let us first consider the use of a balance for simple weighing activities.

Balances and Weights

A balance may be based on mechanical principles, such as the position of a loaded beam or the extension of a spring element, or on electrical principles linked to the measurement of force, or on a combination of both. Types of balances are discussed in detail elsewhere in this encyclopedia (see **BALANCES**), so here we shall concentrate on the method of use of a balance, rather than on the detail of its design and operation.

A mechanical balance may involve a balanced beam offering two weighing pans, one on either side of the balance point, or it may be of single-pan design. An electronic balance will usually be a single-pan device. On a two-pan balance it is possible to measure the difference between the masses of two bodies directly, while a single-pan device can be used to give an indication of the mass of a single body. The single-pan device may also be used to measure mass difference by sequentially weighing one object and then the other. Typically a two-pan balance is used to compare the mass of an unknown weight with that of a calibrated reference weight, whereas with the single-pan device the reference weight is used to calibrate the balance, after which the unknown weight can itself be calibrated. The majority of balances currently available for purchase are single-pan electronic devices, so these will be examined in depth. But a large number of two-pan balances are still in use, so these are also discussed below.

Two-Pan Balances. Two-pan balances may be *damped* so that the swing of the beam quickly dies away, but more commonly they are *free-swinging*, so that the beam continues to swing for a period of time before finally coming to rest. On a damped balance a reading can be obtained quickly but the sensitivity to very small changes of load may be much reduced. The free-swinging balance may be allowed to come to rest before a reading is taken, but that is very time-consuming. An alternative

approach is to record the extremes of the beam swing in each direction and predict the final resting point from these. With care this method can produce satisfactory results with a significant saving of time. The same number of swings must be recorded for each weighing. Typically the first swing in either direction is ignored to allow the oscillating system time to establish a free oscillation; then three swings to one extreme interspersed with two to the other extreme are recorded. Analysis of the resulting data predicts the eventual rest point of the beam.

The most common method of use for a two-pan balance is to place a known reference weight in one pan and the unknown weight in the other pan. If the two weights are similar, the balance will indicate the mass difference, usually by means of a mechanical pointer or an optical scale. The balance must be calibrated so that the user can interpret the reading of the scale in terms of mass difference, and the calibration must be repeated regularly because the mechanical characteristics of the balance may change with time. Calibration of a two-pan balance is most easily achieved by observing the change in reading when a small, known mass is added first to one pan and then to the other.

If the lengths of the arms of the balance beam differ, the measurement result will be biased to one pan or the other. On a perfect balance, if the loads in the two pans are exchanged, the balance indication should remain the same except for a change of sign. But it is rarely the case that a balance is so perfectly set up, so it is common to take two readings, exchanging the weights between the pans before the second reading. The resulting mass difference is the mean of that obtained from the two readings. However, this method places some reliance on the stability of the balance during the time of the weighings, and two-pan mechanical balances are subject to a number of time-dependent effects. Therefore, for the smallest uncertainty of measurement the second weighing can be made a second time, before the weights are returned to their original pans and a fourth weighing is made. This *double-double weighing* method offers the best measurement uncertainty for a balance of this type.

Single-Pan Balances. Single-pan balances may be purely mechanical in operation, or involve electric force measurement. They may have a pan suspended below the weighing system, in which case access to the pan is hindered by the suspension mechanism, or be of "top-pan" design where the loading receptacle is freely accessible from above. Regardless of the operating principle, the method of use is similar.

The balance will be equipped with an indicator system to display the measured value. This may be as simple as a pointer attached to a spring or lever, observed against a printed scale, or as complex as an electronic digital display. The display mechanism will probably indicate in a unit of mass; nevertheless, all such systems require calibration to ascertain the difference between the displayed value and the true mass of the weight in the scale pan. Calibration can be performed using a suitably calibrated external weight, or may be dependent on the use of one or more weights built in to the weighing system (the "auto-cal" feature of many modern electronic balances). In either case it is necessary to know the uncertainty of the calibration weight, and where internal weights are used it is prudent to calibrate the device, or at least make occasional checks, using an external reference weight.

It is important to realize that, once calibrated, a single-pan balance is making measurements of force, and the indicated value will be dependent on the buoyancy effect, but more importantly on the value of gravitational acceleration. For this reason the balance should be calibrated in the location in which it will be used. (One major distribution company, selling balances manufactured and calibrated high in the mountains of China, received complaints of inaccuracy from a purchaser in a sea-side town in northern England.) The use of an internal reference weight to reset the *span* (the reading corresponding to full load) helps to overcome these problems, but if the internal weight is used to set the span of the balance, it cannot also be used to provide a calibration. Transportation of a balance is likely to affect its performance, hence adding further impetus to the requirement to calibrate in the location of use.

Other factors affecting the performance of single-pan balances are *hysteresis* (a difference in reading for the same load approached from a smaller load and from a larger load) and *eccentricity* (a difference in reading arising from placing the load toward one side of the pan, rather than centrally). These issues can be investigated during the calibration of the balance, while at the same time the sensitivity, linearity, and repeatability of loading are assessed.

The measurement uncertainty obtained using a calibrated balance to make measurements of unknown loads can often be improved significantly by using the same balance, together with an appropriate set of calibrated reference weights, to make comparison measurements. If done properly this approach can eliminate the uncertainties arising from nonlinearity, hysteresis, and long-term drift of the measuring system, and careful loading will minimize eccentricity effects. A measurement of the unknown weight is followed immediately by measurement of one or a group of reference weights of similar value. The closer the value of the reference weight(s) to the unknown, the less reliance is placed on the linearity of the balance. Repeated measurements allow assessment of repeatability, and comparison with a second reference weight adds confidence in the measurement. This method of substitution weighing is often referred to as *Borda's method*.

Weights. Weights, or reference mass standards, are available in a range of accuracy classes. The most widely recognized international classification of weights is that recommended by the International Organization of Legal Metrology in International Recommendation 111, "Weights of Classes E₁, E₂, F₁, F₂, M₁, M₂, and M₃," (8). Class E₁ weights are of the very highest metrological quality, typically used only by calibration laboratories of the highest order. These weights are made of non-magnetic stainless steel, and handled with extreme care, using tongs or gloved hands to avoid contamination. Class E₂, and F₁ weights, also made of nonmagnetic stainless steel and handled only with gloved hands, are commonly used as transfer standards to calibrate weights of lesser classes. Class F₂ weights are typically used as working standards in an industrial laboratory, while class M weights, commonly made from cast iron, are used as working standards in the field.

Weights of any class require periodic recalibration to ensure that they have not become damaged or contaminated. Weights should be handled with an appropriate amount of care—greater for the higher-class weights. Weights of class F₂ and below will often be recorded as being within a given tolerance of their nominal value, and this must be taken into account when assessing

the uncertainty of measurement of any weighings made with reference to them. Weights of the higher classes will usually be accompanied by a certificate stating their measured conventional values with associated uncertainties. The certified values should be used in all calculations associated with these weights.

Industrial and Process Weighing

For routine weighing of a variety of objects it is a simple task to purchase a balance and a set of weights from a supplier. But for weighings associated with industrial processes it is necessary to design the weighing apparatus as an integral part of the process equipment, which requires expertise both in the functional issues of the particular process, and in the selection, installation, use, and maintenance of appropriate force-measuring devices and systems. The remainder of this article is devoted to force measurement systems and transducers and their application to industrial and process control weighing.

FORCE

Force is a measure of the interaction between bodies. Force takes a number of forms, including short-range atomic forces, electromagnetic, and gravitational forces. Force is a vector quantity, with both direction and magnitude. If the forces acting on a body in equilibrium are summed around the periphery of the body, then they add to zero. If there is any resultant force acting, then the body is not in equilibrium and it will accelerate so that the rate of change of the body's momentum is equal to the force. If the body is held stationary in some way, then there will be a reaction acting on the body from the support structure that is equal in magnitude and opposite in direction to the force imposed. Although the definition of force units is based on acceleration of a free body, most force measurements are made on bodies in equilibrium, and are therefore measures of forces within a structure. The basis of most force measurements is that a physical support or link in a structure is replaced with a device that measures the forces acting at that point.

The SI unit of force is the newton (N); this is defined as the force that would give to a mass of one kilogram an acceleration of one meter per second per second. In practice, it is not convenient to accelerate 1 kg at 1 m·s⁻² to generate 1 N. Instead, the practical realization of the unit of force makes use of known masses that, when subjected to the effect of local gravitational acceleration, exert a known force on an earth-located support. The mechanical structure to handle and control such masses is known as a deadweight machine.

FORCE MEASUREMENT SYSTEMS

A force measurement system is made up of a transducer and associated instrumentation. The transducer is subjected to the force to be measured, and some resultant change in the sensing element of the transducer is measured by the associated instrumentation. The instrumentation may power the transducer in some way and may also process the output from the transducer before it is shown on an indicator. For many types of force measurement system, the term "load cell" is in common usage in place of "force transducer."

The instrumentation may be as simple as a dial gauge or as complex as a computer with associated analog-to-digital converters and excitation circuitry. The indicated value is the output of the force measurement system, which may be in units of force or other units such as volts. If the indicated value is not in units of force, then the user may need to perform a calculation based on a calibration to calculate the force value.

For the measurement made using a force device to be of any use, the uncertainty of that measurement must be known. Performance criteria for a force measurement system are described by the manufacturer in terms of a specification, which gives limits within which the behavior of the instrument can be expected to fall. However, this performance is generally given for a production series, and cannot be relied upon for a specific device unless verified by calibration.

Characteristics of Force Measurement Systems

Many physical principles are involved in the operation of different force measurement systems. The performance of these systems can, however, be described by a number of common characteristics, and the behavior of the system may be expressed graphically by plotting the indicated output value against the force applied to generate that value. It is then usual to fit a straight line to the resulting curve; deviation from this line is referred to as *nonlinearity*, and generally the largest such deviation is the value given for nonlinearity in the system specification. The difference of readings between the increasing and decreasing forces at any given force is called the *hysteresis*. The largest value of hysteresis is usually at the midrange of the system. Sometimes it may be useful to combine the nonlinearity and the hysteresis in a single figure. This is usually done by drawing two lines parallel to the best-fit line such that these lines enclose the increasing- and decreasing-force curves. The difference of output between the lines is halved and stated as combined error.

Any difference between the indicated value of force and the true value is an *error* of measurement, and may be expressed as percentage of actual force applied or percentage of full-scale output. *Full-scale output*, also known as *span* or *rated output*, is the output at rated capacity minus the output at zero applied force. *Sensitivity* is defined as the full-scale output divided by the rated capacity of a given load cell.

The ability of a force measurement system to measure force consistently is covered by the concept of *repeatability*. Repeatability is defined as the measure of agreement between the results of successive outputs of a force measurement system for repeated applications of a given force. The tests should be made by the same observer, with the same measuring equipment, on the same occasion, without mechanical or electrical disturbance, and calibration conditions such as temperature, alignment of loading, and the timing of readings held constant as far as possible. Although many manufacturers quote a value for repeatability as a basic characteristic of a transducer, it should not be considered as such. The value obtained for a transducer in a given machine will depend not only on the characteristics of the device itself, but also on temperature gradients, the resolution and repeatability of the electric measuring equipment, and the degree to which the conditions of the tests are held constant, all of which are characteristics of the test procedure.

In contrast to repeatability, *reproducibility* is defined as the closeness of agreement between the results of measurements

of the same force carried out under changed conditions of measurement. Note that a valid statement of reproducibility requires specification of the conditions changed and may include changes in the principle of measurement, method of measurement, observer, measuring instrument, reference standard, location, conditions of use, and time.

A force measurement system will take some time to adjust fully to a change in force applied, and *creep* of a force transducer is usually defined as the change of output with time following a step increase in force from one value to another. Most manufacturers specify the creep as the maximum change of output over a specified time after increasing the force from zero to the rated force. *Creep recovery* is the change of output following a step decrease in the force applied to the force transducer, usually from the rated force to zero. For both creep and creep recovery, the results will depend on how long the force applied has been at zero or the rated value respectively before the change of force is made.

The frequency response of a force transducer is affected by the nature of the mechanical structure, both of the transducer itself and of its mounting. A force transducer on a rigid foundation will have a natural frequency of oscillation, and large dynamic errors occur when the frequency of the vibration approaches the natural frequency of oscillation of the system. If dynamic forces (ones that vary with time) are to be measured, the frequency response of the force measurement system must be capable of following the changing force. If a transducer is to be used with fluctuating forces applied, its fatigue life should also be considered.

The temperature coefficients of both the output at zero force and the sensitivity quantify the effect of temperature changes on a given system. A force measurement system may need to be kept at constant temperature, or set up well in advance to settle into the ambient conditions, if high-accuracy measurements are required. In some cases the temperature gradients within the measurement installation can create a problem even when the average temperature is stable.

Other influence quantities such as humidity, pressure, electrical power changes, or radio-frequency interference may have analogous effects to those of temperature and may be considered in a similar manner.

In general, a force transducer has two interfaces through which a force is applied. These may be the upper and lower loading surfaces of a compression force transducer or the upper and lower screw threads of a tension device. At each interface, there will be a force distribution, which will depend on the end loading conditions. A change in these loading conditions may therefore cause a change in the force distribution, resulting in a change of the sensitivity of the transducer, even though the resultant force at the interface remains unchanged. Depending on the design of the transducer, the change of sensitivity caused by a change of end loading conditions can be significant in comparison with all other effects.

The long-term stability of the sensitivity of force measurement systems is clearly important if they are to be used to compare the magnitude of forces at different times, perhaps months or years apart. This stability will be determined by several factors, including the stability of the force transducer's many components, the protection of the sensing components or other parts against humidity, and the conditions under which the system is stored, transported, and used.

FORCE TRANSDUCERS

There are many types of force transducer, and these may be used with instrumentation of varying complexity. Most force transducers employ some form of elastic load-bearing element or combination of elements. Application of force to the elastic element causes it to deflect, and the deflection is then sensed by a secondary transducer, which converts it into an output. The output may be in the form of an electric signal, as in strain-gauge and linear variable differential transformer (LVDT) load cells, or of mechanical indications, as in proving rings and spring balances.

The method of measuring the distortion of the elastic element varies considerably. The most often used method is to make measurements of the surface strain; when this is undertaken with electrical resistance strain gauges, such a transducer is known as a strain-gauge load cell. Strain-gauge load cells are the most common commercially available type of force transducer.

Another commonly used type of force transducer is based on the piezoelectric phenomenon exhibited by certain crystalline materials, where an electric charge is generated on the crystal surface, the amount of charge being proportional to the applied force. Many other physical relationships are also used, and force transducers based on these are described later.

Strain-Gauge Load Cells

Each such load cell contains an elastic element to which a number of electrical resistance strain gauges are bonded. The geometry and modulus of elasticity of the element determine the magnitude of the strain field produced by the action of the force. Each strain gauge responds to the local strain at its location, and the measurement is determined from a combination of these individual strain values.

The rated capacities of strain-gauge load cells range from 1 N to more than 50 MN, and they can be used with high-resolution digital indicators as force transfer standards. The shape of the elastic element used in load cells depends on a number of factors, including the range of force to be measured, dimensional limits, final performance, and production costs.

Figure 1 shows a selection of different elastic elements and gives their typical rated capacities. Each element is designed to measure the forces acting along its principal axis, and to be unaffected by side forces. The arrows in the figure indicate the principal axis of each element.

The material used for the elastic element is usually tool steel, stainless steel, aluminum, or beryllium copper. The ideal material exhibits a linear relationship between stress and strain with low hysteresis and low creep in the working range. There also has to be high level of repeatability between force cycles to ensure that the load cell is a reliable measuring device.

In electrical terms, all electrical resistance strain gauges may be considered as a length of conducting material, such as a wire. When a length of wire is subjected to a tension within its elastic limit, its length increases, with corresponding decrease in its diameter and change of its electrical resistance. If the conducting material is bonded to an elastic element under strain, then the change in resistance may be measured and used to calculate the force from the calibration of the device.

The foil strain gauge is the most widely used type. It has significant advantages over all other types of strain gauge and

is employed in the majority of precision load cells. It consists of a metal foil pattern mounted on an insulating backing or carrier, which provides electrical insulation between the foil and the elastic element, facilitates handling, and presents a readily bondable surface. A large variety of foil gauges are now commercially available to the transducer designer and general user.

Semiconductor strain gauges are manufactured from strips of semiconducting silicon of either the *n* or the *p* type. The output sensitivity of a semiconductor gauge is very high compared to a foil gauge. It is also, however, nonlinear in strain and highly temperature-sensitive. The gauges exhibit essentially no creep or hysteresis and have an extremely long fatigue life. This type of gauge is widely used on small transducers such as accelerometers and pressure sensors, whose sensing element may be micro-machined out of a single piece of silicon.

Thin-film strain gauges are produced by sputtering or evaporating thin films of metals or alloys onto the elastic element. A thin-film strain-gauge system may have up to eight layers of material. There are a number of such force transducers available, covering a range of 0.1 N to 100 N, in the form of a single- or double-bending beam configuration. These devices are highly cost-effective when produced in large quantities due to the manufacturing techniques involved. This makes them ideally suited for use in large-volume products such as shop scales and pressure transducers.

The wire strain gauge was the original type of resistance strain gauge, though now widely replaced by cheaper foil or thin film types. However, the wire strain gauge is still used extensively for high-temperature transducers and stress analysis, and is available in a wide range of materials.

The nominal resistance of a strain gauge varies with the type and application. Wire gauges have resistances in the range of 60 Ω to 350 Ω , foil and semiconductor gauges from 120 Ω to 5 k Ω , and thin-film types around 10 k Ω . Selection criteria may include size, self-heating, and power requirements. If several load cells are to be connected together, then matched resistance may be important.

The strain gauge converts the strain produced by the applied force into resistance change. To maximize the response of the load cell, it is normal to connect one or more strain gauges aligned to respond to a compressive strain and another set aligned with a tensile strain. When connected electrically as a Wheatstone bridge configuration, this has the added advantage of minimizing the effects of environmental changes, such as temperature, which act equally on all the gauges. The resistance change is detected by measuring the differential voltage across the bridge.

A Wheatstone bridge is normally formed by four strain gauges, although it is not uncommon to use two strain gauges for a *half bridge*, or more than four on elastic elements of complex shape. In the latter case, half the total number of strain gauges are subjected to compressive strains and the other half to tensile strains. The voltage output from the bridge, when excited by an input voltage, is linearly related to the resistance change of the strain gauges and is therefore a function of the force applied to the element. The rated output of a load cell is commonly standardized to a nominal level, usually 2 mV/V (2 mV output at rated load per 1 V input), but it can range from 1 mV/V to 5 mV/V.

A shunt resistor may be needed to perform a shunt calibration, and this is often fitted in the signal conditioning unit to be

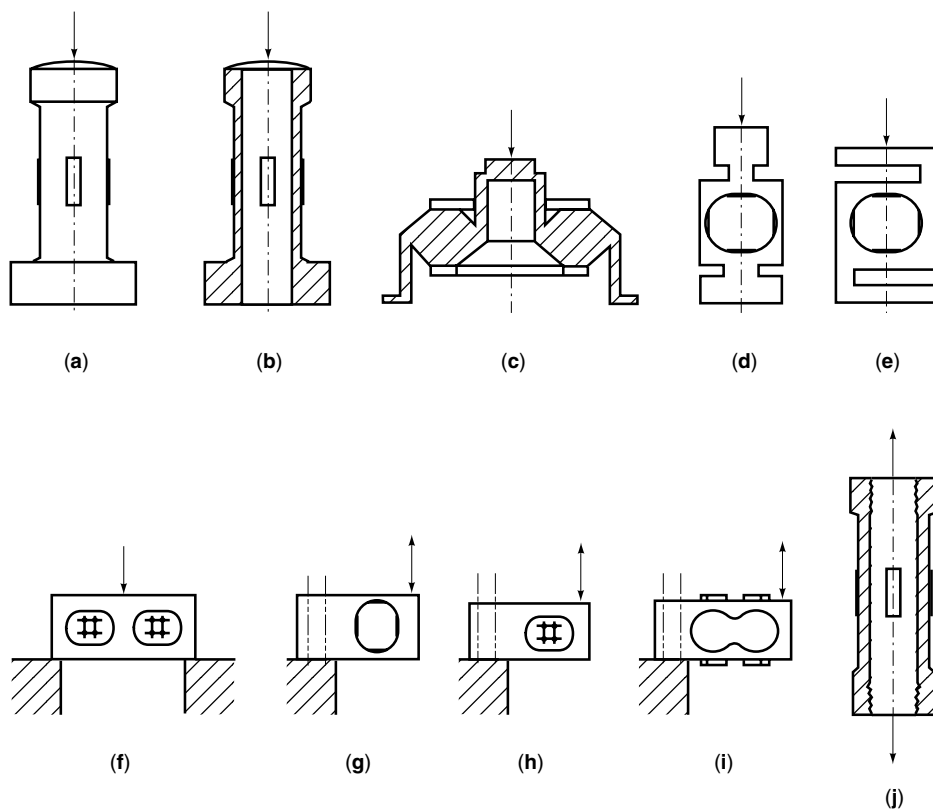


Figure 1. Various types of load cell: (a) compression cylinder, 50 kN to 50 MN; (b) compression cylinder (hollow), 10 kN to 50 MN; (c) toroidal ring, 1 kN to 5 MN; (d) ring, 1 kN to 1 MN; (e) S beam (bending or shear), 200 N to 50 kN; (f) double-ended shear beam, 20 kN to 2 MN; (g) double-bending beam (simplified), 500 N to 50 kN; (h) shear beam, 1 kN to 500 kN; (i) double-bending beam, 100 N to 10 kN; (j) tension cylinder, 50 kN to 50 MN.

switched across one arm of the bridge when required. Typical shunt calibration resistor values used for a 350 Ω bridge are 40 k Ω and 80 k Ω , which equate to approximately 90% and 45% of full load on a 2 mV/V transducer.

A load cell is part of a measurement chain, and it requires an excitation voltage to be supplied, as well as amplification and conditioning of the output signal, before it can be meaningfully displayed or used in a control system. Normally a system excites the load cell with direct current (dc) voltage and amplifies the output through an instrumentation amplifier. This chain features wide frequency bandwidth, high stability, and relatively low cost.

In industrial applications, the distance between the load cell and the measuring instrument may be considerable, possibly hundreds of meters. The voltage drop along the connecting cable and its dependence upon temperature can contribute to the system error. This additional error can be remedied by the use of a six-wire connection technique. The excitation voltage is measured at the load cell, rather than at the instrument, and maintained at a constant preset level.

An alternative to dc excitation is the alternating current (ac) system, which excites the load cell with an ac signal. The output is processed through an ac amplifier, a synchronous demodulator, a filter, and a dc amplifier. A high-level dc signal is obtained, suitable for direct analog display or for conversion to a digital display with the use of an analog-to-digital converter. Such a system offers higher immunity to thermal effects in the instrumentation and thermoelectric effects in the transducer, high noise rejection, good zero-force output stability, and ease of achieving isolation between the signal output and the load cell. However, in view of the complex measuring chain, these systems tend to be costly.

A strain-gauge load cell has two distinct temperature coefficients, one on zero-load output and one on full-scale output. The former is caused by the effect of temperature on the strain-gauge resistance, the mismatch between the expansion coefficient of the gauge and the spring element to which it is bonded, and the out-of-balance condition in the bridge wiring. The effect on full-scale output is mainly caused by the temperature dependence of the modulus of elasticity of the spring element material. Temperature compensation gauges can reduce the load cell's temperature coefficients to very low levels.

It is usually necessary to protect the elastic element, its strain gauges, and associated electronic components, since many load cells are exposed to harsh industrial environments. The way the load cell is housed depends on its intended application, and there are a wide range of shapes and sizes, including cylinders, rectangular and circular beams, and tension devices like turnbuckles. The housing must allow for suitable fixings such as bolts or cradles and is critical in determining the correct transfer of the force to the elastic element. The housing used for this purpose also fulfils other functions, such as limiting the side forces on the element and protecting the various electric components.

Piezoelectric Crystal Force Transducers

When a force is exerted on certain crystalline materials, electric charges are formed on the crystal surface in proportion to that force. To make use of the device, a charge amplifier is required to integrate the electric charges to give a signal that is proportional to the applied force and big enough to measure. These devices are known as piezoelectric crystal, or quartz, force transducers.

These sensors are different from most other sensing techniques in that they are active sensing elements. No power supply is needed, and the deformation to generate a signal is very small. This provides the advantage of high-frequency response of the measuring system without introducing geometric changes into the force measuring path.

When packaged as a load washer and compressed under a force of 10 kN, a typical piezoelectric transducer deflects only 0.001 mm. The high-frequency response (up to 100 kHz) enabled by this stiffness and the other inherent qualities of the piezoelectric effect makes piezoelectric crystal sensors very suitable for dynamic measurements.

Extremely fast events such as shock waves in solids, or impact printer and punch press forces, can be measured with these devices. Piezoelectric sensors operate with small electric charge and require high impedance cable for the electrical interface. It is important to use the matched cabling supplied with a transducer.

Piezoelectric crystal sensors are primarily designed for applications using a pretensioned bolt that allows the measurement of forces in both tension and compression. The preloading is important to ensure optimum linearity, and the sensor must be calibrated after mounting. An extension of this principle is the use of force-measuring pins, which are placed within the structure of a machine and respond to the forces within the structure.

There is a small leakage of charge inherent in the charge amplifier, which is called *drift* of the signal. So while piezoelectric force transducers are ideally suited for dynamic measurements, they cannot perform truly static measurements.

Piezoelectric crystal sensors are suitable for measurements in laboratories as well as in industrial settings. The measuring range is very wide, and the transducers survive high overload (typically > 100% of full-scale output). The sensors' small dimensions, large measuring range, and rugged packaging make them very easy to use. They can operate over a wide temperature range and survive temperatures of up to 350°C.

Hydraulic Load Cells

The hydraulic load cell is a device filled with a liquid (usually oil), which has a preload pressure. Application of the force to the loading member increases the fluid pressure, which is measured by a pressure transducer or displayed on a pressure gauge dial via a Bourdon tube.

When used with a pressure transducer, hydraulic load cells are inherently very stiff, deflecting only about 0.05 mm under full force conditions. Although capacities of up to 5 MN are available, most devices fall in to the range of 500 N to 200 kN. The pressure gauge used to monitor the force can be located several meters away from the device by the use of a special fluid-filled hose.

Hydraulic load cells are self-contained and need no external power. They are inherently suitable for use in potentially explosive atmospheres and can be tension or compression devices, but are sensitive to temperature changes.

Pneumatic Load Cells

The operating principles of the pneumatic load cell are similar to those of the hydraulic load cell. The force is applied to one side of a piston or a diaphragm of flexible material and balanced by pneumatic pressure on the other side. This counteracting

pressure is proportional to the force and is displayed on a pressure gauge.

Other Elastic Devices

The *loading column* is probably the simplest elastic device, being simply a metal cylinder subjected to a force along its axis. The change in length of the cylinder is measured directly by a dial gauge and lever system, or other technique. The *proving ring* is functionally very similar except that the element is a circular ring, and the deformation is usually measured across the inside diameter. These transducers have the advantage of being simple and robust; their main disadvantage is the strong effect of temperature on the output.

An LVDT may be used within a load cell to measure the displacement of an elastic element instead of using strain gauges. The LVDT is essentially a transformer that provides an ac output voltage as a function of the displacement of a separate movable magnetic core. The lack of friction and the low mass of the core result in high resolution and low hysteresis, making this device ideal for dynamic measurement applications.

Capacitive load cells use a capacitance sensor to sense the displacement of an elastic element. In most cases the sensor consists of two parallel plates standing opposite each other. The changing length of a spring member produces a change in the gap between two plates, and hence a change in electric capacitance. In the case of small weighing instruments, such as domestic scales, the spring also provides parallel guidance of the scale's platform.

An optical strain gauge can be formed in a manner similar to a wire strain gauge by the use of optical fibers. The deflection of the elastic force-bearing member with the optical strain gauge bonded to it will result in length changes in the optical fibers. If monochromatic light is used to feed two optical strain gauges experiencing different strains, then the phase difference between the two beams emerging from the gauges is a function of the applied force.

The interference optical load cell uses a high-resolution displacement measuring method. A fork-shaped spring is deformed by the force, the deformation being in the region of 40 μm , and the change of the aperture of the fork is measured by a Michelson interferometer. For the same resolution, the maximum elastic deformation, and with it the strain of the material, need not be as large as in the case of the strain-gauge load cell. The deformation element is made of quartz, which has a very low temperature dependence. The hysteresis and creep of these systems are both particularly small.

Vibrating Elements

In the case of the tuning-fork load cell, the force transducer consists of two parallel band splines, which are connected at their ends and vibrate in opposite directions in resonance. The mode of vibration is like that of a tuning fork and the resonance frequency changes if the element is subjected to a tensile or compressive force. The excitation of the vibration and the reciprocal reception of the vibration signals are carried out by two piezoelectric elements close to the vibration node of the tuning fork.

The vibrating-wire transducer consists of a taut ferromagnetic wire that is excited into transverse vibrations by a drive coil. These vibrations are detected using a pickup coil. Both coils have permanent magnet cores, and once the wire has been

excited to its resonance frequency for a given tension, it is maintained at this frequency by connecting the two coils through an amplifier to form a self-oscillating system. Each resonance frequency is a measure of the wire's tension and, hence, the applied force at that instant. The advantage of the vibrating-wire transducer is its direct frequency output, which can be handled by digital circuitry, eliminating the need for an analog-to-digital converter.

Magnetoelastic Devices

The magnetoelastic force transducer is based on the effect that when a ferromagnetic material is subjected to mechanical stress, the magnetic properties of the material are altered and the change is proportional to the applied stress. Due to its sturdy construction, high signal level, and small internal resistance, the magnetoelastic load cell can be used in rough and electrically disturbed environments such as rolling mills. The rated capacities of these devices are in the range from 2 N to 5 MN.

Dynamic Balance Devices

Gyroscopic load cells exploit the force sensitivity of a gyroscope mounted in a gimbal or frame system. A commercially available gyroscopic load cell incorporates a dynamically balanced heavy rotor on a spindle, itself mounted in the inner frame of a two-gimbal system. The arrangement has three orthogonal axes of rotational freedom and has the axis of origin at the center of gravity of the rotor. The force to be measured by the transducer is applied through the lower swivel, and a couple is produced on the inner frame, causing the gimbals to precess. The time taken for the outer gimbal to complete one revolution is then a measure of the applied force. The gyroscopic load cell is essentially a fast-responding digital transducer and is inherently free of hysteresis and drift.

The force balance uses a feedback circuit to compare the electrical output with the force input. A typical system has attached to the force input member an electric coil, which operates in the flux gap of a permanent magnet. An electric current passed through the coil generates a restoring force in opposition to the applied force. A displacement transducer is used to sense the displacement of the force input member, its output being amplified and used to control the current in the coil until the restoring force exactly balances the applied force and restores the force input member to its original position. The coil current to achieve this balance is proportional to the applied force and is measured as a voltage sensed across a resistor in series with the coil. These types of device have good dynamic performance, small deflection, and relative insensitivity to environmental conditions. They are inherently stable and accurate, and as a result are often considered for use as secondary standards. This type of device is mainly a competitor for the mechanical analytical balance in mass determination.

APPLICATIONS

There are many different applications of force measurement systems in industry, too numerous to describe here. One of the main applications is the weighing of products during manufacturing—this is known as process weighing. Batch mixing, truck loading, packing, and vessel filling of liquids, gases,

or solids are all processes requiring accurate determination of weight. Many process weighing situations use load cells to measure forces, from which the weight of product is derived. One pitfall is that, while determination of the force may be straightforward, the interpretation of the gravitational component acting on the mass may not be so simple, as described in the following example of vessel weighing.

If a vessel can be mounted at all support points on force transducers, and if the associated pipework and fittings do not take some of the force, the weight may be derived from the total force measured. Often, however, to reduce cost, the vessel may be mounted on one load cell and two dummies, or two load cells and two dummies, and the force is assumed to be equally distributed between the active and nonactive points.

Weighing of a three-legged tank/vessel with one active force transducer in this way leaves the measurement at the mercy of variations in center of gravity due to nonlevel loads, nonuniform vessel cross section, agitation of the contents, and external wind forces. Accuracy of weighing with these systems is usually poor (although acceptable sometimes for level measurement) even if the material in the tank is self-leveling or a liquid. The effects of pipework, fittings, and nonlinear deformation of the tank will cause movement of the center of gravity between the empty and full conditions, as well as affecting the total force seen by the measurement devices (9).

System Design

The choice of force measurement system can only be made after considering the specific details of the application. The designer should consider the range of force to be measured, the number of loading points, the direction of the forces (and whether in tension, compression, or both), and the duration and rate at which the force is applied to the transducer. As a result of these considerations, the choice can be made on what type of transducers can be used, how many will be required and of what capacity, whether uni- or bidirectional, and whether single or multiaxis.

The choice of instrumentation also needs to be made. The frequency and number of data to be collected need to be considered, as well as the nature of the indicator and any links to a data acquisition or control system. The environment in which the force measurement system has to work can have a significant effect on its performance. Parameters to be considered include the temperature, vibration, humidity, corrosive nature of the atmosphere, variation in ambient pressure, and freedom of movement required by the displacement of all the force measurement transducers.

The design should also ensure that there are no parallel load paths (known as force shunts) that may take some of the force that is supposed to be measured by the transducer. Force shunts can be present all of the time (as through pipework connections) or may occur when the movement of the system is arrested by overload stops. The end fittings of the transducer should be designed both to minimize side loads and to allow the force to be passed to the transducer along a well-defined axis, normally the principal axis of the transducer.

Calibration and Traceability

Even with good transducers and a good system design, the measurement cannot be relied upon without some check on the performance of the system. Unexpected installation effects, degradation of the equipment over time, or user abuse may all

lead to a measurement that has a greater uncertainty than the designer planned. Consequently, calibration is required to ensure that the force measurement meets the needs of the user and achieves the required degree of uncertainty.

Traceability to authoritative standards is the best way to ensure accuracy of measurements, and leads to consistency of measurements among users at different times and locations. A standard force is less easy to visualize than a standard mass and, by its nature, must be created by a machine rather than having a separate existence as an artifact. Not only must the magnitude of the force be known to the required uncertainty, but also its direction and the conditions of its application to the transducer under test (10,11).

Machines capable of undertaking force calibrations are known as force standard machines, and they may be categorized as either primary or secondary. Primary standards in force measurement are generally deadweight machines whose uncertainty can be directly verified through physical principles with respect to the fundamental base units of mass, length, and time. Secondary standards are machines that can reliably reproduce forces and can be compared with primary standards by the use of a force transfer standard. Examples of secondary standards are lever, reference load cell, hydraulic amplification, and hydraulic strain-gauged column machines. The instrumentation used during the calibration must itself have full traceability, either by calibrating the complete force measuring system of transducer and instrumentation together, or by calibrating the instrumentation separately.

Primary standards are usually defined at a national or international level by a single authoritative measuring instrument known as a national standard or international standard. Traceability normally means that the calibrations have been undertaken by the national standards laboratory, by an accredited laboratory, or by a national or accredited laboratory in another country with which there is a reciprocal recognition agreement.

Most documentary standards in the area of force measurement were written to satisfy a need for traceability in materials testing. The procedures contained in these standards are also used for the calibration of force measurement systems in a wide range of other industries. The calibration, or verification, of materials testing machines is covered by ISO 7500 (12) and other national standards. The document dealing with the calibration and classification of transfer standards (or force-proving instruments) is ISO 376 (13).

The above standards have been developed for the calibration of systems to measure static forces acting along a single well-defined axis. Calibrations are also required for multiaxis situations and for systems that measure dynamic forces.

Multiaxis calibration is similar to single-axis calibration, but done once for each axis. The calibration of multicomponent force sensors is more demanding on the equipment, although the principles remain the same.

At the time of writing, dynamic force calibration is not yet an established procedure. The statically derived force transducer sensitivities are assumed to be applicable for dynamic force measurements. This can lead to significant errors, and work is ongoing to produce a dynamic calibration standard.

There are three main calibration options available for establishing the uncertainty of the force measurement system. The first is to leave the force transducer in its permanently installed position and use a transfer standard to carry out the calibrations. The second is to calibrate the force transducer prior to

installation and remove it as required for further calibrations. The final option is to calibrate the force transducer prior to its permanent installation in the force measurement system and then not to recalibrate during the life of the installation.

The user also needs to consider how often to calibrate the device, whether to calibrate the whole system or just the transducer, whether adjustment of the instrument is required, and what uncertainty level is required. The end-loading and temperature conditions during the calibration should be similar to those experienced in the application.

Before taking any calibration readings the transducer should be preloaded up to its rated force and back to zero several times. This is to ensure that any mechanical or electric connection, offsets, or mismatches have been allowed to settle in. When taking measurements, the applied force should always be approached from the same direction, and allowed to stabilize before a reading is taken. It is also important that zero force be a well-defined mechanical point at the same level and direction, taking account of backlash and the like.

Analysis of the calibration data will often be carried out by the calibration laboratory and a calibration certificate supplied to the user when the transducer is returned. There are other checks the user can make, such as checking that the worst-case error is within the process needs, checking for any anomalies or bias, checking that the linearity is within process needs, and reviewing the history of the device's calibrations.

QUALITY ASSURANCE OF FORCE AND MASS MEASUREMENTS

Many organizations are granted a recognized certification or accreditation for their activities, covering their overall quality management system. However, a generalized quality certification does not address the details of specialised technical practices. Accreditation, on the other hand, means that certain aspects of their business have been independently assessed, and that they comply with given criteria of competence and quality.

In the United Kingdom, for example, the United Kingdom Accreditation Service (UKAS) undertakes the National Accreditation of Measurement and Sampling (NAMAS). NAMAS accreditation of a calibration or testing laboratory provides assurance that measurements are carried out to the highest standards of technical competence, traceable to recognized national or international standards, using agreed methods, and with realistic statements of uncertainty. The NAMAS regulations are based on international standards for the operation and accreditation of laboratories, such as EN 45000 (14) and ISO Guide 25 (15). NAMAS certificates are widely recognised and accepted throughout the United Kingdom and also world-wide.

Counterparts of NAMAS exist in many other countries, and in many cases are recognized as equivalent. Formal agreements provide for the mutual recognition of certificates from different national accreditation schemes. The European Co-operation for Accreditation (EA) is the body that is the focus for multilateral recognition among national measurement accreditation schemes in Europe.

BIBLIOGRAPHY

1. *Guide to the Measurement of Mass and Weight*, London: Institute of Measurement and Control, 1998.

2. R. D. Connor, *The Weights and Measures of England*, London: Her Majesty's Stationery Office, 1987, (out of print).
3. M. Plassa, The international prototype kilogram, a successful standard, the history of its choice and realisation, *Basic Metrology and Applications*, Levrotto and Bella (eds.), Turin, 1994.
4. R. J. Bell (UK ed.), *SI, The International System of Units*, London: Her Majesty's Stationery Office, 1993.
5. G. W. C. Kaye and T. H. Laby, *Tables of Chemical and Physical Constants*, 16th ed., London: Longman, 1995.
6. R. S. Davis, Equation for the density of moist air, 1981/91, *Metrologia*, **29** (1): 67–70, 1992.
7. *International Recommendation No. 33, Conventional Value of the Result of a Weighing in Air*, Paris: International Organization of Legal Metrology, 1973.
8. *International Recommendation 111, Weights of Classes E_1 , E_2 , F_1 , F_2 , M_1 , M_2 and M_3* , Paris: International Organization of Legal Metrology, 1994.
9. *Guide to the Measurement of Force*, London: Institute of Measurement and Control, 1998.
10. *A Code of Practice for the Calibration of Industrial Process Weighing Systems*, London: Institute of Measurement and Control, 1996.
11. *A Procedure for Calibration and Testing of Strain Gauge Load Cells Used for Industrial Process and Force Measurement*, London: Institute of Measurement and Control, 1993.
12. ISO 7500 International Standard on Metallic Materials—Verification of Static Uniaxial Testing Machines, Geneva: International Organisation for Standardization, 1986.
13. ISO 376:1987(E) International Standard on Metallic Materials—Calibration of Force-Proving Instruments Used for the Verification of Uni-axial Testing Machines, Geneva: International Organisation for Standardization, 1987.
14. BS EN 45000 (BS7501:1989) British Standard—General Criteria for the Operation of Testing Laboratories, British Standards Institution, 1989.
15. ISO Guide 25, 1990, General Requirements for the Competence of Calibration and Testing Laboratories, Geneva: International Organisation for Standardization, 1990.

David R. Armitage
 Andy J. Knott
 National Physical Laboratory

GRAVIMETERS

Gravity measurements reflect the earth's gravitational attraction, its centrifugal force, tidal accelerations due to the sun, moon, and planets, and other applied forces. They are invaluable to oil and mineral prospectors. They are used by physical scientists to determine the exact size and shape of the earth and they contribute to the gravity compensations applied to inertial navigation systems. If the earth was a sphere of uniform density, then its gravitational attraction on small bodies located on its surface would be constant everywhere. In fact, the earth's centrifugal forces of rotation have flattened it at the poles, making its polar radius approximately 21 km less than its equatorial radius. The outward centrifugal acceleration at the equator (which is nonexistent at the poles) decreases inward equatorial gravity accelerations by approximately 3400 mgal (1 mgal = 10^{-5} m/s² $\cong$ 1 μ g or 1 part per 10^6). Equatorial gravity measurements also reflect a greater attraction to the whole earth (owing

to the lower latitude bulge), and this results in an increase of approximately 4900 mgal. Polar gravity measurements reflect being closer to the center of mass which accounts for an increased measurement of approximately 6600 mgal. Taken collectively, polar gravity is roughly 5100 mgal stronger than equatorial gravity. Gravity measurements also decrease by about 1 mgal for every 3 m increase in height above mean sea level. Local and regional subterranean density variations also produce gravity changes as large as 200 to 300 mgal, although these tend to be smaller.

As explained by Einstein, the Equivalence Principle does not permit gravity to be measured at a point. What is measured is the *specific* force impressed on the instrument. The *gravitational* force is inferred from knowledge (and assumptions) of the acceleration of the instrument. Accelerometers and gravimeters are very similar devices because they are sensitive to the same types of inputs. They are, however, optimized for different measurement regimes in terms of dynamic range, frequency response, and operating environments.

The first gravity measurement device was the pendulum clock, invented by Huygens in 1656 (Newton's 14th year). The pendulum's period (T) and the acceleration of gravity (g) are inversely related by $T = 2\pi(I_m/mgh)^{1/2}$ where I_m is the pendulum's moment of inertia about its pivot axis, m is its mass, and h is the distance between its center of mass and its pivot point. French astronomers soon noticed such clocks lost time at the equator when compared to Paris-based observations. This was the first direct evidence that gravity lessens as latitude lessens. Pendulum measurements are time consuming and require elaborate executions and corrections. Moreover, neither I_m nor h can be measured with great precision. A pendulum's mechanical properties also change with time and transport. These mechanical changes create changes in the pendulum's period that are difficult to calibrate. Owing to these problems, pendulums have been completely replaced by two classes of high-precision, high-accuracy gravity measurement devices: (1) absolute gravity apparatuses (both portable and stationary) which use lasers and atomic clocks to measure and time freely falling body distances and (2) relative gravity meters (or gravimeters), which measure the force required to rebalance the gravity force acting on a leveled proof mass attached to a spring *against* the force exerted by the spring, as the meter is moved from one measurement point to the next.

ABSOLUTE GRAVITY MEASUREMENTS

Free-fall Acceleration Measurements

Neglecting air resistance, if a freely falling body is a distance x_0 from an overhead origin at time t_0 and moving with velocity v_0 m/s, then subsequent x_i distances occur t_i seconds after t_0 with time and distance related by (assuming g is constant from x_0 to x_i)

$$x_i = x_0 + v_0 t_i + \frac{1}{2} g t_i^2 \quad (1)$$

Designating t_1 as a starting time, t_2 an intermediate time, and t_3 a final time; and taking combinations of $x_i - x_j$ differences

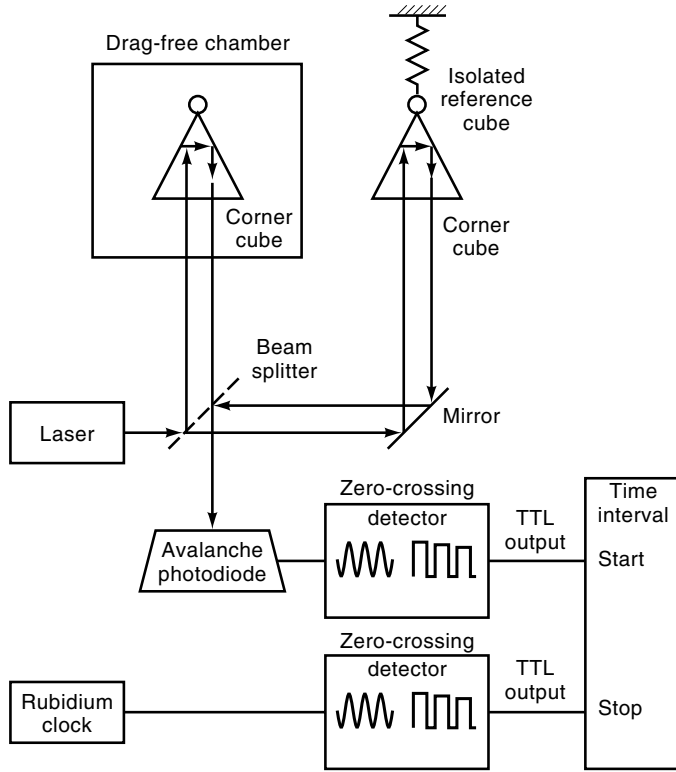


Figure 1. Block diagram of the Joint Institute for Laboratory Astrophysics free-fall method, patterned after Niebauer et al. (2).

yields

$$g = 2 \left[\frac{\Delta x_3 - (\Delta t_3 / \Delta t_2) \Delta x_2}{\Delta t_3^2 - \Delta t_2 \Delta t_3} \right] \quad (2)$$

where $\Delta x_3 = x_3 - x_1$, $\Delta x_2 = x_2 - x_1$, $\Delta t_3 = t_3 - t_1$, $\Delta t_2 = t_2 - t_1$, and it's noted Eq. (2) is free of x_0 and v_0 . Precise and accurate measurements of distance *and* time are required to accurately measure the absolute value of g in this manner. Today's most sensitive free-fall measuring instruments are portable and rely on stabilized lasers and atomic (rubidium) clocks to provide length and time standards. The drifts in these standards are low enough that they can be used for months without drift errors contributing at the parts per 10^9 level. The standards also are minimally affected by transit vibrations and environmental temperature changes (1). Since a mass dropped from rest falls approximately 5 m in 1 s, time and distance measurements precise to 1 part per 10^9 yield absolute gravity measurements with precisions of a few μgal ($1 \mu\text{gal} = 10^{-8} \text{ m/s}^2 \cong 10^{-9} g$).

Figure 1 is a simplified diagram of the free-fall, absolute gravity meter at the Joint Institute for Laboratory Astrophysics (JILA) (2). A stabilized laser illuminates a Michelson interferometer formed between one light beam, reflected from a falling corner cube (a mirror that reflects the laser directly back, regardless of the cube's orientation) and a second light beam reflected from a stationary reference corner cube. The free-falling cube resides in a drag-free vacuum chamber to eliminate air resistance. The interference of the two reflected beams makes moving fringes (light or dark bands) sensed by an avalanche

photodiode, which activates the timing devices. An electronic scaler, linked to the atomic clock, determines the time between whole numbers of fringes to within a nanosecond (10^{-9} s). Since the wavelength of the laser is also accurately known, the distance traveled by the free-falling cube is accurately determined by counting the number of fringes passing during the time intervals Δt_3 and Δt_2 in Eq. (2). Because each fringe covers a distance of one-half the laser's wavelength (λ), the actual determination of g is from a modified version of Eq. (2), namely

$$g = \lambda \left[\frac{f_3 - (\Delta t_3 / \Delta t_2) f_2}{\Delta t_3^2 - \Delta t_2 \Delta t_3} \right] \quad (3)$$

where f_3 and f_2 are the counted fringes during Δt_3 and Δt_2 . The most accurate and precise free-fall absolute gravity measurements taken to date are discussed in Niebauer et al. (1). Accuracies of $2 \mu\text{gal}$ and repeatabilities of $1 \mu\text{gal}$ are claimed using a portable device weighing approximately 300 kg and possessing a 3 m^2 footprint.

Engineering challenges related to free-fall devices include making the vacuum chamber as free of electrostatic and magnetic effects as possible and ensuring that the laser beam reflected to the falling cube is accurately vertical. The latter is addressed by placing a dish of mercury under the point where gravity is measured. The optics base is then positioned and adjusted to make a vertical beam using the mercury as a reference. Another challenge is to optimally reduce microseismic motions. This is addressed by suspending the reference cube from the mass of a long-period vertical seismometer. Making the transmitted laser frequencies as stable as possible and improving the accuracies of the reference atomic clocks are ongoing challenges. Design of the release mechanism such that no impulse is applied is also a major challenge.

Symmetric Rise and Fall Absolute Gravity Measurements

Sakuma (3) has developed an up-and-down corner cube absolute gravity measuring system. The reflector is initially catapulted upward, and measurements are made of both the upward deceleration and the downward acceleration. Key advantages of the up-and-down approach are the cancellations of air resistance effects and systematic timing errors. Thus the cube need not reside in a vacuum chamber. Key disadvantages are the mechanical vibrations caused by the upward launching of the mass, its nonportability, and its overall mechanical complexity. The device is permanently mounted on a seismically stabilized platform in Paris.

If the distances x_1 and x_2 (from overhead origin) are passed by the catapulted cube at times t_1 and t_2 , and the free-falling cube later passes x_2 at t_3 and x_1 at t_4 , then the *mean* time values of the x_1 and x_2 passages are equal, that is, $(t_4 + t_1)/2 = (t_3 + t_2)/2$. Using this fact, letting $x_1 - x_2 = \Delta x$, and applying Eq. (1) gives the up-and-down calculation of g as

$$g = \frac{8\Delta x}{(t_4 - t_1)^2 - (t_3 - t_2)^2} \quad (4)$$

Sakuma's nonportable instrument also claims precision and accuracy levels of a few μgals .

RELATIVE GRAVITY MEASUREMENTS

Absolute gravity measurements give acceleration values in terms of the basic units of length and time. Such measurements are important to physicists, astronomers, and geodesists determining the dimensions of the earth. For most other purposes, such as oil and mineral prospecting, understanding the earth's deep structure, and inertial navigation gravity compensations; it is the relatively small variation in gravity from point to nearby point that is important. Such variations can be quickly measured with an easy-to-carry instrument called a gravimeter.

The majority of gravimeters in use today balance the gravity force acting on a so-called proof mass suspended from a metallic spring or quartz fiber *against* the force exerted by the spring. Such a gravimeter is illustrated in Fig. 2. At a starting point where the absolute value of gravity is often known, the gravimeter is leveled and the spring tension is adjusted until it is balanced against the gravitational force acting on the proof mass. The gravimeter is then moved to the next point, leveled, and balanced again. The change in spring tension required to bring the gravimeter back into balance is the actual measurement. This is equivalent to the change in the acceleration of gravity between the two points. Repeated rebalancing (nulling) of the instrument between closely spaced points yields two desirable by-products: (1) lower spring hysteresis (inability to return to its original tension standard after compensating for a tension change caused by an external gravity change) and (2) the gravimeter's beam (b in Fig. 2) is kept close to a horizontal position that reduces sensitivity to leveling errors.

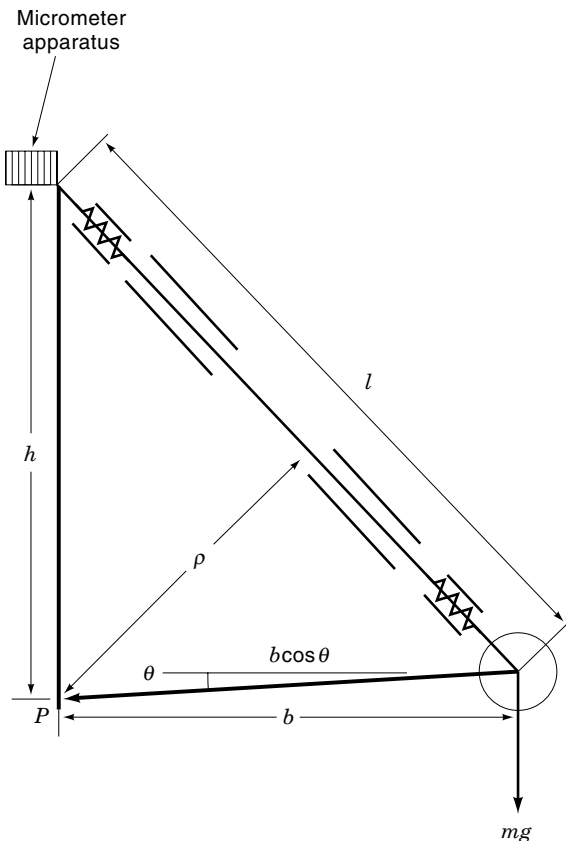


Figure 2. Principles of force balance gravimeter.

Principles of Zero-length Spring, Rebalancing Force, Unstable Gravimeters

A spring is said to be zero-length if its tension is proportional to its actual length. Thus, if all external forces were removed, the spring would collapse to zero length. The key advantage of such a spring is that if its tension supports the beam b and mass m in Fig. 2 in the horizontal beam position, it will support them in any position. Referring to Fig. 2 and the above definition, the spring's tension T is given by

$$T = k(l - l_0) \quad (5)$$

where k is the spring constant, l is the spring's length, and l_0 is the very small length (see discussion below) at which the tension is zero. In a state of true equilibrium or balancing of the two forces, the moments of the weight of the mass (mg) and of the spring's tension (T) about the pivot point P are equal. That is

$$\rho T = mgb \cos \theta \quad (6)$$

where the perpendicular distances ρ and $b \cos \theta$ are shown on Fig. 2, θ being the small angle the beam b makes with the horizontal. Applying trigonometric laws to Fig. 2 and inserting Eq. (5) into Eq. (6) yields

$$l = \sqrt{h^2 + b^2 - 2hb \sin \theta} = \frac{kh l_0}{kh - mg} \quad (7)$$

where h is the distance between the beam pivot point P and the spring end attached to some micrometer screw apparatus (see Fig. 2). Since θ is nearly zero, it follows from Eq. (7) that

$$g \frac{d\theta}{dg} = \left(\frac{l - l_0}{l_0} \right) \left(\frac{h}{b} + \frac{b}{h} \right) \quad (8)$$

From Eq. (8) one sees the gravimeter's sensitivity $d\theta/dg$, that is the change of the beam's angle for a given change in gravity, is greater the smaller l_0 can be made. In practice l_0 is made very small (hence the zero-length name) by winding the coils of a helical spring such that the wire is twisted about an axis in its own length as it is wound (4). Such a gravimeter is classified *unstable* (or *astatic*) because it is a moving system that approaches a point of instability where very small changes in gravity produce relatively larger proof mass displacements.

In zero-length gravimeters the spring is attached to one end of the beam near the proof mass. The spring's other end is attached to the micrometer apparatus. By adjusting the micrometer, the force on the main beam is altered such that when a change in gravity occurs, the beam is returned to the same angle with the horizontal. The change in gravity is shown as an arbitrary scale division on the micrometer's dial, which is easily converted to gravity units. Current zero-length spring gravimeters typically detect changes in gravity at the 1 part/ 10^8 level (0.01 mgal or 10 μ gal). This level of sensitivity requires the spring constant k remain fixed at 1 part/ 10^8 and l_0 be held constant to an even higher degree. Unfortunately both k and l_0 vary with temperature, mainly through the change of the spring's elastic modulus. Therefore gravimeters require a constant temperature environment. This is achieved by housing them in sealed vacuum flasks or in electrically controlled

thermostats. Current gravimeter designs also minimize barometric pressure and magnetic effects.

Gravimeter Range, Accuracy, Precision, Calibration, Drift, and Tidal Effects

A single spring constant k value [see Eq. (5)] cannot give high accuracy measurements over the large range between equatorial and polar gravity. Moreover, if the spring is subjected to large differences in g between force rebalances, it suffers from increased hysteresis. Gravimeter calibration determines k for the specific gravity range to be surveyed. Readings at two or more stations where g is already known gives an average value of k over the range of these stations. The station range must be at least the range of the subsequent survey but not larger than the instrument's range. The latter can be between 5 mgal (geophysical prospecting gravimeters) to over 7000 mgal (global geodetic gravimeters). If the range to be surveyed is large, the calibration stations should be widely separated in latitude. Gravimeters can also be calibrated by tilting them to measure variable components of the g vector. This approach is much more time consuming than field calibrations (5).

There is an inevitable slow and regular change in the length of any gravimeter's spring. The rate of change or *drift* can be determined by returning to a local base as often as the desired accuracy requires (typically every 3 to 4 h). Repeated readings at the same base station over several days produces an oscillatory-shaped drift curve due to tidal effects. The latter result from changes in the gravitational attraction of the sun and moon as their positions change with respect to the earth. Depending on the solar and lunar positions, tidal effects can produce changes in gravity as large as 0.3 mgal over a period as short as 6 h (4). Tidal corrections can be calculated from knowledge of the positions of the sun and moon. However, these effects vary smoothly and slowly and they usually make up part of the gravimeter drift correction itself (unless the required accuracy dictates they be removed). The (tide-free) drift-rate of a fixed-site, specially modified, zero-length, LaCoste-Romberg gravimeter has been reduced to approximately 0.05 mgal/month. This instrument has measured relative gravity at precision and accuracy levels of 1 μ gal (6). Portable, mass produced geodetic gravimeters weigh approximately 2 kg and come with a heating battery to maintain constant temperature. These devices typically measure gravity differences at precision (repeatability) levels of 5 μ gal to 10 μ gal.

RESEARCH AND DEVELOPMENT AREAS

Superconducting Gravimeters

Ultrasensitive cryogenic gravimeters (accelerometers) are being developed wherein the mechanical spring is replaced by a magnetic field produced by current flow in superconducting coils. In the Goodkind (7) device, the field supports a small sphere whose position, determined by the balance between the gravity field and the magnetic field, is monitored electronically. The signal-to-noise ratios of these measurements remain high over a very broad frequency range, covering both the tidal spectrum and very-low-frequency seismic bands (as low as 1 cycle/annum). This instrument has determined major tide components to accuracies better than 0.01 μ gal. The nontidal signal along very low frequencies is mainly due to atmospheric

pressure variations (which can be independently recorded and removed) and the so-called Chandler component of the centrifugal force, a roughly 4 μ gal signal having a period of around 435 days. Superconducting gravimeters remain stable to a few μ gals over such periods and can lead to a better understanding of the Chandler phenomena. Paik et al. (8,9) are developing a superconducting six-axis accelerometer that senses the three translational and three rotational rigid body motions of a single levitated niobium proof mass. This device constitutes a complete inertial measurement unit that is free of drift-prone gyroscopes. Such an instrument could become the core of a passive, nonjammable, virtually drift-free inertial navigation system of the future.

Accurate Shipborne and Airborne Gravity Measurements

When a gravimeter is placed on a moving ship it measures (along its sensitivity axis) the resultant accelerations due to gravity as well as the ship's roll, pitch, and yaw motions. Assuming a velocity of 10 knots, a shipborne gravimeter is typically subjected to high frequency (>1 Hz) horizontal and vertical ship vibration accelerations as large as 0.01 g or 10,000 mgal, 0.01 Hz to 1 Hz ship heaves causing vertical accelerations as high as 0.1 g or 100,000 mgal, and 0.001 Hz to 0.01 Hz ship fishtail or swaying motions which mainly produce horizontal accelerations (4). Therefore only long-period mean gravity values (typically corresponding to frequencies <0.001 Hz) can be precisely and accurately extracted from shipborne gravimeter surveys (using low-pass filtering and averaging techniques over the applicable times and distances). All the motional accelerations cited above can be minimized by placing the gravimeter near the ship's metacenter and mounting it on a shock absorbent, gyroscopically stabilized platform, that is more or less kept level or perpendicular to the average vertical. The stabilized platform also controls cross couplings between the horizontal forces and the highly sensitive vertical response.

In addition to averaging out the ship motion accelerations, one must apply the so-called Eotvos correction to account for the east-west component of the ship's motion. The outward directed centrifugal force component at an earth surface point is given by $R_E \omega^2 \cos^2 \phi$, where R_E is the earth's radius, ω is its angular velocity, and ϕ is the geodetic latitude. At the equator this acceleration is approximately 3400 mgal. If the gravimeter itself has a velocity then the centrifugal force acting on it will be different than if it is stationary. An eastward component of gravimeter velocity numerically adds to the earth rotation effect. This increases the outward centrifugal force and decreases the gravity reading (dominated by the inward attraction of the earth's matter). A westward component of gravimeter velocity has the opposite effect. For an arbitrary gravimeter velocity vector of magnitude V , making a heading angle α with astromonic (true) north, the Eotvos correction is given by

$$E = 2\omega V \cos \phi \sin \alpha + \frac{V^2}{R_E} \quad (9)$$

An east-west velocity of 20 km/h along the equator produces an Eotvos correction of approximately 80 mgal. From Eq. (9) one sees that dE errors in the computed Eotvos corrections are related to dV velocity and $d\alpha$ heading errors by

$$dE = k_1 V \cos \phi \cos \alpha d\alpha + (k_2 \cos \phi \sin \alpha + k_3 V) dV \quad (10)$$

where the k_i are constants. Equation (10) implies an east-west course is most sensitive to velocity errors and a north-south course is most sensitive to heading errors. Therefore, in addition to minimizing the gravimeter's errors caused by the ship's motions and ocean waves, accurate estimates of the ship's positions, velocities, and orientations are required. A shipborne, multiple-antenna (attitude determining) global positioning system (GPS) receiver affordably provides such navigation data. For higher accuracy navigation, GPS data can be integrated with the gyroscopic and accelerometer outputs of an onboard inertial navigation system (INS). The potential of extracting vehicle-borne gravity vectors from integrated GPS/INS data has been examined (10). Therein one pursues the INS's insensitivity to gravity and the GPS's sensitivity to it.

Extraneous accelerations related to high-velocity airborne gravimeter surveys are much more problematic. In particular, large and rapid changes in altitude, linear acceleration, roll, and heading often occur. Even at a low flight speed of 100 km/h, the Eotvos correction in Eq. (9) is around 400 mgal. Thus a 1% error in the estimated speed or a 1 degree error in estimated azimuth introduce gravity errors of approximately 5 mgal. Since gravity decreases by about 1 mgal for every 3 m increase in altitude, and changes in elevation with time produce vertical accelerations, accurate vertical positioning is critical. A GPS or GPS/INS system can potentially meet these navigation requirements. Significant error sources coming from the aircraft's vertical and horizontal accelerations can be further minimized by using gyroscopically stabilized or gimbal suspended gravimeters, flying during periods of low atmospheric turbulence, and increasing the averaging time over which measurements are made (which reduces the resolution of the final mean gravity values). Brozena et al. (11) claimed 2 mgal to 3 mgal accuracies averaged over 20 km.

Gravitational Field Sensors

Jekeli (12) gives a semantically engaging explanation on how a single gravimeter *reacts* to the earth's centrifugal force and *applied* forces such as the one opposite and equal in magnitude to the gravitational attraction of the earth's matter; tidal accelerations caused by the sun, moon, and planets; atmospheric pressures; and any host vehicle accelerations. A gravimeter is insensitive however to a gravitational field (be it the earth's, sun's, moon's, etc.). A gravity gradiometer is a gravitational field sensor. The simplest gradiometer is made up of two closely spaced gravimeters (baseline lengths typically 15 cm to 30 cm) with sensitivity axes pointing in the same (or opposite) directions. By *differencing* (or adding) the two gravimeter outputs, the external forces cancel (at least to a first degree) and one is left *sensing* a gradient of the total gravitation of the solar system along the baseline. The earth's gravitational field clearly dominates those of other bodies and the measured gradient includes a centrifugal force effect. Short-baseline gravity gradient signals are very weak and reflect the higher spatial frequency components of the earth's gravitational spectrum. Engineering challenges abound in the design, fabrication, calibration and mobile operation of ultrasensitive, short-baseline gravity gradiometers. For an overview of these challenges see Jekeli (13). An omnidirectional gradiometer (or full-tensor gradiometer since it measures the full second-order tensor of the gravitational potential scalar) is an elaborate array of 12 or more gravimeters. Gravity gradient measurements interest oil and

mineral prospectors and their line integrals can become real-time gravity vector compensations applied to an accompanying inertial navigation system.

Another type of gravitational field sensor is a satellite-to-satellite electromagnetic ranging device. Like gravity gradients, intersatellite range-rate changes (or line-of-sight accelerations) are functionally related to the difference of the gravity vectors at the two satellite positions. Since satellites are in free-fall and are usually separated by significant distances, onboard gravimeters (accelerometers) can measure the nongravitational forces present at each satellite position (such as atmospheric drag and solar radiation pressure). These external effects can then be removed from the Doppler based satellite-to-satellite range-rate measurement.

BIBLIOGRAPHY

1. T. M. Niebauer et al., A new generation of absolute gravimeters, *Metrologia*, **32**: 159–180, 1995.
2. T. M. Niebauer, J. K. Hoskins, and J. E. Faller, Absolute gravity: A reconnaissance tool for studying vertical crustal motions, *J. Geophys. Res.*, **91** (B9): 9145–9149, 1986.
3. A. Sakuma, A permanent station for the absolute determination of gravity approaching one micro-gal accuracy, *Proceedings of the Symposium on the Earth's Gravitational Field and Secular Variations in Position*. Sydney: University of New South Wales, 1974, pp. 674–684.
4. L. L. Nettleton, *Gravity and Magnetism in Oil Prospecting*, New York: McGraw Hill, 1976.
5. W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied Geophysics*, New York: Cambridge University Press, 1990.
6. K. Lambeck, *Geophys. Geodesy*. Oxford: Clarendon Press, 1988.
7. J. M. Goodkind, Continuous measurement of nontidal variations of gravity, *J. Geophys. Res.*, **91** (B9): 9125–9134, 1986.
8. H. J. Paik, M. V. Moody, and H. A. Chan, Superconducting gravity gradiometer for space and terrestrial applications, *J. Appl. Phys.*, **60** (12): 4308–4315, 1986.
9. H. J. Paik, J. W. Parke, and E. R. Canavan, *Development of a Superconducting Six-axis Accelerometer*, U.S. Air Force Research Laboratory report GL-TR-89-0181, 1989.
10. D. M. Gleason, Extracting gravity vectors from the integration of Global Positioning System and Inertial Navigation System data, *J. Geophys. Res.*, **97** (B6): 8853–8864, 1992.
11. J. M. Brozena and M. F. Peters, An airborne gravity study of eastern North Carolina, *Geophysics*, **53**: 245–253, 1988.
12. C. Jekeli, Does a gravimeter sense gravitation? *Manuscripta Geodaetica*, **17**: 365–372, 1992.
13. C. Jekeli, A review of gravity gradiometer survey system data analyses, *Geophysics*, **58** (4): 508–514, 1993.

DAVID M. GLEASON
United States Air Force Research
Laboratory

VISCOMETERS

The commercial value of numerous fluids produced industrially depends strongly on their viscosity. That is the case when viscosity has a direct effect on how well the fluid performs in its intended application. For example, viscosity plays a significant role in how effectively paint adheres to a wall, and in

the uniformity of the coating. Viscosity also plays a key role in the effectiveness of a lubricant in reducing friction between surfaces, in the power required for pumping sludge from waste reclamation facilities, and so on. The quality of a number of products is specified in terms of well-defined values of viscosity. Such is the case for lubricants, where stringent viscosity specifications have been established since the early 1920s, and also for numerous industrial polymers, as well as inks, paints, and other coating materials.

Definition of Viscosity

When a force is applied to a fluid, the molecules of the fluid are displaced from their positions and move past each other. A velocity gradient is established because molecules located at various points develop different velocities in response to the applied force. This causes a deformation of the fluid: regions of high velocity can be easily imagined to cause a local stretching of the material with respect to regions of lower velocity. Clearly, the extent of the deformation can be quantified in terms of the observed variations of velocity at different locations, or more precisely, in terms of the developed velocity gradient. In some fluids the applied force causes only minor velocity changes, while for other fluids the same force may cause very large changes. The former are commonly referred to as *thin* fluids because they deform easily, posing little resistance to the applied force, while the latter are called *thick* because they are more difficult to deform. If the force is applied to a fluid surface parallel to the direction in which the material is flowing, it is known as a *shear force*, and the rate at which it causes the fluid to deform is called the *shear rate*. The shear rate produced by a given shear stress depends upon the *viscosity* of the fluid, which is a measure of its thickness. In qualitative terms, one may say that the viscosity of a fluid is the property that determines the extent to which the material resists deformation.

A quantitative definition of viscosity is more easily formulated in terms of the system shown in Fig. 1, where a fluid is contained between two parallel plates. A force F is applied to the upper plate, while the lower plate is kept stationary. As a result, the top plate and the fluid layer immediately next to the plate move with a velocity u , while the fluid layer next to the bottom plate remains at zero velocity. Clearly, the applied force has caused a deformation pattern in the fluid. If the surface area of the fluid is A , then the force per unit area, $\tau = F/A$, is called the *shear stress*. Furthermore, the gradient of the velocity field, $\dot{\gamma} = du/dy$, is called the *shear rate* (also called *shear strain rate*), and is interpreted as a measure of the rate of deformation of the fluid. The viscosity of a fluid is denoted

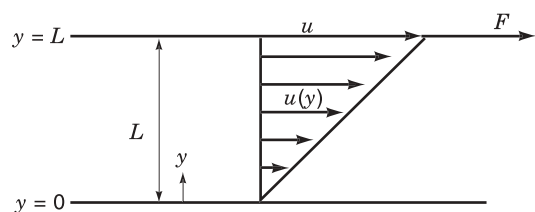


Figure 1. A fluid is sheared between two parallel plates separated by a distance L . While the lower plate is stationary, a force F is applied to the upper plate, causing a deformation in the velocity field such that the top fluid layer flows with velocity u while the bottom layer has velocity zero. The surface area of upper plate is A .

by the symbol η , and is defined as the ratio of the shear stress to the shear rate,

$$\eta = \frac{\tau}{\dot{\gamma}} \quad (1)$$

where in SI units the viscosity η is measured in pascal-seconds (Pa·s), the shear stress τ is measured in pascals (newtons per square meter), and the shear rate $\dot{\gamma}$ is measured in reciprocal seconds (s^{-1}). A common unit of viscosity is the *poise* (P), and $10 \text{ P} = 1 \text{ Pa}\cdot\text{s}$. In turn, the *centipoise* (cP) is defined so that $100 \text{ cP} = 1 \text{ P}$. The *kinematic viscosity* $\nu = \eta/\rho$ is defined as the ratio of the viscosity to the density of the fluid and is measured in units of m^2/s . For the velocity field shown in Fig. 1 it is clear that the shear rate is given by $\dot{\gamma} = u/L$, where L is the distance between the two parallel plates; therefore, application of the general definition (1) to this simple flow system yields that the viscosity is given by the expression $\eta = L\tau/u$. For more complicated shear flow systems, the viscosity is often determined by imposing a shear stress τ , measuring the resulting shear rate $\dot{\gamma}$, and then taking the ratio of these quantities to calculate the viscosity as prescribed by the general definition given by Eq. (1).

In general, the viscosity is a function of the temperature, pressure, and chemical composition of the fluid. The viscosity is also affected by the physical composition of the material, namely, the number and type of different thermodynamic phases present. For example, a multiphase system consisting of water (liquid) and air (gas) will have a very different viscosity from that of the homogeneous water system or the homogeneous gas system, depending upon the relative concentrations of air and water in the multiphase mixture. The viscosity of common fluids is available in tables of physical properties of materials. For example, at 20°C and 1 atm (10^5 Pa) of pressure the viscosity of air is approximately $1.8 \times 10^{-5} \text{ Pa}\cdot\text{s}$, while the viscosity of water is $1.0 \times 10^{-3} \text{ Pa}\cdot\text{s}$, and of ethyl alcohol is $1.8 \times 10^{-3} \text{ Pa}\cdot\text{s}$. Some materials have exceedingly high viscosities; for example, molten glass at 500°C has a viscosity of $1.0 \times 10^{40} \text{ Pa}\cdot\text{s}$. In practical applications fluids are exposed to a wide range of shear-rate values. For example, lubrication oils operating between rotating metal walls in an automobile axle can be exposed to shear rates as large as 10^7 s^{-1} , whereas the sedimentation of particles initially suspended in a fluid exposes the walls of the falling particles to shear rates as small as 10^{-6} s^{-1} .

Newtonian and Non-Newtonian Fluids

A fluid is said to be *Newtonian* if the viscosity remains constant for all shear rates at constant temperature, pressure, and chemical composition of the material. Many common fluids exhibit Newtonian behavior at room temperature. However, other fluids of interest, such as polymer melts, coatings, foods, and many multiphase mixtures, are non-Newtonian in that the viscosity is a function of shear rate. Non-Newtonian fluids can be time-independent or time-dependent, depending on whether the viscosity at constant shear rate remains constant or changes with respect to time. Time dependence of viscosity can be the result of chemical changes (e.g., polymer degradation) or structural changes in multiphase systems (e.g., ketchup, mayonnaise) that occur during shearing. Finally, some non-Newtonian fluids exhibit elastic shear properties, that is, part of the deformation produced by the stress can be recovered when the

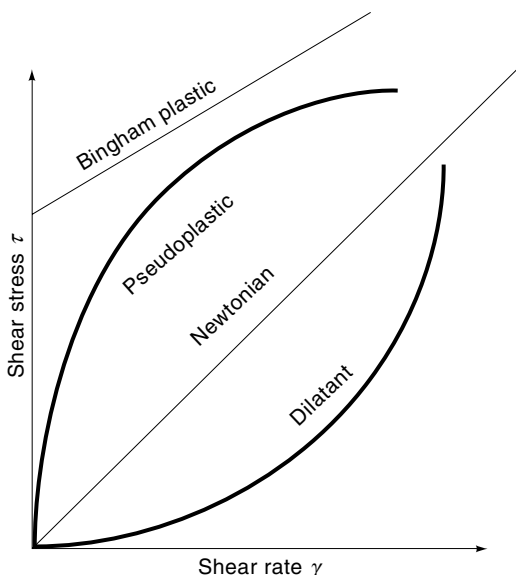


Figure 2. Shape of four typical curves of shear stress versus shear rate for time-independent fluids. The viscosity, given by the ratio of shear stress to shear rate, is constant for Newtonian fluids. Bingham plastics show no deformation for shear-stress values below a minimal nonzero value called the yield stress. Dilatant and pseudoplastic fluids are strongly non-Newtonian materials.

stress is removed. Such materials are called *viscoelastic* fluids. The response of a viscoelastic fluid to shear is strongly dependent on the shear rate. Obviously, all viscoelastic fluids are non-Newtonian, and a complete description of their response is extremely complex. However, in general the viscous portion of the response can be characterized separately from the elastic component.

Figure 2 illustrates several relevant types of viscous responses normally encountered for time-independent fluids. The constant slope of the shear-stress versus shear-rate curve shown in the curve for a Newtonian fluid reflects the fact that the viscosity is constant for all shear rates. Simple homogeneous fluids and most lubricants are Newtonian. If the viscosity decreases with increasing shear rate, the fluid is described as *pseudoplastic*; if it increases, *dilatant*. Nearly all polymers and polymer solutions are pseudoplastic, while many multiphase materials are dilatant. For some fluids, a certain level of shear stress, called the *yield stress*, must be exceeded before the fluid begins to flow. After this stress is exceeded, the fluid may behave as a pseudoplastic or dilatant material, or may adopt a behavior similar to a Newtonian material (as is the case of the Bingham plastic behavior shown in Fig. 2). Many food products and pastes exhibit yield stress.

Measurement of Viscosity

A *viscometer* is an instrument used to measure the viscosity of a fluid. These instruments are also called *rheometers*, since rheology is the science that studies the flow of materials under stress. A viscometer can be used in a laboratory facility, or directly in the industrial production line where the fluid of interest is processed. The laboratory method is an *off-line* technique which often involves a time delay because the sample of the fluid must be transported from the production facilities to

an adjacent laboratory environment. The off-line measurement has the advantage that the viscosity can be determined under highly controlled temperature and pressure conditions, and for a wide range of shear rates, resulting in a more accurate characterization. The delay associated with off-line measurements, however, can result in poor process-control performance, and the production of large quantities of material of undesirable characteristics. Furthermore, factors such as corrosiveness and toxicity of the fluid often complicate the use of off-line methods. Also, for materials such as slurries, the rheology of which is important for many mineral processing operations, off-line measurements may be nonrepresentative of the actual sample in process, due to particle settling and the possible occurrence of irreversible particle interactions such as coalescence and agglomeration.

When measurement delays cannot be tolerated, the viscometer can be installed directly in the path of the fluid being processed (*in-line* installation) or in a bypass line that is directly connected to the main path of the fluid (*on-line* installation). Diagrams illustrating in-line and on-line deployment of a viscometer are shown in Fig. 3. Both in-line and on-line instruments aid in maintaining desired product quality. In-line viscometers measure the viscosity at the temperature, pressure, and flow rate of the flowing fluid. Furthermore, since the process conditions (temperature, pressure, concentration, etc.) are usually measured and values are available, no extra hardware and control effort in addition to the in-line viscometer is needed to characterize the sample being tested. However, the sensor is subject to harsh process conditions and disturbances in the process line may damage the sensor. Also, testing is limited to the process conditions. A good instrument must allow for the fact that process streams are often subject to perturbations in temperature, pressure, and flow rate, which can have a great influence on the viscosity measurement. An additional disadvantage of an in-line installation is that the process flow must be interrupted when the viscometer requires service, causing a discontinuity in the production rate and possibly affecting processing equipment located downstream.

In contrast, on-line installations have the advantages that instrument service and repairs can be done by isolating the bypass line without affecting the flow of the process fluid.

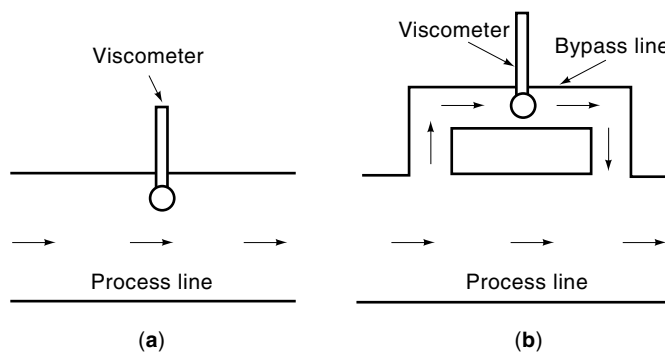


Figure 3. Viscometers deployed in an (a) in-line and (b) on-line fashion. In-line and on-line instruments are designed for use in the production environment and measure the viscosity of the fluid as it flows through its processing lines. The on-line configuration can include flow-control valves and heat transfer equipment to provide stable thermodynamic conditions on the bypass line. Viscometers designed for use in laboratory facilities are known as off-line instruments.

Furthermore, the temperature and flow rate of the sampled fluid can be more closely controlled with the installation of small heat exchangers and flow valves, which can improve the accuracy of the viscosity measurement. However, on-line installations often require a higher capital investment because of the additional cost involved in the installation of the bypass lines, flow valves, and heat exchangers.

In the food and dairy industries, where hygiene is a dominant concern, in-line installations are often avoided because they introduce an additional maintenance overhead due to the need for frequent cleaning and disinfection. However, in many applications in-line installations are necessary in order to maintain adequate measurement conditions, such as laminar flow.

The desirability for in-line or on-line measurement of viscosity, particularly for multiphase materials and for materials from natural sources, has been recognized for at least 20 years. Also, the economic incentives for in-line blending of fluids to reduce inventory and labor costs, for better control of product uniformity, and for increased rates of production have been recognized, and are expected to drive the continuous development of new viscometers and the identification of new applications for the instruments. Implementation has been retarded by the complexity and costs of the instruments, by the range of design modifications necessary for general application, and by the moderate resources available to most viscometer manufacturers, who are constrained by the fact that any particular design will address only a limited market segment. Nevertheless, several in-line and on-line instruments are currently available and are beginning to be used with increasing regularity.

Viscometers can be grouped into the following major types according to the principle used for measurement: (1) falling-element, (2) rotational, (3) pressure-flow, and (4) vibrational. Viscometers of these four types may be designed for use in off-line, in-line, or on-line deployment modes. Each measuring principle involves specific governing equations that relate the viscosity to measurable parameters. Other principles of measurement are possible, but they are less common and are therefore not discussed here. A viscometer must be designed and operated so that flow within the instrument is laminar. In all cases, the approach consists in subjecting the fluid to a known or controlled shear rate and measuring the shear stress, or vice versa.

For accurate viscosity measurements, most instruments require calibration against fluids of known viscosity. This is most often performed with commercially available standard oils that are Newtonian fluids of precisely known viscosity. The calibration procedure consists in measuring the viscosity of the standard fluid and then adjusting one or more parameters in the governing equations until the instrument readings match closely the viscosity of the standard.

Because of specific mechanical constraints, each instrument has limits on the shear rates and shear stresses that can be attained; hence, particular care must be taken in the selection of an instrument for a given application. This is most particularly important for non-Newtonian fluids, since the viscosity may be a strong function of shear rate. Furthermore, viscosity is a bulk transport property, and care must be exercised to ensure that measurements are free of artifacts introduced by the instrument geometry; this concern is particularly significant for the measurement of the viscosity of multiphase materials.

MAJOR TYPES OF VISCOMETERS

This section presents various types of viscometers grouped according to the principle of measurement used, and discusses the underlying physical relationships and governing equations associated with each instrument. For reference, Table 1 lists several commercial viscometers available for off-line use, and Table 2 lists instruments designed for in-line or on-line use. Table 3 gives contact information for several major viscometer manufacturers.

Falling-Element Viscometers

Falling-element viscometers involve dropping a solid weight inside a fluid contained in a vessel and measuring the terminal velocity of the weight as it falls. The viscosity is then determined from a governing equation that uses the terminal velocity as a primary input, and that also makes use of a number of known variables such as the weight and dimensions of the falling element, the densities of the element and of the sample fluid, and other geometric factors. The shear rate produced is the result of a combination of the drag force on the moving surface of the falling element and the

Table 1. List of Selected Off-Line Viscometers That Are Commercially Available

Viscometer Type	Manufacturers	Specific Features	Target Industries
Capillary	Goettfert Contraves Instron Shimidzu	Shear-stress control. Shear-stress control. Shear-rate control. Shear-rate control.	Polymer industry
Rotational	Bohlin Rheometrics Haake Contraves Brookfield Rheologica	Vary depending upon geometry of viscometer as well as fluid properties. Limited to lower shear rates.	Used in all industries covering a wide range of viscosities
Vibrational	Nametre	Newtonian fluids only limited to one frequency. Very broad viscosity range.	
Sliding element	Cambridge Stony Brook Scientific	Fluid need not be Newtonian; used mainly for quality control. Fluid must be particulate-free.	Petroleum, specialty chemicals

Table 2. List of Selected Viscometers Designed for In Line or Off Line Use

Viscometer Type	Manufacturers	Specific Features	Target Industries
Capillary	Rheometrics Seiscor Micro Motion Kayeness	Applicable to non-Newtonian and high-viscosity fluids. Pressure drop required.	Polymers, food, petroleum, pulp and paper
Rotational	Brookfield BTG		
Vibrational	Nametre Dynatrol BTG	Wide viscosity range. Large volume of fluid required for accurate measurement.	Broad applications
Sliding element	Cambridge	Relatively inexpensive. Only clean Newtonian fluids with low viscosity (<200 cP) for accurate measurement.	Petroleum and marine fuels, simple solutions

pressure flow due to the motion of the fluid as it is displaced by the falling element. Therefore, even though the shear rate is significant in one direction only, the flow pattern is nevertheless complex. The most common falling-element viscometers utilize a cylinder or a sphere as weights. The Gardener viscometer uses a perforated disk as the falling element, but although this instrument was widely used in the past, it is now obsolete (1).

Although falling-element viscometers are predominantly used as off-line instruments, several instruments are now available commercially for utilization in an on-line mode with clean, Newtonian fluids. In these instruments a sample from the flowing stream is drawn into the viscometer's vessel by pumping. Then the sample is isolated from the flowing stream by closing an inlet valve, and a measurement of the terminal velocity of the falling element is taken. When the measurement has been completed, the sample is discharged back into the flowing stream, a new sample is introduced, and the process is repeated. An example of such type of on-line falling-element viscometer is an instrument marketed by Cambridge Instruments.

Falling-element devices can be used to measure the viscosity of Newtonian fluids accurately, but are of limited value for non-Newtonian fluids. The problem is that these instruments operate at a fixed shear rate, and therefore cannot describe the shear-rate-dependent viscosity of non-Newtonian fluids. Moreover, the governing equations for the physics that describe the fall, including a situation where the falling element is off center, have not been solved successfully for non-Newtonian fluids. Therefore, use of these devices should be limited to measurements under conditions where the fluid exhibits Newtonian behavior.

Falling-Cylinder Viscometer. Falling-cylinder viscometers are the most commonly used falling-element viscometers. These devices consist of a solid cylinder that falls vertically through a hollow cylindrical vessel of larger radius (2). The sample fluid contained within the hollow vessel is sheared by the solid cylinder, which falls downward over a fixed vertical distance under the influence of gravity, or a combination of gravity and an externally applied force (3). The time required for this distance to be traversed is a function of the fluid's viscosity. As with all falling-element viscometers, the higher the viscosity of the fluid, the lower the velocity of the falling element. When the relative gap width, defined as $(r_0 - r_c)/r_0$, is very small, the vis-

cosity of the fluid can be calculated using the relationship (4)

$$\eta = \frac{F(r_0 - r_c)}{2\pi r_c h u} \quad (2)$$

where F is the force driving the fall of the element, r_0 is the radius of the cylindrical vessel, r_c is the radius of the falling cylinder, and u is the terminal velocity of falling cylinder. The terminal velocity u is easily determined by measuring the time the falling cylinder takes to travel between two reference points in the vessel separated by a known distance. This measurement can be made using optical or electrical detectors that sense the instant when the falling element passes through the reference points. When the cylindrical element falls under the influence of gravity alone, the force shown in Eq. (2) is given by the expression $F = \pi r_c^2 h(\rho_c - \rho_f)g$, where h is the length of the falling cylinder, u is its terminal velocity, ρ_c is its density, ρ_f is the density of sample fluid, and g is the gravitational acceleration. The cylinder's fall can also be driven by electromagnetic forces provided by an electrical coil wrapped around the outside wall of the vessel. In that case the force F used in Eq. (2) must include the prescribed electromagnetic force in addition to the gravitational force. Equation (2) is valid for non-Newtonian as well as Newtonian fluids, provided that the relative gap width is small.

The accuracy of the viscosity measurement determined by the falling-cylinder method may be degraded by several undesirable effects. In particular, the following potential problems are of relevance: (1) end effects in the flow pattern near the top or the bottom of the cylindrical vessel can lead to significant errors in the viscosity measured, and must be taken into account; (2) tumbling may occur while the cylinder is falling through the fluid; (3) a horizontal shifting of the solid cylinder may result in an off-center descent in the vessel. These problems can be alleviated to a certain extent by decreasing the radius of the falling cylinder while increasing its length (2). The net objective is to achieve a larger value of the length-to-diameter ratio (L/D) of the instrument, so that horizontal shifts and end effects are minimized.

Falling-Sphere Viscometer. Falling sphere viscometers operate on the principle of measuring the terminal velocity of a solid sphere as it falls through a sample fluid contained in a cylindrical vessel. As in the case of the falling cylinder, the terminal velocity is simply determined from the measurement of the

Table 3. Contact Information for Selected Viscometer Manufacturers

Company	Laboratory Models: Features	In-Line/On-Line Models: Features
Bohlin Instruments Inc. 2540 Rt. 130 Cranbury, NJ 08512 USA Ph: (609) 655-4447 Fax: (609) 655-1475 http://www.bohlin.com	<i>Visco-88</i> : Coaxial cylinders Cone and plate	
Brookfield Engineering Labs Inc. 240 Cushing St. Stoughton, MA 02072-2398 Ph: (800) 628-8139 Fax: (617) 344-7141 http://www.brookfieldengineering.com	<i>Model DVI+</i> , <i>DVII+</i> : Coaxial cylinders Cone and plate <i>Model SSB</i> : Coaxial cylinders <i>CAP 1000</i> , <i>CAP 2000</i> : High-shear-rate visc. Cone and plate	<i>TT 100</i> , <i>TT 200</i> : Coaxial cylinders <i>TT 100VS</i> : Variable speed <i>TT 220</i> , <i>PVT 230</i> : Control system
BTG 2364 Park Central Blvd. Decatur, GA 30035-3987 Ph: (770) 981-3998 Fax: (770) 987-4126 http://www.btghome.com		<i>VISC 21 5E</i> : Parallel-plate rotation <i>VISC 2000E</i> : Coaxial cylinders <i>MBT 4000</i> : Oscillating plate
Cambridge Instruments 196 Boston Ave. Medford, MA 01255 Ph: (781) 393-6500 Fax: (781) 393-6515 http://www.cambridge-applied.com	<i>ViscoLab 3000</i> : Electromagnetic det.	<i>SPC 301</i> , <i>311</i> : Falling cylinder <i>SPC 372J</i> : Falling cylinder Low flow rate jacket sensor <i>SPC 392</i> : Falling-cylinder flange <i>SPC 401</i> , <i>501</i> : Falling cylinder
Dynatrol Div., Automation Products, Inc. 3032 Max Roy St. Houston, TX 77008 Ph (USA and Canada): (800) 231-2062 Fax: (713) 869-7332		Vibrating probe
Geotfert Inc. 488 Lakeshore Parkway Rock Hill, SC 29730 Ph: (803) 324-3883 Fax: (803) 324-3993 http://www.goettfert.com	Capillary rheometers Mooney viscometers	Capillary rheometers
Haake 53 W. Century Rd. Paramus, NJ 07652 Ph: (800) 631-1369 Fax: (201) 265-1977 http://www.haake-usa.com	<i>Viscotester</i> , <i>VT 550</i> : Rotational viscometer, varying geometries	
Instron Corporation 100-TR Royall St. Canton, MA 02021-1089 Ph: (800) 373-6978 Fax: (781) 575-5751 http://www.instron.com	Capillary rheometers	
Kayeness Polymer Test Systems 115 Thousand Oaks Blvd., Suite 101 P.O. Box 709 Morgantown, PA 19543 Ph: (610) 286-7555 Fax: (610) 286-9396 http://www.dynisco.com	Capillary rheometers	Capillary rheometers Slit rheometers
Micro Motion, Inc. 7070 Winchester Circle Boulder, CO 80301 Ph: (303) 530-8400 Fax: (303) 530-8596 http://www.micromotion.com		<i>Coriolis flow meter</i> : Measures pres- sure drop and flow rate

Table 3. *Continued*

Company	Laboratory Models: Features	In-Line/On-Line Models: Features
Nametre 101 Liberty St. Metuchen, NJ 08840 Ph: (732) 494-2422 Fax: (732) 494-8916 http://www.nametre.com	Vibrational Torsional oscillation Spherical and cylindrical	<i>Viscoliner</i> : Vibrational Torsional oscillation Spherical and cylindrical
Norcross 255 Newtonville Ave. Newton, MA 02158-1898 Ph: (617) 969-7020 Fax: (617) 969-3260 http://www.viscosity.com		<i>In-Line Viscometers M10, M20, M24, M50</i> : Atmospheric-pressure viscometers to be mounted on tanks
Rheometrics Scientific Inc. One Possumtown Road Piscataway, NJ 08854 Ph: (732) 560-8550 Fax: (732) 560-7451 http://www.rheosci.com	Dynamic stress rheometer Differential thermal rheometer Bench-top viscometers	

time taken by the sphere to traverse a fixed distance (5). The viscosity can be calculated from Stokes's equation (6)

$$\eta = \frac{2(\rho_s - \rho_f)r_s^2g}{9u} \quad (3)$$

where ρ_s is the density of the sphere, ρ_f is the density of the fluid, r_s is the radius of the sphere, u is the terminal velocity of the sphere, and g is the gravitational acceleration. Stokes's equation is applicable when (1) the sample fluid is Newtonian and (2) the radius of the sphere is much smaller than the radius of the cylindrical vessel.

Rotational Viscometers

Rotational viscometers utilize the principle of shearing a fluid between two walls (one moving in a rotational motion and one stationary) to create a velocity profile in the liquid as well as to impose a force on the stationary wall. The resulting type of flow is known as *drag flow*. The walls, or boundaries, are arranged with specific geometries that allow the viscosity of the liquid contained between them to be determined through measurement of the force applied on the stationary wall. The best-known drag-flow instruments are the cone-and-plate viscometer, the parallel-plate viscometer, and the coaxial-cylinder viscometers.

Cone-and-Plate Viscometer. The cone-and-plate viscometer is a rotational drag-flow device in which the sample fluid is sheared between a rotating cone and a fixed plate (Fig. 4) (2). The cone is usually situated above the plate, with a point of contact at the vertex of the cone and the midpoint of the plate (7). In order to achieve a shear rate that depends only on the direction of rotation (i.e., that is independent of the radial direction), the angle between the cone and the plate surfaces is small (usually smaller than 3°) (8).

At constant angular velocity, the shear stress in the fluid is proportional to the torque exerted by the fluid on the cone. For small angles between the cone and the plate surface, the shear rate within the fluid is constant. The viscosity of the sample at

a particular shear rate can be calculated as (3)

$$\eta = \frac{3\tau\alpha}{2\pi R^3\Omega} \quad (4)$$

where τ is the torque exerted on the cone, α is the angle between the cone and plate, Ω is the rotational velocity of the cone, and R is the radius of the cone. Equation (4) is valid for all incompressible fluids, provided there is no slip at the surface (2).

Cone-and-plate viscometers are easy to operate; however, there are several sources of error that can affect the measurement, including (1) fluid inertia and secondary flow, (2) edge effects, (3) nonuniformity of the shear field due to a large cone angle, (4) viscous dissipation, and (5) nonideal geometry.

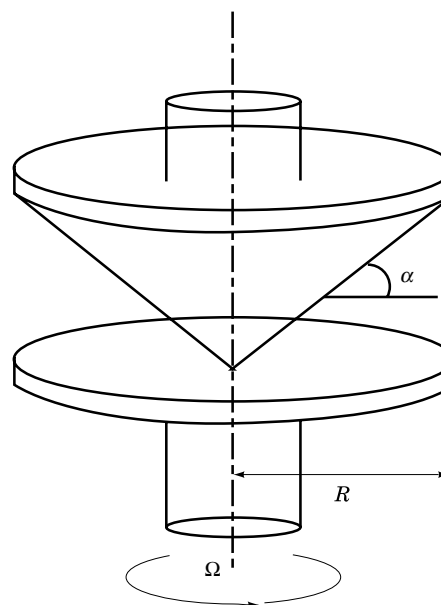


Figure 4. Schematic of a cone-and-plate viscometer geometry where the plate rotates at a constant angular velocity while the cone is fixed. The angle between the cone and plate is typically smaller than 3° . The fluid is confined between the surfaces of the cone and the plate.

Collyer and Clegg (2) or other pertinent references should be consulted for information on appropriate procedures to avoid or minimize these errors. The operation of cone-and-plate viscometers is restricted to creeping flow conditions; therefore their use is limited to low shear rates, typically smaller than 20 s^{-1} .

Parallel-Plate Viscometer. A parallel-plate viscometer is simply a cone-and-plate viscometer in which the angle between the cone and the plate is zero and there is no longer a contact point between the two surfaces. This instrument is particularly useful for the measurement of the viscosity of multiphase materials; however, it is restricted to measurements at very low shear rates, and the shear rate is no longer constant throughout the fluid. In a given sample, the viscosity at a particular shear rate (the shear rate at the edge of the fluid) can be calculated as (3)

$$\eta = \frac{\tau h}{2\pi \Omega r^4} C_F \quad (5)$$

where τ is the torque exerted on the plate, Ω is the angular velocity of the moving plate, r is the plate radius, h is the distance of separation between plates, and C_F is a correction factor defined as

$$C_F = 3 + \frac{d \ln(\tau/2\pi r^3)}{d \ln(\Omega r/h)} \quad (6)$$

In general, the parallel-plate viscometer is subject to the same limitations as the cone-and-plate viscometer.

An instrument for on-line viscosity measurement using parallel-plate or cone-and-plate geometries is available through BTG. This instrument may yield relative viscosity values that are useful for process control purposes, but it is unlikely to yield accurate on-line measurements, since the flow does not have a free surface at the boundary between the plates. For measurement accuracy there must be virtually no flow through the gap between plates during a measurement; however, some flow is necessary to make measurements of viscosity periodically on fresh samples of the flowing fluid.

Rotating-Cup Viscometer. This viscometer consists of two coaxial cylinders that contain the fluid to be measured in the annular gap between them. One cylinder (called the *stator*) is held stationary while the other cylinder (called the *rotor*) is rotated. Two configurations are possible. In one configuration the cup rotates inside the hollow stator, and in the other the immobile stator is located inside the rotating cup. The fluid being measured is sheared in the annular gap. The shear stress can be directly determined by measuring the torque delivered by the motor that drives the rotating cup, or by measuring the torque required to hold the stator immobile. The shear rate can in turn be directly determined by the rotational speed of the cup. For the configuration that features the stator positioned inside the cup, the shear stress and shear rate are given by

$$\tau = \frac{1}{2\pi R_o^2 L_c} T \quad (7)$$

$$\gamma = \frac{4\pi C_F}{1 - (R_i/R_o)^2} N \quad (8)$$

where T is the measured motor torque, N is the rotational speed of the cup measured in revolutions per second, the geometrical parameters R_i and R_o are the inner and outer radii of the gap, and L_c is the effective length of the stator. The parameter C_F is an experimental curvature-flow-index correction factor that takes into account the fact that the curved walls of the gap are not exactly parallel plates and that the fluid may be non-Newtonian (9–12). Using the viscosity definition (1), the working equation for the rotating-cup viscometer reduces to

$$\eta = \frac{1 - (R_i/R_o)^2}{8\pi^2 R_o^2 L_c C_F N} T \quad (9)$$

When the correction factor is arbitrarily set equal to $C_F = 1$, ignoring the effects of curvature, then Eq. (7) yields the *apparent shear stress*, Eq. (8) gives the *apparent shear rate*, and Eq. (9) the *apparent viscosity*. For accurate viscosity determinations, the instrument is calibrated with standard fluids to determine effective length L_c for use in Eq. (9) (4,13).

A typical approach for determining viscosity using the governing Eq. (9) is to first fix the rotational speed N at a desired operating value, and then rewrite Eq. (9) in the form

$$\eta = \phi T \quad (10)$$

where ϕ is a parameter determined experimentally by measuring the torque attained when the instrument is loaded with a standard calibration oil of precisely known viscosity. Equation (10) is then used to calculate the viscosity of any fluid at the given rotational speed. Other specialized types of rotational instruments, such as the Stormer viscometer (1), were widely used in the past, but are now superseded by coaxial cylinder viscometers for laboratory use.

It must be emphasized that the flow induced within the gap by drag must be laminar. If the inner cylinder is the rotor, the curl in the fluid caused by the rotation of the inner cylinder can induce secondary rotational flows imposed on the laminar flow. This can introduce serious error. Finally, viscous dissipation within the fluid can introduce errors due to temperature changes if the fluid is sheared for extended periods. The useful shear-rate range varies with the instrument and fluid, but is generally limited to 600 s^{-1} to 800 s^{-1} if the inner cylinder is the rotor, and 1000 s^{-1} to 1300 s^{-1} if the outer cylinder is the rotor. Rotational viscometers are popular measurement devices because they require small sample sizes, are reasonably easy to use, deliver accurate measurements if used correctly, and can be used for non-Newtonian as well as Newtonian fluids. Their principal disadvantage is that in order to protect the motor and electronics from the fluid, the moving part uses a seal that may require regular check and maintenance.

A commercial rotating-cup viscometer produced by Brookfield Engineering is shown in Fig. 5. The schematic shown is for an in-line instrument that features a rotating cup and a stator mounted vertically inside the cup while the process fluid flows along a horizontal pipe. Most of the material flows around the external walls of the rotor, bypassing the instrument to reach the outlet port. A fraction of the fluid enters the sampling chamber from the bottom and flows upwards into the measuring gap, where it is sheared by the rotating cup. The measured fluid then leaves the measurement gap by flowing upwards and exiting through openings at the top of the rotor. The pressure flow through the gap superimposes an axial shear-rate component

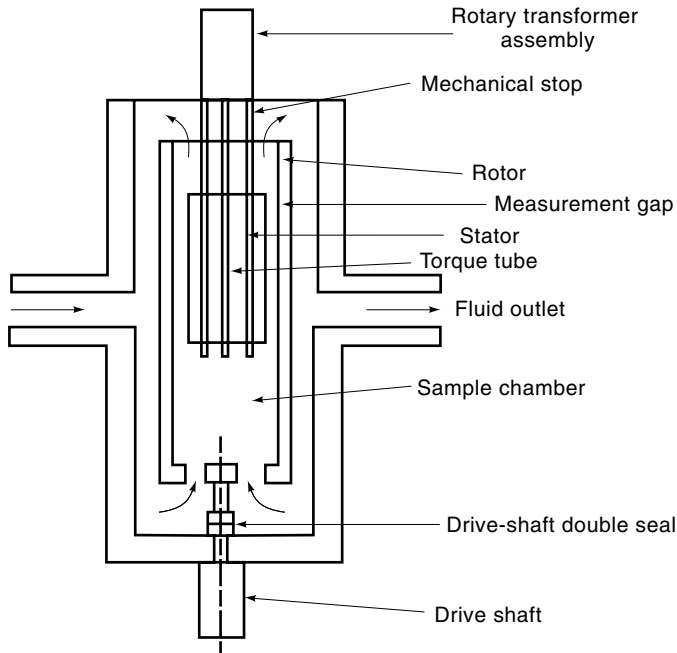


Figure 5. Schematic of an in-line rotational viscometer. The torque required to maintain a specified rotational speed is related to the viscosity of the sample.

on the rotational component. Care must be exercised to ensure that the stress resulting from pressure flow through the instrument gap is negligible compared to the stress resulting from the drag flow. This can be controlled by varying the bypass area for flow, by varying the total flow rate, or by varying the instrument dimensions.

BTG has also marketed a coaxial cylinder viscometer for in-line or on-line measurements. In the proposed design the entering flow is directed through the gap, which leads to superimposing a large pressure flow stress, unless the flow is interrupted for measurement. With continuous flow, this instrument can give only relative values, even with constant flow rate.

Pressure-Flow Viscometers

Several methods for measuring viscosity are based on the properties of axial flow through a channel, where the flow is driven by the pressure difference between the ends of the channel. Ideally, the pressure difference should be measured for fully developed laminar flow through a channel of circular cross section (i.e., a tube); however, one-dimensional flows through a slit have also been used successfully. Usually, the pressure difference is measured over the entire channel length; nevertheless, this practice is not recommended, because it can introduce errors due to entry and exit effects. These adverse effects can be minimized by using tubes with large L/D ratios.

Pressure-flow viscometers are most often constructed using a *capillary*, which is simply a tube of small diameter. The pressure drop needed to maintain a sustained liquid flow rate can be supplied mechanically (using a pump or a plunger, for example), or simply by the fluid head due to gravity. In the latter case, pressure-flow viscometers allow the determination of only the kinematic viscosity, $\nu = \eta/\rho$, where ρ is the density of the fluid. The viscosity η of the fluid can be found from the

measured kinematic viscosity provided that the density of the material is known or is measured using a densitometer. The most common types of pressure-flow instruments include classical capillary viscometers capable of accurate measurements, glass capillary viscometers in widespread use for characterizing Newtonian fluids, and specialized capillary viscometers that have been standardized for use in selected fluids.

Classical Capillary Viscometers. A number of pressure-flow viscometers featuring a capillary tube are available for making accurate viscosity measurements for a wide range of fluids. The laboratory versions of these instruments can be configured in a stress-controlled or a shear-rate-controlled mode, depending on whether the instrument operates at a constant shear stress or at a constant shear rate. For reference, selected instrument manufacturers are listed in Table 1. In response to market demands, a large number of the instruments available commercially have been developed for the study of highly viscous materials. Generally, end effects are minimized by using long capillaries ($L/D > 50$), or by making measurements with capillaries of different lengths and the same diameter and using the results to make appropriate corrections for end effects (14–17). With such techniques accurate measurements of viscosity as a function of the shear rate at the wall of the capillary can be made. When the flow is laminar, the shear rate at the wall of the capillary is given by the expression (3,4,18)

$$\dot{\gamma}_w = \frac{Q}{\pi r^3} \left(\frac{3n+1}{n} \right) \quad (11)$$

with

$$n = \frac{d \ln \tau_w}{d \ln Q} \quad (12)$$

where Q is the volumetric flow rate through the capillary, r is the radius of the capillary, τ_w is shear stress at the wall, and $\dot{\gamma}_w$ is shear rate at the wall. The parameter n defined in Eq. (12) is known as the *flow index*, and for Newtonian fluids it takes the value $n = 1$. The shear stress at the wall of the capillary is in turn given by the expression (3,4,14)

$$\tau_w = \frac{r \Delta P}{2h + 4r C_F} \quad (13)$$

where ΔP is the pressure drop across the capillary, r and h are respectively the radius and the length of the capillary, and C_F is a correction factor that must be determined experimentally. From Eq. (1) it follows that the viscosity is given by the expression

$$\eta = \left(\frac{\pi r^4}{2h + 4r C_F} \frac{n}{3n+1} \right) \frac{\Delta P}{Q} \quad (14)$$

An experimental measurement of viscosity using a capillary viscometer consists of measuring the volumetric flow rate Q along with the pressure drop ΔP across the capillary, and then using Eq. (14) to compute the viscosity. The greater the L/D ratio of the capillary, the more accurate the measurement, due to decrease in end effects. In every case, the flow rate through the capillary must be adjusted to ensure laminar flow.

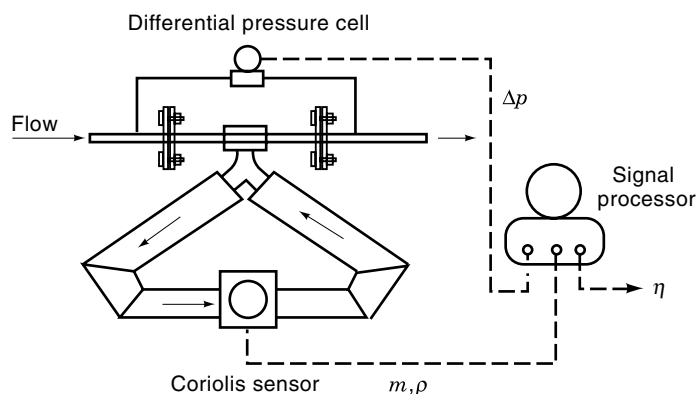


Figure 6. Schematic diagram of an on-line capillary-Coriolis viscometer geometry. The Coriolis sensor measures the mass flow rate m and the density ρ while the differential-pressure cell measures the pressure drop Δp . A signal processor combines the three measurements to produce the viscosity output η .

In general, off-line versions of classical capillary viscometers require relatively small volumes of fluid samples, and permit making measurements at shear rates of $30,000 \text{ s}^{-1}$ or higher. In most cases, the commercially available laboratory instruments of this type are designed for fluids with high viscosity (typically in excess of 600 cP). Furthermore, laboratory versions of these instruments are typically inexpensive; in contrast, in-line and on-line versions of pressure flow viscometers are expensive, and therefore have found applications in manufacturing operations where there are significant economic incentives to justify the higher capital expenditure, such as in the processing of polymer melts. On-line and in-line instruments are marketed by several providers, including Kayeness and Rheometrics, among others (Table 2).

A recent variation of a capillary viscometer that is being successfully used in a number of on-line applications consists of combination of two sensors, namely, (1) a Coriolis mass-flow meter that measures the density and the mass flow rate of the fluid, and (2) a differential pressure cell that measures the pressure drop across a capillary tube (19). Recognizing that the volumetric flow rate Q and the mass flow rate m are related by $Q = m/\rho$, where ρ is the density, it is possible to rewrite Eq. (14) in the form

$$\eta = \phi \frac{\rho \Delta P}{m} \quad (15)$$

where ϕ is an instrument constant that in principle is equal to the factor inside the parentheses on the right-hand side of Eq. (14). In practice, ϕ is determined experimentally for the specific fluid of interest, and may be a function of temperature. A capillary-Coriolis viscometer arrangement similar to that marketed by Micro Motion Inc. is shown in Fig. 6. It is preferred that the pressure drop be measured across a straight tube upstream or downstream of the Coriolis meter to minimize turbulence and end effects. For common combinations of Coriolis and pressure-cell sensor pairs, the instrument can operate in a range of viscosities differing by a factor of 6.

Glass Capillary Viscometers. Pressure-flow viscometers featuring a glass capillary, such as the Cannon, Ostwald, and Ubbelohde viscometers, have been used for many years in

off-line environments to measure the viscosity of Newtonian fluids. Since in these devices the pressure driving force is supplied by the fluid head, the instruments measure the kinematic viscosity. The principle of measurement consists of determining the time needed for a given volume of fluid to pass through the glass capillary. This lapse of time is known as the *time of efflux*, and can be used to calculate the kinematic viscosity via the equation

$$\nu = \phi_1 t + \frac{\phi_2}{t} \quad (16)$$

where ϕ is the kinematic viscosity, ρ is the density, t is the time of efflux, and ϕ_1 and ϕ_2 are constants determined by calibration against fluids of known viscosity. These glass capillary viscometers are useful for Newtonian fluids only. Usually, the instrument is used to measure viscosities of fluids for which the term ϕ_2/t is negligible.

Specialized Capillary Viscometers. Specialized pressure-flow viscometers that carry out measurements under conditions specified by standards organizations have been used for many years for specific materials. For example, the *Zahn cup*, a fluid-head device with an extremely short L/D ratio, has been used to control the viscosity of paints, inks, dyes and other materials as specified by ASTM D4212-93 (1990). Also, the *Saybolt viscometer*, a device with a short L/D ratio, incorporates a load-bearing plunger on the chamber containing the fluid to add a significant pressure difference to the fluid head. The time for a known volume of fluid to be forced through the capillary under the specified load is measured, and the viscosity is reported in arbitrary time units commonly referred to as *Saybolt viscosity units*. Specifications for the test are given in ASTM D1238-90b (1990). Saybolt viscometers have been used so extensively that tables for converting Saybolt viscosity measurements to poise and other conventional units have been developed for Newtonian fluids and can be found in standard references (20).

Another important specialized viscometer of pressure-flow type is the *melt indexer*, which is used extensively for polymer melts. The melt contained in the viscometer chamber is driven through a capillary of small L/D by a fixed load to give a constant pressure difference. The measured mass of polymer that flows through the capillary in a specified period of time is determined, and is reported in grams as the *melt index* (or *melt flow* for some polymers). Complete specifications for the use of this instrument are given in ASTM D445-88 (1990).

Vibrational Viscometers

Vibrating elements immersed in a fluid can be used to measure the viscosity, because the energy required to sustain a given frequency and amplitude of oscillation depends upon the viscosity of the fluid. A wide variety of vibrational viscometers, which differ primarily in the geometry of the vibrating element, have been developed. The most commonly used elements are spherical or cylindrical probes, as well as wires, rods, or plates. The vibrating element is fixed at one end, allowing motion about a pivotal point. Typically, the element is made to oscillate by an induced magnetic force. Vibrational viscometers are used in a variety of industries, due to their versatility.

The principle of operation of these instruments is based upon either controlling the amplitude or the frequency of oscillation

of the element. In an amplitude-controlled vibrational viscometer, a force is applied to the element in such a way as to make it oscillate at a high frequency but at constant amplitude. The power required to maintain a particular amplitude is directly related to the viscosity (21). For a given amplitude, the power required increases as the viscosity of the sample increases. Frequency-controlled viscometers operate using an analogous principle, but keep the frequency at a user-specified constant value.

The physics of oscillation of a spherical probe in a fluid medium at a constant amplitude is described by the equation (6)

$$\eta = \frac{1}{2} \delta^2 \omega^2 \rho^2 \quad (17)$$

where ω is the vibrational frequency that yields the required amplitude, ρ is the density of the fluid, and δ is the distance the oscillatory wave propagates through the sample medium until its amplitude falls to $1/e$ of its original level, where e is the base of natural logarithms. Therefore, there is a fundamental relationship between the viscosity and the frequency. The specific governing equation for a practical viscometer must take into account a number of additional factors, including the geometry of the vibrating elements as well as the compartment size in which the element vibrates. Note that for a spherical probe the mass of the vibrating element can be considered to be concentrated at a point, whereas for a vibrating rod the mass is distributed throughout the entire element.

A commercial vibrational viscometer for in-line or on-line use, produced by the Nametre Company, is shown in Fig. 7. The vibrating sensor is located inside the measuring chamber. The fluid flows past the instrument from the bottom, after colliding with a deflector plate that protects the sensor from direct fluid impact. The material flows through the chamber and exits

through the outlet port at the top of the instrument. The instrument can be fitted with a heating jacket in order to maintain a constant temperature. The Nametre Company viscometer operates on an controlled-amplitude mode; hence, under normal operation, the transducer that causes the sensor to vibrate is driven in such a way that the amplitude of motion of the sensor is held constant at approximately $1 \mu\text{m}$, and the average electrical power P required to accomplish this constant amplitude is measured. It can be shown that the product of the viscosity of the fluid times its density is given by the expression

$$\eta\rho = \frac{2}{\omega} \frac{P^2}{\phi} \quad (18)$$

where η is the viscosity, ρ is the density, ω is the frequency of vibration, P is the power, and ϕ is an experimental factor specified by the manufacturer and stored in the instrument's on-board electronics. Note that the viscosity of the fluid can be obtained from the viscosity-density product measured by the instrument if the density of the material is known or is measured using an on-line densitometer. The shear rate is given by $\dot{\gamma} = \omega$, where ω is measured in radians per second. Equation (18) gives accurate estimates of the viscosity-density product of fluids that do not show viscoelastic behavior, such as Newtonian fluids, for example. When the sample fluid is viscoelastic, the viscosity appearing on the left-hand side of Eq. (18) is interpreted as a *nominal viscosity* because it is affected by the energy-storage behavior of the elastic modes of the fluid.

Vibrational viscometers for on-line measurements are marketed by a number of companies. These devices can also be used for laboratory measurements. In particular, Nametre Company, BTG, and Dynatrol market vibrational viscometers. The measurements are fairly independent of the flow rate of the fluid being measured; in fact, the flow can even be moderately turbulent. Most instruments operate at a single frequency of oscillation, hence they operate at a single shear rate, a disadvantage when characterizing the flow of non-Newtonian fluids. During operation of the instrument, extreme care must be taken to keep the vibrating element free of deposits that could affect the total mass of the element and lead to biased measurements.

Challenges in On-Line Viscosity Measurement

On-line measurements using the traditional principles of operation introduce many technical challenges, and appropriate remedial measures are necessary to obtain reliable measurement and control. The problems and solutions depend on the particular process at hand. For example, if a rotational device is used, with a superimposed axial flow, a spiral flow within the instrument results, and this can introduce errors in measurement at high flow rates. If a capillary-Coriolis device is used, the flow rate must be high enough to produce an appreciable pressure drop between the points of measurement, but the flow regime must still be laminar. With a falling-element device, axial flow can force the falling element to hit the wall of the instrument, and this can result in malfunction of the instrument. There is also the possibility of corrosion of the equipment or deposits on elements of the instrument due to continued exposure to the process conditions. This can change critical dimensions or change the response of measuring elements.

These are problems of a general nature. In addition, every application gives rise to a unique set of challenges. For example,

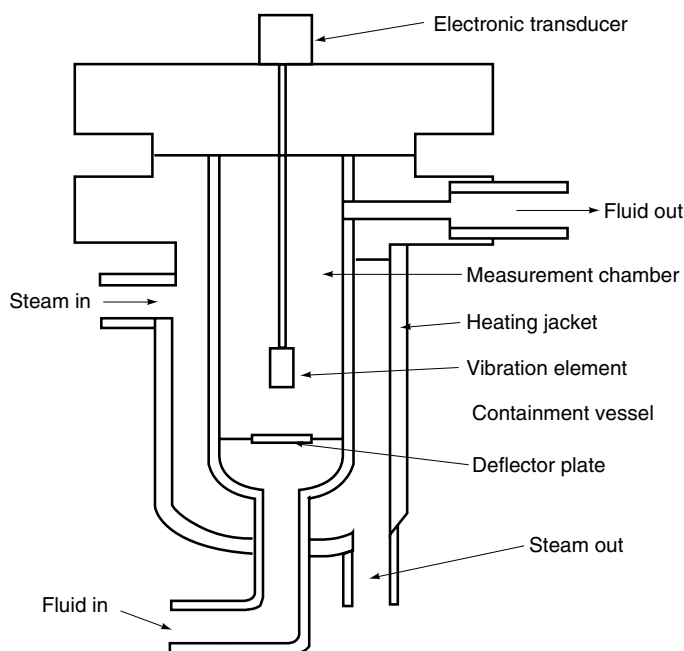


Figure 7. Schematic of an in-line vibrational viscometer. The power required to maintain the vibration element at a constant amplitude is related to the viscosity.

preserving the phase distribution of the sample in a multiphase flow (as in suspensions, slurries, and pastes) during measurements is a major challenge.

Reported Applications of On-Line Viscometers

Many application studies have been reported in the literature on using in-line and on-line viscometers in process industries. The following abridged list gives an overview into the range of applications of on-line viscometers. See Table 2 for an overview of commercial in-line and on-line viscometers available in the market, and Table 3 for contact information. Kalotay (19) and Bates (22) discuss the principles and various applications of on-line capillary–Coriolis and on-line rotational viscometers, respectively, for measurement and control. See also De Laney et al. (23). Essegheir et al. (24) discuss the development of in-line and on-line sensors to monitor blend rheology and morphology for microstructure monitoring in compounding polymer blends. Baker et al. (25) discuss an on-line rheometer to monitor the change in melt index of polymer products. Broadhead et al. (26) used an in-line rheometer for closed-loop viscosity control of reactive extrusion. Kawatra et al. (27) studied the effect of slurry viscosity on hydroclone classification by installing an on-line viscometer on a hydroclone system. Shi and Napier-Munn (28) have developed a new procedure for obtaining a full shear-rate–shear-stress curve by using an on-line viscometer. Kawatra and Bakshi (29) have used vibrating-sphere and rotational on-line viscometers for measurements on mineral suspensions. Pillo et al. (30) describe using a viscometer in closed-loop control of recycled high-density polyethylenes during extrusion; the melt viscosity was measured with an on-line rheometer and used as the indicator of polymer quality. Speight et al. (31) use in-line process viscosity monitoring for injection molding control.

The advantages of using an on-line viscometer are also being realized in the food and the pulp and paper industries. For example, in the manufacture of skim milk powder, it is known that using an on-line viscometer to control the degree of evaporation leads to a significant reduction of steam consumption and stack losses, and to the minimization of fouling problems (32). Furthermore, maintaining a constant viscosity also results in a more consistent product, as the mean powder particle diameter is known to be directly proportional to the viscosity of the skim milk concentrate (33).

It is also known that using on-line viscometers in the pulp and paper industry can assist in obtaining enormous energy savings, and studies have been done to evaluate various principles of viscosity measurement (34,35) for this purpose.

Future Developments

There are continuing efforts to develop new types of on-line rheometers for various applications. For example, see Todd et al. (36) for a helical-barrel rheometer, and Arola et al. (37) for an NMR-based method in which nuclear magnetic resonance imaging and fundamental principles of capillary flow are combined.

Rapid advances in the development of viscometers have been made in the last two decades, and the pace of development and applications is accelerating, as is seen from the examples cited. However, much more is needed. The viscosity range that is of interest for industrial fluids varies by about five orders of magnitude. Also, the rheology of the fluid is an additional complication. Technology is available now for accurate on-line

measurement of Newtonian, single-phase fluids and for time-independent non-Newtonian single phase fluids. However, accurate measurements can be made now for only a limited number of multiphase fluids, and time-dependent viscosity is extremely difficult (virtually impossible) to measure on line at present. Corrosive process conditions add to the complexities.

On-line viscometers are moderately expensive devices. That they are nonetheless being adapted for measurement and control in such a diverse array of industries is the best indication of their potential value.

IMPORTANT ISSUES IN VISCOSITY MEASUREMENT AND CONTROL

Challenges Posed by Non-Newtonian Fluids

Incompressible Newtonian fluids at constant temperature are completely characterized by two *material constants*: density and viscosity. For non-Newtonian fluids the rheology leads to *material functions* that are dependent on shear rate, frequency, composition, and time (3). These complexities make it significantly more difficult to measure and control the rheological properties of non-Newtonian materials. However, a wide class of industrially important materials such as polymer solutions, polymer melts, motor oils, varnishes, adhesives, coatings, paints, and foods are non-Newtonian.

From an application viewpoint, the most important rheological property is the non-Newtonian viscosity. Viscosity has very significant effects on heat transfer, fluid transport, droplet and bubble formation and breakup, foaming, lubrication, coating and calendering, extrusion, mixing, and molding, just to mention a few commercially important operations. Proper measurement of viscosity on line can lead to major improvements in these operations and processes.

In addition, the viscosity of the fluid is an important performance specification for the final product. Direct on-line measurement can lead to vastly improved control of product uniformity. This possibility is particularly important for foods, coatings, and lubricants. It is, of course, complicated by the wide range of viscosity exhibited by industrially important fluids. Even a single fluid can exhibit very large variations in viscosity during processing due to temperature, concentration, and shear-rate changes. For example, Kraft black liquor obtained from pulping only one wood species can vary in viscosity by more than three orders of magnitude in the course of processing (38–41). At high concentration, the viscosity of such liquors can vary by more than an order of magnitude at the same temperature, concentration, and shear rate due to variations in composition and interaction of the constituents. This is not uncommon for concentrated solutions or suspensions derived from natural sources.

The phenomenological behavior of any one of these complex fluids is uniform, even though the viscosity varies widely. Various theories proposed to explain the rheological behavior of non-Newtonian fluids have led to models of varying complexity, ranging from generalized Newtonian models such as the power-law model (3), the Cross model (42), and the Carreau–Yasuda model (43,44), to more complex nonlinear viscoelastic models (3). These models yield the general response. For limited ranges of shear rate, two measurements in the shear-sensitive region can be used to define the viscosity

behavior. Thus, it is possible to control the viscosity of Newtonian and simple non-Newtonian time-independent fluids with one or two on-line measurements if the general rheological behavior of the fluid is known.

Challenges Posed by Multiphase Fluids

The problem of measuring viscosity is more challenging in the case of multiphase materials. The shear-rate-dependent viscosity and the flow behavior change markedly with phase distribution and particle size distribution. Many industrially important products are in this category. Examples include milk products, mayonnaise, peanut butter, unleavened dough, chocolate, cheese spreads, latex paints, and coating materials such as inks and paints. Measurement is usually difficult, and on-line control is virtually impossible. For example, see Kawatra and Bakshi (45) for an extensive discussion of various problems involved in the on-line measurement of slurry viscosity using rotational, capillary, and vibrational viscometers. However, if the slurry can be shown to behave as a bulk fluid by obtaining off-line measurements made in different viscometer geometries that are in agreement with each other, on-line measurement and control is possible.

Challenges Posed by Fluids with Time-Dependent Viscosity

A third major challenge is time dependence of the viscosity of materials. For example, the viscosity of tomato juice decreases with time of shearing, whereas that of egg white increases. The usual approach to this problem is to measure viscosity under identical and controlled conditions in the laboratory. On-line measurement and control is still an open issue.

BIBLIOGRAPHY

1. J. H. Perry, *Chemical Engineers' Handbook*, 3rd ed., New York: McGraw-Hill, 1950.
2. A. A. Collyer and D. W. Clegg, *Rheological Measurement*, New York: Elsevier Applied Science, 1988.
3. R. B. Bird, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids*, New York: Wiley, 1977.
4. J. M. Dealy, *Rheometers for Molten Plastics*, New York: Van Nostrand Reinhold, 1982.
5. Y. I. Cho, J. P. Hartnett, and W. Y. Lee, Non-Newtonian viscosity measurements in the intermediate shear rate range with the falling-ball viscometer, *J. Non-Newtonian Fluid Mech.*, **15**: 61, 1984.
6. D. Y. C. Fung and R. F. Matthews, *Instrumental Methods for Quality Assurance in Foods*, New York: Dekker, 1991.
7. R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*, New York: Wiley, 1960.
8. J. H. Prentice, *Measurements in the Rheology of Foodstuffs*, New York: Elsevier Applied Science, 1984.
9. A. H. P. Skelland, *Non-Newtonian Flow and Heat Transfer*, New York: Wiley, 1967.
10. I. M. Krieger and H. Elrod, Direct determination of the flow curves of non-Newtonian fluids. II. Shearing rate in the concentric cylinder viscometer, *J. Appl. Phys.*, **24**: 134–136, 1953.
11. I. M. Krieger and S. H. Maron, Direct determination of the flow curves of non-Newtonian fluids. III. Standardized treatment of viscometric data, *J. Appl. Phys.*, **25**: 72–75, 1954.
12. T. M. T. Yang and I. M. Krieger, Comparison of methods for calculating shear rates in coaxial viscometers, *J. Rheol.*, **22**: 413–421, 1978.
13. D. J. Highgate and R. W. Whorlow, End effects and particle migration effects in concentric cylinder rheometry, *Rheol. Acta*, **8** (2): 142–151, 1969.
14. E. B. Bagley, End corrections in the capillary flow of polyethylene, *J. Appl. Phys.*, **28**: 624–627, 1997.
15. B. A. Toms, in F. R. Eirich (ed.), *Rheology*, 2nd ed., New York: Academic Press, 1958.
16. A. G. Frederickson, *Principles and Applications of Rheology*. Englewood Cliffs, NJ: Prentice-Hall, 1964.
17. H. K. Kim, A. Co, and A. L. Fricke, Viscosity of black liquors by capillary measurements, *AIChE Symp. Ser.*, **77**: 207, 1981.
18. A. A. Zaman, Instructional module on techniques in rheological measurements: Fundamentals and applications, *ERC for Particle Science & Technology*, Univ. of Florida, Gainesville, FL, 1998.
19. P. Kalotay, On-line viscosity measurement using Coriolis mass flowmeters, *Flow Meas. and Instrum.*, **5**: 303–308, 1994.
20. R. C. Weast (ed.), *Handbook of Chemistry and Physics*, 51st ed., Boca Raton, FL: CRC Press, 1970.
21. R. Opie, Getting into the thick of things, *Control and Instrum.*, **20** (1): 69–71, 1988.
22. R. H. Bates, Rotational viscosity measurement applications in on-line monitoring and control, *Proc. Instrum. Soc. Amer.*, **29**: 47–56, 1996.
23. D. E. De Laney, V. Morrow, and J. Greene, Broad spectrum on-line rheology, *Proc. Soc. Plastics Eng.*, **1**: 1095–1099, 1996.
24. M. Essegir et al., Morphology and rheology sensors for in-line microstructure monitoring in compounding, *Proc. Soc. Plastics Eng.*, **3**: 3556–3560, 1996.
25. F. Baker, J. Martins, and J. Shekhtmeyster, Efficient integration boosts on-line rheometer benefits, *Modern Plastics*, **74** (6): 87–89, 1997.
26. T. O. Broadhead, W. I. Patterson, and J. M. Dealy, Closed loop viscosity control of reactive extrusion with an in-line rheometer, *Polymer Eng. and Sci.*, **36**: 2840–2851, 1996.
27. S. K. Kawatra, A. K. Bakshi, and M. T. Rusesky, Effect of slurry viscosity on hydrocyclone classification, *Int. J. Mineral Process.*, **48**: 39–50, 1996.
28. F. N. Shi and T. J. Napier-Munn, Measuring the rheology of slurries using an on-line viscometer, *Int. J. Mineral Process.*, **47**: 153–176, 1996.
29. S. K. Kawatra and A. K. Bakshi, On-line measurement of viscosity and determination of flow types for mineral suspensions, *Int. J. Mineral Process.*, **47**: 275–283, 1996.
30. A. Pillo, W. I. Patterson, and J. M. Dealy, Closed loop control of recycled HDPE crosslinking using an in-line rheometer, *Proc. Soc. Plastics Eng.*, 1228–1231, 1994.
31. R. G. Speight et al., In-line process monitoring for injection moulding control, *Proc. Inst. Mech. Eng.*, **211** (E2): 115–128, 1997.
32. C. O'Donnell, N. Herlihy, and B. McKenna, Use of an in-line viscometer in the manufacture of skim milk powder, in T. Yano, R. Matsuno, and K. Nakamura (eds.), *Developments in Food Engineering, Proc. 6th Int. Congr. Eng. Food*, New York: Blackie Academic & Professional, 1994.
33. C. O'Donnell, N. Herlihy, and B. McKenna, In-line viscosity measurement in food processing: Challenges and opportunities, *Proc. Food Processing Automation IV Conf.*, Chicago, 1995, p. 259.
34. A. L. Fricke and O. D. Crisalle, Development of viscometers for Kraft black liquor, Summary Report—Phase I, U.S. DOE Contract No. DE-FCO2-93CH10564, December 1994.
35. A. L. Fricke and O. D. Crisalle, Development of viscometers for

- Kraft black liquor, Summary Report—Phase II and IIA, U.S. DOE Contract No. DE-FCO2-93CH10564, November 1996.
36. D. B. Todd et al., Monitoring process viscosities with a new on-line rheometer, *Plastics Eng.*, **53** (3): 107–108, 1997.
 37. D. F. Arola et al., Use of nuclear magnetic resonance imaging as a viscometer for process monitoring, *Chem. Eng. Sci.*, **52**: 2049–2057, 1997.
 38. A. A. Zaman and A. L. Fricke, Viscosity of black liquors up to 140°C and 80–85% solids, *AIChE 1991 Forest Products Symp.*, 1991, pp. 59–77.
 39. A. A. Zaman and A. L. Fricke, Shear flow properties of high solids Kraft black liquors: Effects of temperature, solids concentrations, lignin molecular weight and shear rate, *Chem. Eng. Commun.*, **139**: 201–221, 1994.
 40. A. A. Zaman and A. L. Fricke, Viscoelastic properties of high solids softwood Kraft black liquors, *Ind. & Eng. Chem. Res.*, **34** (1): 382–391, 1995.
 41. A. A. Zaman and A. L. Fricke, Effects of pulping conditions and black liquor composition on the zero shear rate viscosity of softwood Kraft black liquors, *Ind. & Eng. Chem. Res.*, **35** (2): 590–597, 1996.
 42. M. M. Cross, Rheology of non-Newtonian fluids: A new flow equation for pseudoplastic systems, *J. Colloid Sci.*, **20**: 417–437, 1965.
 43. P. J. Carreau, Rheological equations from molecular network theories, *Trans. Soc. Rheol.*, **16**: 99–127, 1972.
 44. K. Y. Yasuda, Shear flow properties of concentrated solutions of linear and star branched polystyrenes, *Rheol. Acta*, **20** (2): 163–178, 1981.
 45. S. K. Kawatra and A. K. Bakshi, On-line viscometry in particulate processing, *Mineral Process. and Extractive Metallurgy Rev.*, **14**: 249–273, 1995.

V. R. BASKER
United States Air Force Research
Laboratory
DAVID W. MARTYNY
National Physical Laboratory
ARTHUR L. FRICKE
OSCAR D. CRISALLE
University of Florida

COMPASSES

Modern life is unthinkable without extensive physical movement of humans and goods around the globe. Since the advent of Sputnik in 1957, extraterrestrial space has been added.

The general problem of developing instrumentation to aid in this transfer has been recognized for thousands of years. It has led from simple observation of geographic and astronomical features to use of magnetic compasses, global positioning systems with the aid of earth-bound satellites, and to inertial guidance systems. A qualitative step from the observation of external features like the stars to an internal device is the recognition by the Chinese and the Vikings in the eleventh century that pieces of magnetized iron point in the direction of the magnetic north when allowed to orient themselves freely.

As a result of these developments, the instruments became at the same time more precise and more accurate. For example, millimeter precision can now be achieved for differential changes of position on the earth's surface by using the global positioning system (GPS) with interferometric techniques, with accuracies in the meter range, and directions can be measured

with large ring lasers with precisions reaching subarcseconds, with accuracies in the arcsecond range. Here, *accuracy* means the usual sense of difference between measured value and true value, whereas *precision* means scattering of measured values around the measured average.

One notes that the angular accuracy and precision achievable through astronomical observations at optical wavelengths with instruments mounted at sea level is limited through air turbulence to 0.1 arcsec. It can only be further reduced by special techniques.

POSITION AND DIRECTION ON EARTH, COMPASSES AND GYROS

Position and direction are in principle two very different categories, and the instrumentation required is correspondingly different.

Positioning of a Fixed Point

In positioning a fixed point, the position is to be recorded with respect to a coordinate system that should satisfy the following requirements, among others:

- Realizable and easy interpretable coordinates
- Good time stability
- High resolution or precision
- High reliability and repeatability, or accuracy

Instruments need to be developed that are ideally capable of a resolution and accuracy commensurate with the level of accuracy that the coordinate system is able to afford. These requirements are not self-evident. A nonspherical earth in the large introduces systematic errors when a purely spherical coordinate system is applied. Also ultimately, effects like the continental drift introduce errors in a time-independent system.

A variety of phenomena is at present used with positional accuracies in the meter range. On the earth's surface, starting with well-established stable surveying fix points (geodetic positions), triangulation instruments used in geodetic surveys can establish millimeter accuracy over short distances, up to kilometer distances. For larger distances, astronomical references are used. A large variety of astronomical bodies have historically served as references, starting with the sun for navigational purposes. The moon, with positional data versus time (ephemerides) of centimeter accuracy, is capable of yet higher accuracy. Among the stars, Polaris (North Star or Pole Star, the super giant Ursae Minoris of stellar magnitude +2 in the constellation Ursa Minor) describes a circle around the celestial North Pole with a radius of only 49.5 arcmin and is thus well suited to locate latitude with arcsecond accuracy. It is a preferred reference for navigation, including extraterrestrial. One notes that 1 arcsec of latitude on the earth's surface corresponds to about 30 m.

Finally, for geodetic surveys requiring submillimeter accuracy over global distances, quasars are excellent fiducial radiation sources whose positions can be considered fixed for positioning purposes within our galaxy. For geodetic studies, their radio emission is monitored with a globally distributed set of radio antennas. By study of the correlation of the radio noise of the

receiving antenna dishes, submillimeter accuracy is achieved, allowing the measurement of even continental drift, which is generally on the scale of millimeters per century. Such observations require, however, considerable expense of equipment. Quasars are suitable references mainly on account of their luminosity and their large distances from earth, up to several gigaparsecs (1 parsec = 3×10^{16} m), with corresponding very good positional stability, whose radio emission can be detected with radio dishes of sufficient size. In terms of angular resolution, coordinated radio dishes with dish diameters in the 10 m range achieve milliarcsecond accuracy as opposed to positioning of visible astronomic sources.

Direction at a Point

The establishment of a direction in a given coordinate system is a fundamentally different problem. Mathematically, position is a scalar property at a point that can be established by measuring the three positional coordinates, while the direction is a vector with three vector components. It is, however, possible to evaluate a given direction, for example, the course (clockwise angle relative to the direction true north), by accurately measuring two successive positions of it and evaluating the vector connecting the two positions; this can be done with instruments of sufficient precision (not necessarily accuracy), for example, with GPS.

The time-honored process of dead reckoning used in navigation is roughly the inverse of the process just described: knowing the starting position, the instantaneous direction of the vessel, its instantaneous speed, and the time of travel for each leg, the position of the vessel can be evaluated at any future time by a vector sum where each vector represents the direction and distance traveled with a given speed and direction at each leg.

There are a variety of physical principles available that are able to provide a reference direction independent of the position at which the measurement is performed. Such instruments may either evaluate a direction locally by differential analysis, or else supply an internally generated reference direction that is independent of time and motion of the carrier on which this instrument is mounted. Naturally, great care has to be exercised to render such an instrument insensitive to any rotation of the carrier vessel.

The discussion of such instruments is the main concern of this article.

A variety of physical effects have been employed in directional devices.

Magnetic Compasses. These instruments use the ability of a magnetized needle to align itself in the earth's magnetic field. Since in the present era the horizontal components of the magnetic induction of this field point with reasonable accuracy to a north geomagnetic pole that is close to the (geographic) North Pole, the direction of such a needle is, at least at latitudes far away from the polar regions, reasonably aligned with a meridian on the earth surface, with deviations (declinations, see glossary at the end of this article) generally amounting to less than 10° .

Internally Defined Direction References. For instruments of this kind, any physical principle that establishes a direction in space that is constant is usable. A practically important class

of instruments employ mechanical gyros. They work with the principle that the rotation axis of a rotating massive body maintains its direction under any translation if no torque is applied to it.

Electronic and Nuclear References. Besides electronic interferometers, a recent version of nuclear interferometers is the use of nuclei that have an angular momentum or spin associated with them. When techniques are employed that overcome the randomizing thermal effect such that sufficient nuclei can be aligned in a probe, a readout mechanism will then show a constant direction of such an assembly. Such instruments are called *nuclear magnetic resonance gyroscopes*.

Ring Lasers. These lasers operate with two or more beams, that is, photons, that circulate in opposite directions around a given area defined by three or more mirrors or, more commonly, by an optical fiber. The thus defined area can be represented by an area vector that provide the desired reference direction. The mechanism of excitation of the beams is in principle irrelevant. Besides the classic excitation via inversion through dc current with internal electrodes, high-frequency excitation is used with external electrodes, but modern approaches employ Brillouin stimulation and other methods as well (1).

Any rotation around the area vector or axis results in an output frequency that is proportional to the rate of rotation or the angular velocity. This principle can be used to define directions in various ways. A direct way is its use as a north finder, where a ring laser mounted on the surface of the rotating earth is tilted until its area vector is parallel to the earth rotation axis, which is then verified by measuring a maximum output rate of the ring laser. This vector then points to the geographic North Pole.

In a "strapped-down" version, the ring laser is clamped on the carrier. It records then any rotational component of it. Thus, given the original direction of the ring-laser axis, the integrated output is then used to evaluate the new direction. Since one ring laser can only measure angular velocity components around its own axis, one needs three such rings to completely evaluate the direction vector. Commercial devices with one to three rings in one instrument are on the market, including tiny gyro sensors for image stabilization in binoculars (2), inertial navigation systems in modern airliners, submarines, and missiles, and for establishing stable platforms (3).

DIRECTION FINDING IN THE EARTH'S MAGNETIC FIELD, ERROR SOURCES, AND GYRO OPERATION

Direction Finding via the Earth's Magnetic Field

The Earth's Magnetic Field. The magnetic flux density $\mathbf{B}$ of the earth's field is a vector field whose flux lines define a direction anywhere within a few earth radii (the average earth radius is 6371 km). Its direction can be sensed by placing a freely movable compass needle into it. The torque exerted by the earth's field on the compass dipole becomes zero when the dipole is aligned with the field at that location. Observations show that the earth's field itself can be described to a good approximation by a dipole (Fig. 1) with magnetic moment $m_E = 8.0052 \times 10^{22}$ A·m² (4). This dipole is fixed at the earth's center and its axis is tilted by about 11° against the earth's axis (1996) (the number in parentheses is the year in which these

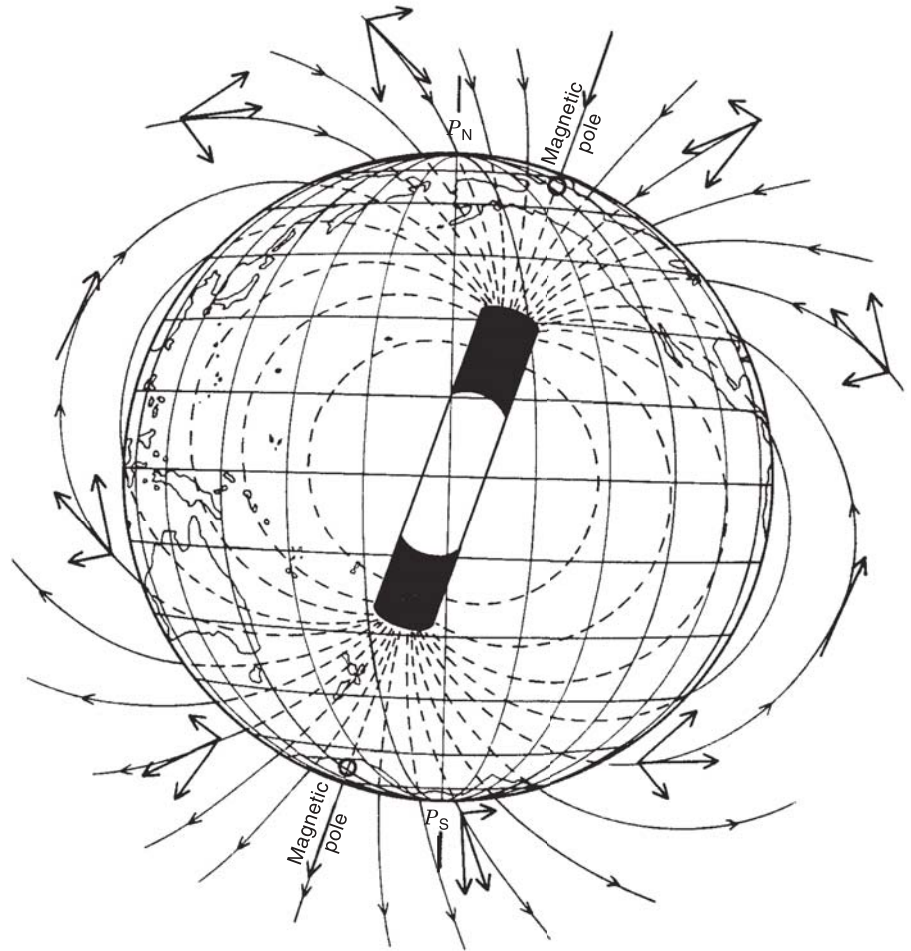


Figure 1. The earth's magnetic field. Dipole approximation. The vertical axis is the earth's kinematic axis, with poles P_N and P_S . [Source: Maloney, Ref. 7, p. 62.]

coordinates were observed). An aligned needle whose movement is restricted such that it only shows the component parallel to the earth's surface will then point to the coordinates 79° North, 105° West (1996), which is called the north magnetic pole. The magnetic poles, by virtue of the fact that the magnetic flux lines at these points are normal to the earth's surface, constitute also perfect portholes for entry of charged particles from space (5). The following comments qualify the statements above:

1. There is no theory of earth magnetism that is able to predict the field's magnitude, direction, or time evolution sufficiently for navigational purposes.
2. The dipole approximation is good but a closer description of the actual earth field in terms of associated Legendre polynomials has several hundred terms in addition to the basic dipole term (4). The justification of using the dipole approximation stems from the fact that the dipole term in such an expansion is at least an order of magnitude larger than any subsequent term. The actual field distribution on the earth's surface is shown in a Mercator projection in Fig. 2.
3. The south magnetic pole is not at the homologous point in the southern hemisphere; it is at 65° South, 138° East (1996).

4. The nature of the dipole-dipole interaction is such that opposite poles attract each other. In order to keep the notion that the magnetic pole closest to the geographic north pole is called north magnetic pole, the end of the compass needle that points north would have to be the needle's South Pole.
5. Local disturbances due to the presence of, man-made or other, magnetic depositions measurably distort the local field.
6. The field is slowly time dependent, both in magnitude and in direction. Daily changes exist. Figure 3 shows a measurement with a diurnal peak-to-peak excursion of 40 nT (6). These changes include a daily movement of the magnetic poles in elliptical paths with a major axis of about 50 miles (7).

The yearly change in magnitude is of the order of a few tens of nanoteslas depending on the location.

Paleomagnetic studies show that the field is known to have reversed itself approximately once per one million years (7a). The movement of the north magnetic pole during the last 430 years is shown in Fig. 4 (4). The data are, however, not as smooth as the figure implies, as an estimate of the position in 1360 is 66° N, 291° E (69° W) (8). In recognition of this movement, modern atlases indicate only a "North magnetic pole area" (1998: near Bathurst Island, Canada).

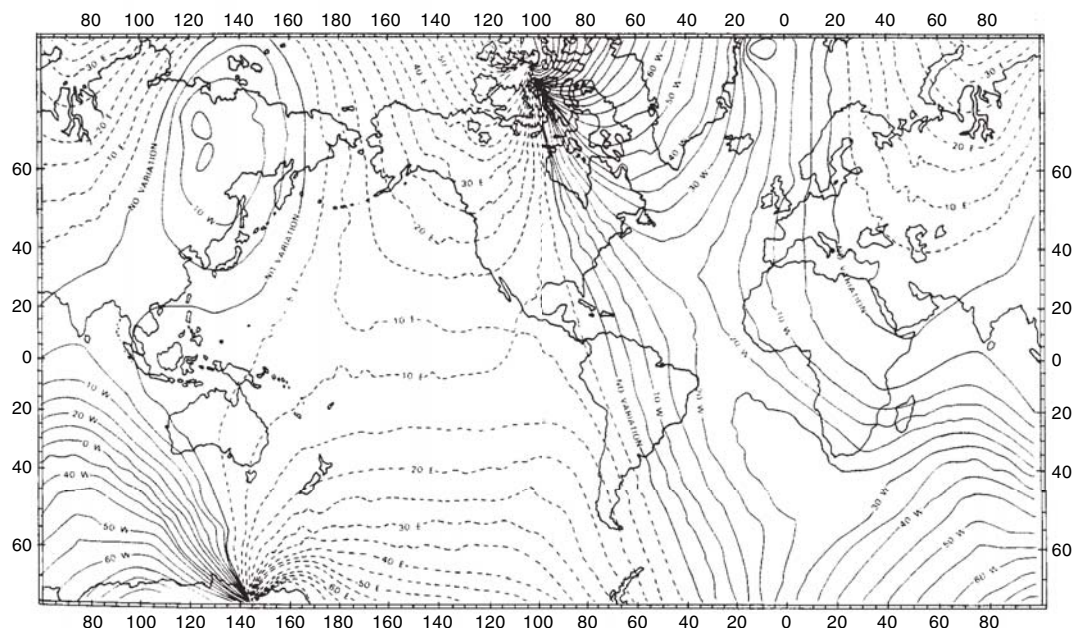


Figure 2. Deviation of horizontal compass directions from geographic South–North directions, or Variation. The poles in northern Canada and in Antarctica are clearly visible. [Source: Maloney, Ref. 7, p. 65.]

7. Magnetic storms add additional time dependence. Their magnitudes may range up to $5 \mu\text{T}$ (6).

From detailed studies of the earth's magnetic field it can be concluded that the accuracy of compasses is limited to about $\pm 0.5^\circ$, due to the lack of stability of the earth's field, local disturbances, and due to the difficulty of eliminating man-made fields in the vessel. This error is achievable with a compass whose vicinity has small residual stray fields that are properly compensated for, and whose declination is updated at regular intervals (at least annually).

It follows that a compass mounted in precision instruments, such as surveyor's instruments, that are capable of ± 20 arcsec resolution can only give cursory initial guidance. Also, since

0.5° corresponds to about 55 km on the earth's surface, it is impossible to direct, for example, a vessel with subkilometer accuracy on the globe using compasses alone. Meter accuracy of positioning can be achieved with GPS, and arcsecond accuracy of direction is achieved with inertial guidance systems, for example, with ring-laser gyros (RLG).

Analysis and Reduction of Disturbing Fields from Sources other than the Earth's Field. Given that the earth's field has at present an average magnitude of $53 \mu\text{T}$, but man-made saturation fields of permanent magnets reach magnitudes of 1 T, the necessity of elimination of or compensation for fields other than the earth's field is an important consideration during installation and use of compasses.

Any extraneous ac fields can easily be removed by enclosure in a nonmagnetic metallic screen. This problem is, however, of lesser importance because the compass that is to sense the dc earth field is usually made with a large time constant of several seconds, thereby integrating to zero any ac fields with frequencies well above 1 Hz, for example, 50 Hz or 60 Hz fields generated by nearby power lines.

Time Constant. The large time constant for the needle is usually obtained by immersion in a liquid. While this method is very convenient, it is nevertheless not optimal from the point of view of minimizing the mechanical fluctuations or the mechanical noise. The latter is achieved by introducing feedback via a lossless mechanism, which, however, introduces an effective damping into the needle movement and thus lowers the noise temperature of the needle. Approaches of this kind are well known in electrical engineering but not in connection with magnetic compasses. The reason for this lack of development is probably the inherent, much larger limitation of precision due to the fluctuations of the earth's field that make further development of the precision of magnetic compasses redundant for any but the shortest observation times.

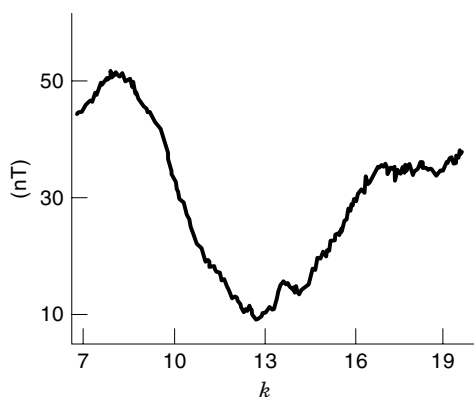


Figure 3. A measurement of a daily variation of the earth's magnetic field strength measured by nuclear magnetic resonance at Jussy, Switzerland, on 11 March 1959. A peak-to-peak change of 40 nT is apparent. [Source: Hochstrasser, Ref. 6, p. 232.]

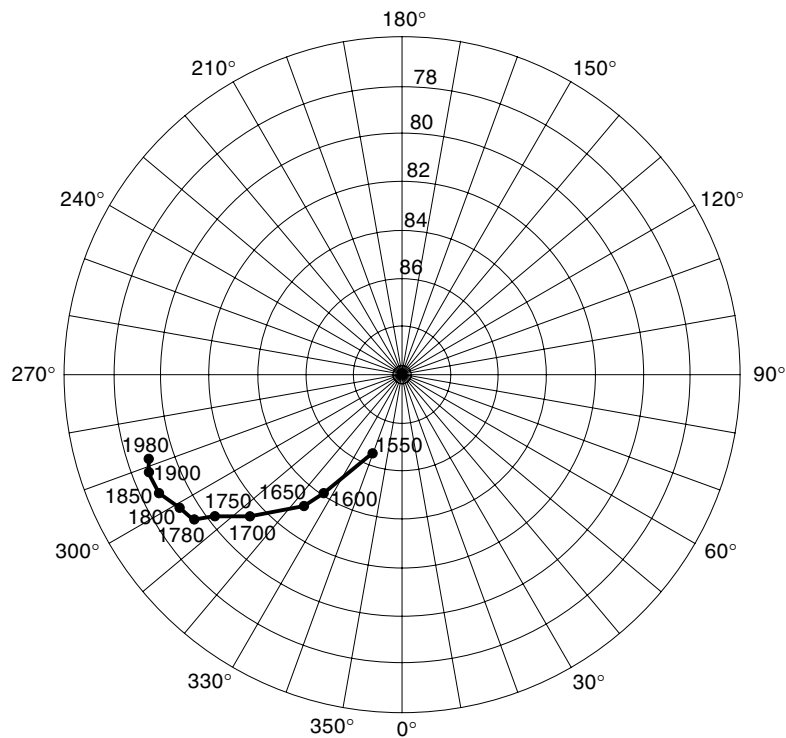


Figure 4. Movement of the geomagnetic North Pole between 1550 and 1980. The movement covers about 7 degrees of latitude and 50 degrees of longitude. Corresponding yearly variations depend on the location on earth. [Source: Reprinted from Ref. 4, p. 389, by permission of the publisher Academic Press Limited, London.]

Magnetic Stray Fields. The compensation of extraneous dc fields can become a formidable problem. There are two simple rules used to assess such fields' magnitudes.

***B* field produced by an electric dc current:** "One ampère produces 200 nT at one meter distance." This rule is derived from a current flowing in an infinite straight line. This field decreases with distance R as R^{-2} .

***B* field produced by ferromagnetic material in the earth's field:** "One gram of iron produces 1 nT at 1 m distance." This rule is derived from the dipole field of a ferromagnet induced by the earth's field. The strength of a dipole field decreases with distance R as R^{-3} . The basis of this rule was first espoused in Ref. 9, p. 139.

The disturbances are such that the field due to the iron mass of a typical automobile or a ship needs to be carefully compensated for if a magnetic compass is to be used in such an environment.

The two rules are heeded to the extreme in high-resolution nuclear magnetic spectra in the earth's field. This method to obtain an accurate mapping of the earth's local magnetic field vector requires a housing of the apparatus that has to be set up at least 100 m away from power lines, and all ferromagnetic materials, including glass and bronze nails with traces of iron in them, have to be avoided in the construction of such sites (6).

Conversely, these rules require careful selection of the site of a compass on a vessel with a typical iron mass of the order of 10,000 tons. The second rule previously given suggests 10 μ T at 10 m distance, of the same order as the horizontal field component (15 μ T). Furthermore, the induced dipole moment of a ship depends not only on the ship's position on the globe but also on its alignment with the local field. Additionally, many types of ferromagnetic material may be present with vastly different nonlinearities, saturation, and remanences. This is dealt

with approximately by distinguishing between a permanent part and a variable part of the disturbing field.

Methods of Compensation

Helmholtz Coils. A rigorous approach is to set up three orthogonal pairs of Helmholtz coils (10) whose currents are controlled by sensors of the three spatial components of the stray field. A pair of Helmholtz coils, with an axial separation equal to the coil radius, produces a rather homogeneous **B** field in the symmetry center with a drop-off at the fringes varying with the fourth power of the distance and is therefore an excellent, precisely adjustable tool. As such it is the preferred tool where compensation down to picoteslas or less are required as in the production of magnetoencephalograms, where the fields to be measured themselves are at the picotesla level. For regular compass applications in the field, this solution is too spacious and too expensive.

Compensating Permanent Magnets (11). The horizontal component of the disturbing field can be split into two orthogonal components. A straightforward method of compensation is therefore to place two small permanent magnets at 90° off the needle such that after any rotation of the vehicle with respect to the direction North, the needle shows in the same direction, that is, North plus the proper declination. The installation and adjustment of such compensating magnets require a proper empirical approach, since the direction and magnitude of the stray field are generally not known.

Flinders Bars (12). This is an ingenious method to compensate for the induced variable component of the magnetic dipole moment of a large vessel. One or more pieces of soft iron (with low remanence) are placed near the compass, such that their induced moment compensates the ship's magnetic moment. Since the field of the ship decreases with R^{-3} , relatively small pieces

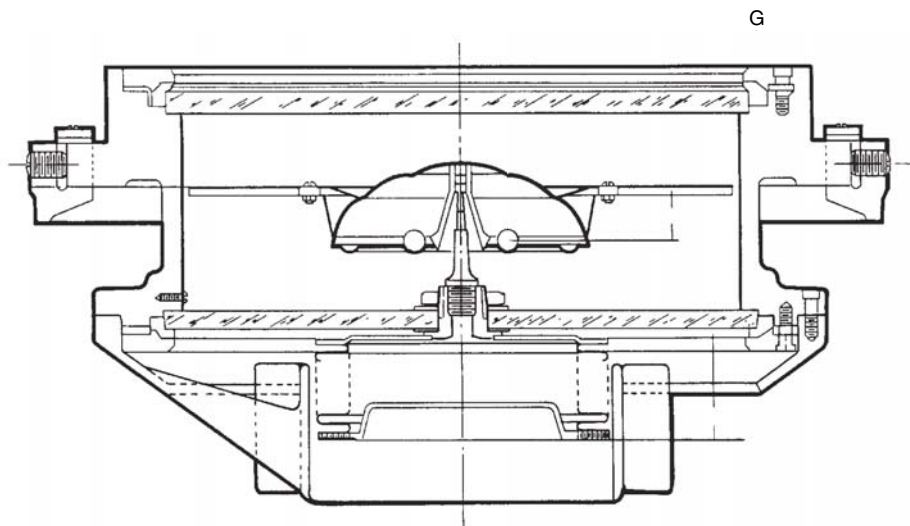


Figure 5. Schematic of a magnetic compass for a ship. The compass card (B) is immersed in a bowl (C) filled with liquid (D) and an airbubble (E). The bowl is centered on a vertical axis, a pivot, which lets the card's magnets stay oriented toward North when the ship changes direction. There is an expansion bellows (F), and a bezel ring (G) to secure the cover. See also text.

in close vicinity of the compass suffice; furthermore, any rotation of the ship or positioning of it at a different magnitude of the earth's field induces a proportional amount into ship and Flinders bars, and thus keeps the compensation valid.

Details for the use and adjustment of compasses are given in the excellent book by Maloney (7).

Degaussing Coils. These coils are used to neutralize the external field of the vessel, for example, to avoid detection by an adversary. Typically, however, the internal field increases as a consequence, and therefore the compensation has to be adjusted depending on whether the degaussing field is switched on or off.

Description of a Practical Magnetic Compass

A widely used instrument is the US Navy Standard No. 1 7.5 inch compass; see Fig. 5 (Ref. 7, p. 67). The housing (C) is made from nonmagnetic material. Two bar magnets (A) of a nickel-cobalt-iron alloy are fixed and aligned with magnetic north on a compass card (B). A ring magnet may also be used. The card is in a bowl filled with alcohol or varsol (D) to provide a passive dampening with a time constant of a few seconds. A mark, the lubber's line, at the compass rim indicates the ship's direction. The graduation on the compass card opposite to the lubber's line indicates the ship's course. Correction pieces of soft iron with negligible remanence and hard iron with a permanent dipole moment are placed in certain configurations outside the compass. The goal of initial correction is to compensate not only for the ship's horizontal dipole moment at various headings of the ship but also to an extent for the vertical magnetic moment of the ship and the earth's magnetic field, since generally only the horizontal component is used for navigation.

Astronomical Direction Finding

A very convenient fact is the position of Polaris close to the celestial north pole, which is equivalent to the direction of the kinematic earth axis. Although Polaris is by no means fixed (13), its motion is known to better than arcsecond accuracy, and for determination of positions or directions of such accuracy, the *American Ephemeris and Nautical Almanac* (14) can be used to correct the observation. The Almanac is updated every year.

An obvious disadvantage of this mode of direction finding is the restriction to nights and cloudless skies.

Gyros

A radical departure from the use of external aids is the use of instruments based on phenomena that provide a stable reference direction independent of external parameters. Gyros can be mechanical, based on nuclear spin, electron spin, or macroscopic quantum systems with an angular momentum given by circulating photons (ring lasers) (15), electrons (16) including superconducting rings, atoms (atomic ring interferometers) (17), or in particular helium atoms (superfluid rings) (18). All gyros possess an internally defined direction, given either by the direction of the angular momentum or spin $\mathbf{M}$ of the spinning particle or quantum fluid, or by the area vector that is circumscribed by the moving particle. For spinning massive particles the time evolution in the absence of a torque $\mathbf{T}$ is described by

$$d\mathbf{M}/dt = \mathbf{T} = \mathbf{0} \quad (1)$$

that is, the vector $\mathbf{M}$ has then a constant direction. In particular, its direction stays constant under any movement of the vessel on which this system is placed if one ensures that the movement does not impart any torque to the system. For mechanical gyros, a Cardan suspension or similar gimbaling structures are designed to effect the decoupling of motions of the carrier from the gyro system; see Fig. 6.

Readout. There must be an engineering solution to the readout problem. For gyros the position of the rotation axis has to be read out with an accuracy that the particular gyro system is capable of. The interaction of the readout and gyro has to be minimized to avoid any backlash that affects the gyro axis. An equivalent problem exists in optical gyros where the required interference of the optical beams at the output may produce backscattering into the cavity, which in severe cases synchronizes the sensing beams inside and thus renders the device useless for interrogation of its rotation status or as a compass.

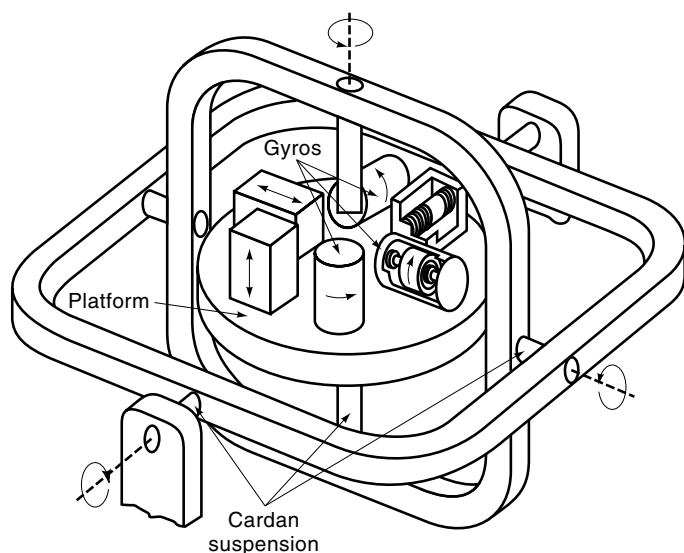


Figure 6. Schematic of an inertial navigation system with mechanical gyros and acceleration sensors on a navigational platform which is placed in a Cardan suspension. The platform keeps its attitude in space while the vessel on which it is mounted rotates freely in all directions. [Source: Der Spiegel, No. 10, 1984, p. 214.]

In modern gyros, a viable system has to have a resolution of 1 arcsec or better. A submarine, to get to its berth after a long time of inertial navigation without external references, needs arcsecond accuracy.

Of the many possible internal directional references, mechanical gyros with various types of suspension and readout capabilities as well as ring-laser gyros and fiber-optic gyros have been developed to engineering standards. These devices are all capable of arcsecond resolution, as far as the principle of operation and the resolution of the readout is concerned. In mechanical gyros, the rotor of a motor constitutes the spinning part; in this case, the coupling to the outside world is basically through the stator's rotating magnetic field, with no friction except for the bearing of the rotor itself. High accuracy gyros employ air bearings.

Superconducting Mechanical Gyro (19). An extreme case of development is the Gravity Probe B where a satellite houses a mechanical gyro which is to measure a secular change of 44 marcsec per year with an error of ± 1 marcsec per year. The superconducting rotor is electrostatically suspended in a vacuum, and the readout is optical. The development of this gyro has so far lasted for about three decades, but the previous specifications are now secured (1997).

Superfluid Gyros. They are based on the rotation of superfluid ^3He or ^4He in a closed circuit. They are in the development stage. Their basic resolution is theoretically far superior to ring lasers, and more so to classical mechanical gyros, but laboratory results (1997), while encouraging, show nevertheless that much work needs to be done to approach the anticipated performance limits.

Ring-Laser Gyros and Fiber-Optic Gyros. The counterrotating beams in an optical cavity form typically a triangle or a square

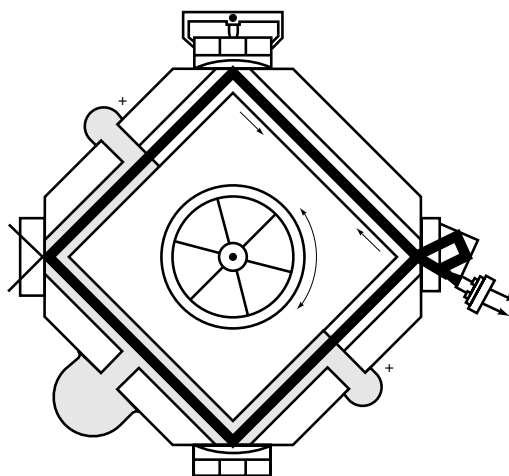


Figure 7. Schematic of a plane square ring-laser gyro with dc excitation. The interferometric readout is visible on the right of the figure. The mirror on top is moved in and out by a piezoelectric device, to compensate for pathlength changes. A dithering motor is mounted in the center to prevent the beams from locking at small rotation rates. [Source: G. T. Martin, *IEEE Spectrum*, Feb. 1986, p. 50. © 1986 IEEE.]

(Fig. 7). The beams develop a phase difference as soon as the cavity is rotated around its area vector in absolute space. If the rotation in three dimensions needs to be known, a three-gyro system as in Fig. 8 is employed. In interferometer fiber-optic gyros (IFOGs, Fig. 9), this phase difference is measured by special readout techniques (20).

In self-oscillating loops, the requirement of phase closure forces a shift of the optical oscillation frequency, which is opposite for beams in opposite directions. The corresponding frequency difference Δf between the beams is then given by the Sagnac formula (15)

$$\Delta f = (1/2\pi)d\phi_o/dt = (4/\lambda L)\mathbf{A} \cdot \boldsymbol{\Omega} = (4/\lambda L)A(d\phi_r/dt) \cos \alpha \quad (2)$$

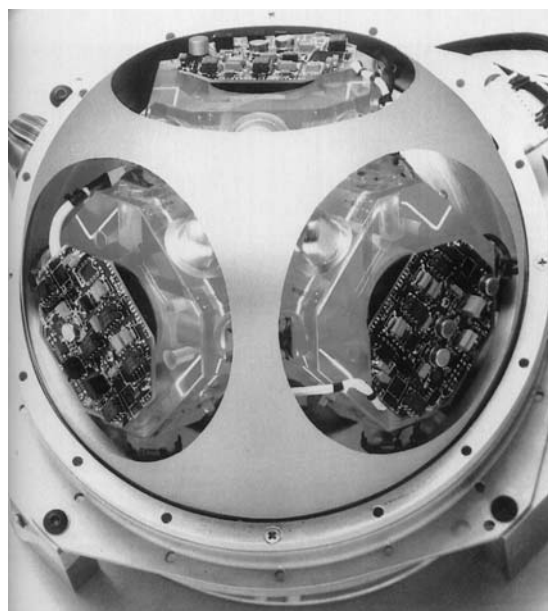


Figure 8. Three-gyro mount to sense three-dimensional rotation. [Source: G. T. Martin, *IEEE Spectrum*, Feb. 1986, p. 49. © 1986 IEEE.]

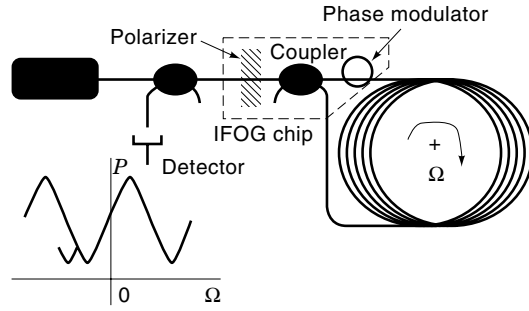


Figure 9. Principle of an interferometer fiber-optic gyro (IFOG). The fiber-optic loops on the right are excited by an external laser on the left whose beam is split into the two counterrotating beams by the coupler. A second coupler, immediately at the laser output, picks up the returning beams and feeds them into the detector where a phase difference is detected. [Source: Hotate, Ref. 3, p. 108.]

where $\mathbf{A} = (1/2)\oint \mathbf{r} \times d\mathbf{l}$ is the area enclosed by the beam path, represented by its normal vector, Ω is the angular velocity vector to which the device is subjected, α is the angle between the two vectors, λ is the vacuum wavelength of the laser light, and L is the perimeter $L = \oint dl$, or the length of the cavity. This general definition of $\mathbf{A}$ and of L includes also nonplanar RLG configurations as in the quadrilateral Raytheon device (21), for example. After integration of Eq. (2), the optical phase ϕ_o of the RLG output becomes proportional to the mechanical angle of rotation, ϕ_r , if the direction of Ω stays constant.

Certain gyros (21) work with four beams, with pairs of counterrotating beams. In those cases, the beat frequency doubles.

Scale Factor. This denotes the sensitivity of the transduction from the mechanical rotation rate to the optical frequency, or conversely the ratio of equivalent electronic phase change of the ring-laser output frequency to the mechanical rotation angle of the device. It is defined as (22)

$$S = 4A/(\lambda L)[\text{Hz}/(\text{rad/s})] \quad (3)$$

In normal usage, the scale factor is, however, given as

$$S' = [4A/(\lambda L)](\pi/180)(1/3600) \quad (\text{counts}/\text{arcsec}) \quad (4)$$

A typical RLG operating at $\lambda = 633 \text{ nm}$, with an area $A = 1 \text{ dm}^2$ and a length $L = 40 \text{ cm}$ then has a scale factor $S' = 0.76 \text{ counts}/\text{arcsec}$.

Sensitivity to Attitude. The scalar product $\mathbf{A} \cdot \Omega = A\Omega \cos \alpha$ in the Sagnac Eq. (2) gives rise to a sensitivity to the tilt angle α ,

$$|d\Delta f/d\alpha| = S \frac{d\phi_r}{dt} \sin \alpha \quad (5)$$

Equation (5) is important for North-finding ring-laser gyros, where the device is oriented on the rotating earth until the output is maximized, or with equivalent strategies. The uncertainty $d\alpha$ is a measure of the achievable error of the RLG as an indicator of the direction North, that is, the direction of the kinetic axis of earth rotation. Since $d\alpha \propto 1/\sin \alpha$, better sensitivity is basically obtained by finding the direction(s) orthogonal to North.

Limits of Accuracy of Ring-Laser Gyros. Contrary to mechanical gyros, ring-laser gyros possess no rotating mass. Errors due to torque, Eq. (1), that give rise to a variety of effects in

mechanical gyros are therefore absent. Absent also are basic time constants when the ring laser changes its rotation rate or its attitude. Equally, no spin-up time is required to initialize the sensor. In practice, the electronics surrounding the gyro cause warm-up times that are typically in the millisecond range.

Basic quantum limits have been reached, which allow, for example, one to measure a rotation as slow as the earth's with very high precision.

Quantum Noise and Optical $1/f$ Noise. The photon statistics in a laser beam produce a frequency fluctuation whose white one-sided power spectral density $S_{\Delta f}$ is given by

$$S_{\Delta f} = hf_0^3/(PQ^2) \quad (6)$$

where h is Planck's constant, $= 6.6 \times 10^{-34} \text{ W}\cdot\text{s}^2$, f_0 is the average optical oscillation frequency (Hz), P is the optical power lost by the beam in the cavity, including the output (W), and Q is the passive quality factor of the cavity in which the beam is created.

For a standard RLG operating with two independent counterrotating light beams, the beat frequency Δf has twice this noise power density, provided that the beams are uncorrelated and each has the same optical power P . In a four-frequency gyro, the noise power quadruples, but there is a net gain in the signal-to-noise ratio of the latter, because the doubled output frequency [see Eq. (2)] is accompanied by a root-mean-square frequency fluctuation that is only $\sqrt{2}$ times larger.

In ring-laser systems a $1/f$ noise component has also been observed empirically (23). Its power spectral density is

$$S_{\Delta f, 1/\nu} \cong 4(f_0^2/Q^4)(1/\nu) \quad (7)$$

where ν is the Fourier frequency of the oscillator. The inverse quartic power dependence of Q suggests, however, only a small contribution to optical cavities with large Q 's as in RLGs.

Time Averaging. When averaging the beat frequency between two beams over a time duration T , the average has an rms error (24)

$$\overline{\Delta f}_{rms} = [hf_0^3/(PQ^2)]^{1/2}(1/T)^{1/2} \quad (8)$$

The error of the average frequency is thus reduced proportional to $1/\sqrt{T}$. The stability of the system with time sets a useful upper limit of measuring time and with it the ultimate limit of accuracy.

An RLG as mentioned before, with a quality factor $Q = 1 \times 10^9$, a power output of $100 \mu\text{W}$, and a measurement time of 1 min has an rms frequency fluctuation of 3.4 mHz from Eq. (8). The equivalent mechanical phase excursion due to rotation follows from S' above as 4.5 marcsec in 1 min of observation time.

As a North finder [Eq. (5)] the compass error of the same device, again with 1 min of observation, is about 6 arcsec on the earth, if measured with $\alpha = 90^\circ$.

RLGs are so far superior to IFOGs on account of their larger quality factors achievable.

Optical Frequency. The beat frequency versus rotation rate, or the scale factor, can be increased by increasing the optical frequency, that is, lowering the wavelength. Gas-ring lasers are using a helium-neon mixture, approximately 7:1, whereby the helium may be actually the isotope ^3He , which achieves a better conversion of input power to optical power than the naturally occurring ^4He , and an approximately equal mixture

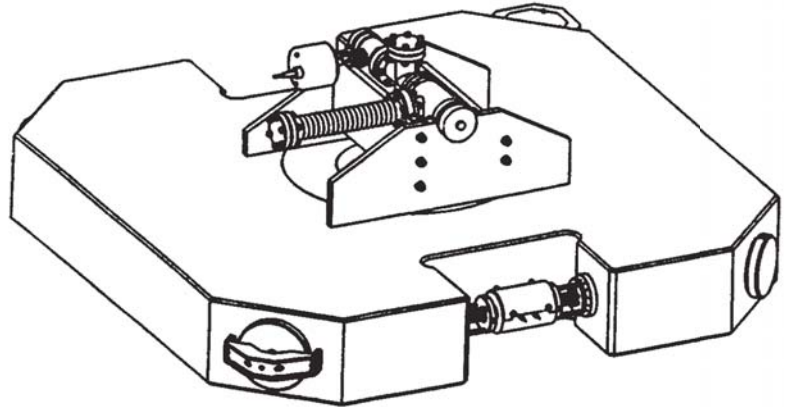


Figure 10. Schematic of a high-resolution research ring laser for detection of fluctuations of the earth rotation. The main differences to RLGs are the large size with a 1 m² area mounted in a stiff 600 kg Zerodur block, a very-high-*Q* cavity, and RF excitation.

of the two naturally occurring neon isotopes ²⁰Ne and ²²Ne to minimize the cross coupling between the two counterrotating beams. Other lasing gases with larger optical frequencies have been found to be inadequate for RLGs.

The optical cavity needs to be excited in one and only one mode clockwise (CW) as well as counterclockwise (CCW). This limits the power, whose upper limit is given by the onset of multimode excitation. Fiber-optic rings are operated at wavelengths in the infrared. The preferred wavelength is 1.55 μm with a fortuitous coincidence of minimum absorption in the fused silica fiber used and europium-doped fibers that can be used as in-line amplifiers.

Area and Path Length. For a given geometry of the ring the area is proportional to L^2 , and it follows from Eq. (2) that larger-area rings are more sensitive. For usual navigational applications decimeter-sized rings are quite sufficient, mainly due to the large quality factor [Eq. (2)] that can be achieved with high-quality mirrors. The latter have reflection losses approaching 1 ppm (1997).

Fiber-optic rings have quality factors that are several orders of magnitude smaller. The effective area is therefore increased by making spools of many turns. With n complete turns the effective area is increased n -fold. The output frequency (or the equivalent phase shift) is then also increased n -fold.

Optimum RLG Geometry. Of all possible planar and nonplanar polygons, plane regular polygons have the largest area to perimeter ratio and therefore the largest scale factor [see Eq. (3)]. Furthermore, given a fixed circumscribed circle for them and employing mirrors with a fixed finesse, the optimum geometry of a gas-ring laser is a square. With this geometry, the gain in the area to perimeter ratio overcompensates the additional loss when going from an equilateral triangle to a square. Square rings also have additional advantages (15). However, in practice triangular-shaped as well as square-shaped RLGs are used.

These geometric considerations do not apply to fiber-optic rings. There the loss mechanisms are distributed in the fiber, and the total loss and therefore the noise power are roughly proportional to the total length L of the fiber. The signal is also proportional to the length; thus the beat frequency to rms frequency fluctuation ratio generally increases in fiber-optic rings with $\sqrt{L}$.

Practical RLG. See Fig. 8. The beam path is encased in a material with a very low coefficient of thermal expansion; at this time (1997) the two materials ultralow expansion (ULE) quartz and Zerodur, with thermal expansion coefficients approaching 0.01 ppm/K, are used. Compared to these artificial materials,

the best-known naturally occurring low-expansion material is fused silica with 0.55 ppm/K, which is, however, inadequate, since the amount of electronics required to keep a single optical cavity mode tuned to the laser gain medium at even modest ambient temperature changes becomes prohibitive.

The maturity of these devices is underlined by the fact that as early as in the mid-1970s a “laser navigation system” was planned for the Sojourner that landed on Mars on 4 July 1997 (25).

Figure 10 shows a high-resolution ring laser specifically constructed for geodetic research (26). The beams circumscribe a plane square area of 1 m². It is designed to resolve the earth’s rotation rate to about one part in 10⁸, which comes close to the known fluctuations of the earth’s rotation vector Ω . Recent developments (June 2000) include a 4 m × 4 m RLG on a Zerodur base with a diameter of 4.2 m, to be housed underground at Wettzell, Germany, for geodetic research.

MORE DETAILED ANALYSIS OF THE EARTH’S MAGNETIC FIELD, COORDINATE SYSTEMS, AND NEW DEVELOPMENTS

Some Details about the Earth Magnetic Field

Units. The SI system is used here as it is throughout this encyclopedia. In order to facilitate the reading of older literature, three hard-to-eradicate non-SI units are translated.

$$\text{One gauss} = 1 \text{ G} = 0.0001 \text{ T} = 100 \mu\text{T}$$

$$\text{One gamma} = 1 \gamma = 1 \text{ nT}$$

$$\text{One oersted} = 1 \text{ Oe} = 1000/4\pi \text{ A/m} = 79.6 \text{ A/m}$$

The Magnetic Dipole Approximation. The magnetic flux density $\mathbf{B}$ of a dipole at far field is given in coordinate-free form by (10)

$$\mathbf{B} = [\mu_0/(4\pi R^3)] \left(3 \frac{\mathbf{R} \cdot \mathbf{m}}{R^2} \mathbf{R} - \mathbf{m} \right) \quad (9)$$

whereby the dipole with magnetic dipole moment $\mathbf{m}$ is placed at $\mathbf{R} = 0$. The far field of $\mathbf{B}$ is then given at the vector distance $\mathbf{R}$.

With the equivalent magnetic moment at the earth’s center pointing in the direction of magnetic north, the far field at the earth surface is given by

$$\mathbf{B} = [\mu_0 m/(4\pi R_E^3)] (\mathbf{a}_R 2 \sin \phi_m - \mathbf{a}_{\phi_m} \cos \phi_m) \quad (10)$$

and its magnitude by

$$B = [\mu_0 m / (4\pi R_E^2)] \sqrt{1 + 3 \sin^2 \phi_m} \quad (11)$$

Here, R_E is the average earth radius, $m = m_E$ is the magnitude of the equivalent dipole approximating the earth's field, ϕ_m is the magnetic latitude (positive for northern latitudes, negative for southern latitudes), $\mathbf{a}_R$ is a unit vector at the field point normal to and pointing away from the earth surface, and $\mathbf{a}_{\theta m}$ is a unit vector pointing to magnetic north. The magnitude near the north magnetic pole ($\phi_m = 90^\circ$) is $62 \mu\text{T}$, whereas at the magnetic equator ($\phi_m = 0^\circ$) it is $31 \mu\text{T}$. The average field at the earth surface is $53 \mu\text{T}$ but the average horizontal component of it is only $15 \mu\text{T}$.

The curl of any $\mathbf{B}$ field is always 0.

Coordinate Systems. In the literature, generally three coordinate systems are in use.

Right-Handed Spherical Coordinate System r, θ, ϕ . Here the radius r_s is measured from the earth's center, the polar angle is measured starting at the (geographic) North Pole with $\theta_s = 0^\circ$ and ending at the (geographic) South Pole with $\theta_s = 180^\circ$, and the azimuth ϕ_s is measured from a reference meridian (Greenwich) onwards east for a full circle, or 360° .

Geographic-Navigational Coordinate System for the Earth Surface ϕ, λ . Here the latitude ϕ is counted from the equator on, positive towards North, negative towards South. The longitude (meridian) λ is counted from Greenwich on positive towards West and negative towards East, so that the meridian opposite the Greenwich meridian has both assignments, $\pm 180^\circ$. For most purposes of navigation, the earth's surface can be considered as a perfect sphere except for local features like mountains and valleys. This coordinate system is not right-handed.

Coordinate Transformations. The piecewise linear transformations between the geographic-navigational and spherical coordinate systems can be written with the aid of the signum function $\text{sgn}(x)$ [$\text{sgn}(x = 0) = 0$, $\text{sgn}(x > 0) = 1$, $\text{sgn}(x < 0) = -1$] as

$$\begin{aligned} \phi_s &= -\lambda + \text{sgn}(\lambda)[\text{sgn}(\lambda) + 1]180^\circ \\ \lambda &= -\phi_s + \text{sgn}(\phi_s - 180^\circ)[\text{sgn}(\phi_s - 180^\circ) + 1]180^\circ \\ \theta_s &= 90^\circ - \phi \\ \phi &= 90^\circ - \theta_s \end{aligned} \quad (12)$$

Magnetic Coordinate System ϕ_m, λ_m . This system is similar to the geographic-navigational system except that the magnetic latitudes $\phi_m = \pm 90^\circ$ are defined by the geomagnetic poles.

The local angle between a magnetic meridian and the corresponding local geographic meridian is the declination D (degrees).

Because of local and temporal changes of the direction of the magnetic field that may amount up to about 0.5° even for observations over one day and/or over short distances (6), the magnetic coordinates, including the most important declination, are not reliable when better accuracies are required. The compass mounted on a level or theodolite for surveying purposes is exclusively used for cursory orientation in the field, not for measurements, considering that these instruments are generally capable of performing at the subarcminute level for civil engineering tasks.

Calculation of Declination. Given the spherical coordinates of the north geomagnetic pole $\theta_{s,Nm}$ and $\phi_{s,Nm}$, a location X with coordinates $\theta_{s,X}$ and $\phi_{s,X}$, and the dipole approximation of the earth's field, the declination D at X becomes, through application of spherical-trigonometric equations,

$$D = \cos^{-1} \left[\frac{\cos \theta_{s,X} - \cos \theta_{s,Nm} \cos \phi_{m,X}}{\sin \theta_{s,Nm} \sin \phi_{m,X}} \right] \quad (13)$$

$$\begin{aligned} \phi_{m,X} &= \cos^{-1} [\cos \theta_{s,Nm} \cos \theta_{s,X} \\ &\quad + \sin \theta_{s,Nm} \sin \theta_{s,X} \cos(\phi_{s,X} - \phi_{s,Nm})] \end{aligned}$$

where $\phi_{m,X}$ is the great circle distance from the north geomagnetic pole to the location X expressed as an arc, in degrees, as seen from earth's center. Care has to be exercised in applications of these formulas, as the relation of the spherical coordinates to the navigational coordinates is only piecewise linear.

Distortion of the Earth's Magnetic Field by Ferromagnetic Materials. The magnetic dipole moment $\mathbf{m}$ of a piece of material with relative permeability μ_r and volume V that is placed in a field $\mathbf{H}$ is estimated by

$$\mathbf{m} = (\mu_r - 1)V\mathbf{H} \quad (14)$$

One kilogram of iron with a mass density of 8580 kg/m^3 placed in a magnetic field $H = 50 \mu\text{T}/(\mu_0) = 40 \text{ A/m}$ with a relative permeability of about 2000 satisfies the rule for distortion by magnetic dipoles given previously.

Size of Torque. Given the magnetic dipole moment $\mathbf{m}$, the torque exerted on this moment in a field $\mathbf{B}$ is

$$\mathbf{T} = \mathbf{m} \times \mathbf{B} \quad (15)$$

The torque on a compass needle is small. With the magnetic moment of a needle of $2 \text{ A}\cdot\text{m}^2$, the torque corresponds to that of a 10 mW electric motor running at 1000 rpm . Here the relation $P = T\omega$ is used, where P is the power generated (W), and ω is the angular frequency (rad/s).

Some Details about Ring Lasers and Other Gyros. A ring laser is an example of an ordered spin system with correlation lengths that may extend to megameters depending on the quality factor of the optical cavity in which the light beam is generated and on the power output. It is one of a very few macroscopic quantum systems. This one uses photons; others are superfluid systems using liquid ^3He or ^4He and electronic superconductors using electrons (Cooper pairs), and magnetic systems (quantum Hall effect). Gyros with low-noise electronic beams are also under active investigation. Of these novel principles, only the RLG and fiber-optic gyros are at this time engineered into mature systems. A recent addition to these tools are atomic interferometers in ring arrangements.

All these systems have in common a closed path enclosing a finite area around which the quantum fluid is circulating. In ring lasers this circulation is internally generated. In other systems the quantum fluid is injected by an external source whereby with specific devices (beam splitters or generally flux splitters), two different paths are created. At the point of

rejoining there is a detection system that detects the phase difference, which is then a measure of the rotation rate.

Frequency Difference and Phase Difference in Optical Gyros; Sense of Rotation. The general problem of obtaining an output signal that is linearly related to the rotation of the device is solved as follows. Given a frequency split of the counterrotating beams $f_{1,2} = f_0 \pm \Delta f/2$ during a rotation, with a corresponding split in the wave numbers $k_{1,2} = k_0 \mp \Delta k/2$, and phase splits after rotation of $\phi_{0,1,2} = \phi_{00} \pm \Delta\phi_0/2$. The split may have a slow time dependence. The two optical beams are overlaid to each other, say in the x direction. The rapidly oscillating electric fields of the beams, with frequencies on the order of 10^{14} Hz to 10^{15} Hz and splits ranging up to 10^6 Hz, having equal amplitudes E_0 , are then

$$E_{1,2} = E_0 \exp[j(2\pi f_{1,2}t - k_{1,2}x)] \quad (16)$$

The observed output power of the combined beams is then

$$P_{out} \propto |E_1 + E_2|^2 = 2E_0^2[\cos(2\pi\Delta f t - 2\pi\Delta kx + \Delta\phi_0) + 1] \quad (17)$$

Observation at $x = 0$ results in a cosinusoidal pattern with frequency Δf (ring lasers) and phase $\Delta\phi$ (nonoscillating Sagnac interferometers) superimposed on a background of constant power. The time dependence is then detected by a low-frequency optical ac detector, typically by a silicon junction device. If the direction of rotation is desired, the beams are slightly misaligned to produce a moving interferometric fringe pattern. Two detectors are then placed a quarter-fringe apart. Their phase differences versus time determine the direction of rotation uniquely.

It also follows from the extreme smallness of the Sagnac frequency relative to the optical frequency that the beam path must be extraordinarily symmetric in clockwise and anticlockwise direction. Any nonreciprocity will lead to an additional non-Sagnac effect, either an additional phase or a frequency bias depending on the nature of the nonreciprocity.

Modifications of Optical Beams. Two new types of optical beams are investigated: One investigation centers on the possibility of correlated emission of the amplifying plasma with the goal of creating beams whose quantum noise is correlated, so that the noise of the interfering beams can be greatly reduced (27). Another approach is to use squeezed light (28), where the quantum noise of the amplitude fluctuations is greatly increased in order to substantially reduce the frequency fluctuation, which keeps the Heisenberg uncertainty relation intact, but affords RLGs with much lower (frequency) noise.

Polarization-Preserving Fiber. When the beams are guided by optical fibers, the original polarization may not be maintained if irregularities are present along the fiber. Special fibers have been manufactured with a slightly elliptical cross section that enforces the same polarization throughout. This is necessary for several reasons, one of the most important is the necessity to have the polarizations of the two beams aligned when they are brought to interference.

Quality Factor Q . This factor enters with its inverse square in the noise and is therefore an important factor in the sensitivity of a gyro, whether as an interferometer or as a self-oscillating device.

In ring lasers with an amplifying gas at low pressure of the order of 1 torr, the overall quality factor Q of the optical

cavity is mainly determined by the losses of the mirrors (29). Given a square cavity with four mirrors, each with a relative power loss $1 - R$, where R is the power reflectivity of a mirror, $Q = (\pi L/2\lambda)/(1 - R)$. In a square research ring laser (Fig. 10) with good mirrors approaching $1 - R = 1$ ppm each, $L = 4$ m, and $\lambda = 633$ nm, a quality factor of 9.9×10^{12} can be achieved (26). Q increases linearly with the perimeter of the cavity, as the losses are localized.

In fiber-optic gyros, the losses are distributed along the fiber. With a total loss of α dB/m (1000α dB/km), the quality factor becomes independent of the path length and is $Q = (2\pi/\lambda)(10/\ln 10)(1/\alpha)$. A good fiber with $\alpha = 0.001$ dB/m (1 dB/km) at $\lambda = 1.55$ μ m gives rise to $Q = 1.8 \times 10^{10}$, almost three orders of magnitude less and independent of the length of the cavity.

GLOSSARY

Agonic line. A line on the earth's surface that connects points with zero magnetic variation.

Cardinal directions. North, East, South, and West.

Compass error. Algebraic sum of variation and deviation.

Declination. Variation.

Deviation. Errors caused in the compass direction by local effects.

Dip. Inclination of a perfectly balanced compass needle (inclination or dip circle) with respect to the local horizontal.

Isogonic line. A line on the earth's surface that connects points with equal magnetic variation.

Point. Subdivision of the compass circle into 32 points. 1 point = 11.25° .

Variation. Angle, in degrees, between the magnetic meridian and the corresponding geographic meridian. It is the direction of the horizontal component of the magnetic field with respect to the geographic coordinate system.

BIBLIOGRAPHY

1. A. E. Siegman, *Lasers*, Mill Valley, CA: Univ. Sci. Books, 1986.
2. *Newsweek*, Feb. 2, p. 12, 1998.
3. K. Hotate, Fiber optic gyros, *Photonics Spectra*, 108–112, April 1997.
4. J. A. Jacobs (ed.), *Geomagnetism*, vols. 1–4, London: Academic Press, 1987.
5. T. Feder, Congress chills hopes for polar cap observatory, *Phys. Today*, 53–54, July 1998.
6. G. Hochstrasser, Détermination de formes et de largeurs de raies très fines en résonance magnétique nucléaire, *Helv. Phys. Acta* **34**: 189–239, 1961.
7. E. S. Maloney, *Dutton's Navigation and Piloting*, 13th ed., Annapolis, MD: Naval Institute Press, 1981.
- 7a. H. Breuer, *dtv-Atlas zur Physik*, Bd.2, 3rd ed. München, Germany: Deutscher Taschenbuch Verlag, 1993, p245.
8. F. Jueneman, Magnetic north by northwest, *R&D Mag.*, 13, September, 1997.
9. S. Chapman and J. Bartels, *Geomagnetism*, vol. I, Oxford, UK: Clarendon Press, reprint 1951.
10. D. K. Cheng, *Field and Wave Electromagnetics*, 2nd ed., Reading, MA: Addison-Wesley, 1992.

11. Defense Mapping Agency Hydrographic/Topographic Center, *Handbook of Magnetic Compass Adjustment and Compensation*, DMAHTC Publication No. 226.
12. Director, Defense Mapping Agency Hydrographic Center, Washington, DC, *American Practical Navigator*, vol. I–II.
13. W. M. Smart, *Textbook on Spherical Astronomy*, 6th ed., revised by R. M. Green, Cambridge, UK: Cambridge Univ. Press, reprinted 1979.
14. J. C. Smith and F. G. Smith, *The American Ephemeris and Nautical Almanac for the year 1980*, Washington DC: U.S. Government Printing Office, 1979.
15. R. Anderson, H. R. Bilger, and G. E. Stedman, “Sagnac” effect: A century of earth-rotated interferometers, *Am. J. Phys.*, **62**: 975–985, 1994.
16. F. Hasselbach and M. Nicklaus, Sagnac experiment with electrons: Observation of the rotational phase shift of electrons in vacuum, *Phys. Rev. A*, **48**: 143–151, 1993.
17. P. Bouyer, T. L. Gustavson, and M. A. Kasevich, Development of an atom interferometer gyroscope, presented at Cairns, Australia, July, 1996.
18. R. E. Packard and S. Vitale, Principles of superfluid-helium gyroscopes, *Phys. Rev. B*, **46**: 3540–3549, 1992.
19. Anonymous, Orbiting gyro test of general relativity, *Phys. Today*, 20–22, May 1984.
20. B. Culshaw and J. Dakin (eds.), *Optical Fiber Sensors*, vol. IV, Norwood, MA: Artech House, 1997.
21. H. Statz et al., The multioscillator ring laser gyroscope, in M. L. Stitch and M. Bass (eds.), *Laser Handbook*, vol. 4, Amsterdam: North-Holland, 1985, pp. 229–332.
22. G. E. Stedman et al., Harmonic analysis in a large ring laser with backscatter-induced pulling, *Phys. Rev. A*, **51**: 4944–4958, 1995.
23. M. R. Sayeh and H. R. Bilger, Flicker noise in frequency fluctuations of lasers, *Phys. Rev. Lett.*, **55**, 700–702, 1985.
24. H. Gerhardt, H. Welling, and A. Güttner, Measurements of the laser linewidth due to quantum phase and quantum amplitude noise above and below threshold, *Z. Phys.*, **253**: 113–126, 1972.
25. R. Jayawardhana, Mars pathfinder, *Muse* **1** (4): 32–33, 1977.
26. H. R. Bilger, U. Schreiber, and G. E. Stedman, Design and application of large perimeter ring lasers, *Symp. Gyro Technol.* 1996, Stuttgart, Germany, October 1996.
27. W. Schleich, M. O. Scully, and H. G. von Garssen, Quantum-noise quenching in the correlated spontaneous-emission laser as a multiplicative noise process, part I, A geometrical argument, *Phys. Rev. A*, **37**: 1261–1269, 1988; part II, Rigorous analysis including amplitude noise, *Phys. Rev. A*, **37**: 3010–3017, 1988.
28. M. A. M. Marte and D. F. Walls, Enhanced sensitivity of fiber-optic rotation sensors with squeezed light, *J. Opt. Soc. Am.*, **B4**: 1849–1852, 1987.
29. J. T. Verdeyen, *Laser Electronics*, 3rd ed., Upper Saddle River, NJ: Prentice-Hall, 1995.
- Director, Defense Mapping Agency Hydrographic Center, Washington DC, *American Practical Navigator*, vols. I–II. The 1977 edition is published by the Defense Mapping Agency Hydrographic Center, Washington, DC 1977. The first edition was published in 1802. These volumes are generally known as “Bowditch.”
- H. Gerhardt, H. Welling, and A. Güttner, Measurements of the laser linewidth due to quantum phase and quantum amplitude noise above and below threshold, *Z. Phys.*, **253**: 113–126, 1972. Equation (8) had been proposed already by Townes, except for a factor $\sqrt{2}$.
- G. Hochstrasser, Détermination de formes et de largeurs de raies très fines en résonance magnétique nucléaire, *Helv. Phys. Acta*, **34**: 189–239, 1961. This paper shows that with the nuclear magnetic resonance technique, the earth’s field can be measured at the earth’s surface down into the natural fluctuations of it, which are of the order of magnitude of 0.1 nT for short-term fluctuations, but can be as large as several tens of nanotesla over one day.
- J. A. Jacobs (ed.), *Geomagnetism*, vol. 1, London: Academic, 1987.
- R. T. Merrill, M. W. McElhinny, and P. L. McFadden, *The Magnetic Field of the Earth: Paleomagnetism, the Core, and the Deep Mantle*, San Diego, CA: Academic Press, 1996.
- S. P. Parker (ed.), *McGraw-Hill Dictionary of Scientific and Technical Terms*, 5th ed., New York: McGraw-Hill, 1994. This dictionary has been used throughout this article for spelling and definition of scientific and technical terms.
- J. C. Smith and F. G. Smith, *The American Ephemeris and Nautical Almanac for the year 1980*, Washington DC: U.S. Government Printing Office, 1979. This is an annually issued volume since the original British edition of 1767. The data are predicted for each coming year. As the somewhat random announcements of leap seconds, up to twice annually, shows, the times in this Coordinated Universal Time System (UTC) are uncertain at a level of a fractional second, and the corresponding angular uncertainties are uncertain by up to several arcseconds. In the Atomic Time System (TAI), this type of error is greatly reduced.
- H. Statz et al., The multioscillator ring laser gyroscope, in M. L. Stitch and M. Bass (eds.), *Laser Handbook*, vol. 4, Amsterdam: North-Holland, 1985, pp. 229–332.

H. R. BILGER
Oklahoma State University

Reading List

- R. Anderson, H. R. Bilger, and G. E. Stedman, “Sagnac” effect: A century of earth-rotated interferometers, *Am. J. Phys.*, **62**: 975–985, 1994.
- S. Chapman and J. Bartels, *Geomagnetism*, vol. I, Oxford, UK: Clarendon Press, reprint 1951.
- B. Culshaw and J. Dakin (eds.), *Optical Fiber Sensors*, vol. IV, Norwood, MA: Artech House, 1997.
- Defense Mapping Agency Hydrographic/Topographic Center, *Hand-*

GONIOMETERS

Goniometry is the science of measuring angles. The instruments used to perform this task are called goniometers. They provide an accurate reference system for angular-dependent measurements in a wide field of applications. In their most elementary form, goniometers have existed for many hundreds of years. Recently, they have assumed an extra role as powerful high-precision and computer-controlled systems capable of positioning an infinite variety of emitters, detectors, and targets in any combination and permutation of geometrical configurations. Their field of application has widened as the need for more precise information from an ever-increasing number of high-precision emitters and detectors has grown and, above all, because the large amount of information generated in the process can now be rapidly and automatically acquired, stored, processed, and analyzed. From a simple hand-held device for the measurement of angles, the goniometer has also become the heart of many of the most sophisticated and intelligent analytical tools available to mankind. Goniometers today find

numerous new applications in research and industry, from the traditional fields of medicine and materials analyses to the more recent fields of remote sensing and space applications. At the same time, they remain the basic element in a large number of precise mechanical positioning systems.

HISTORICAL BACKGROUND

Considering the fact that traditional navigational instruments such as astrolabes, quadrants, octants, and sextants are all examples of goniometers, we can see that the goniometer is one of the oldest scientific instruments known to mankind. The original astrolabe was a hand-held astronomical instrument used by ancient Greeks and others to measure the height above the horizon of celestial bodies. Another version of this was the marine astrolabe, a simple wooden or brass graduated disk used for measuring the angle of the sun or a star from the horizon in order to obtain one's position in terms of degrees of latitude above the equator. The quadrant, a graduated quarter circle with a suspended plumb bob, was another example of a goniometer used to assist in determining latitude by alignment with the sun or a star, typically the Pole Star.

An improved version of the quadrant was Hadley's octant in 1731, which incorporated the use of a mirror to bring the object viewed into coincidence with another. From the inclination of the mirror, a measurement of the angle between the two objects could be obtained. The octant was in time superseded by an improved goniometric instrument, the sextant, which along with the compass has remained the basis of marine navigation until recent times.

In other sciences, such as mineralogy, improvements were also being made to the goniometer. One of the first references to be found on the subject is that of Carangeot's contact goniometer (1), which was used by mineralogists to compare large crystals of different development by measuring the interfacial angles. It consisted of two flat steel bars pivoted together like a pair of scissors. The angle between the bars was read from a graduated semicircle with a precision of about a half degree. The contact goniometer was mainly used for large dull crystals that did not yield reflections. For the study of the morphology of smaller crystals, Wollaston's reflecting goniometer in 1809 was a considerable improvement on the contact goniometer, enabling more precise measurements to be taken. This was a simple, cheap, and portable instrument that affected the interpretation of the structure of crystals and had a great influence on the science of mineralogy.

This type of goniometer was followed by the two-circle or theodolite goniometer in 1876, which was a combination of the vertical- and horizontal-circle types. This eliminated the need to mount and readjust the crystal for each series of measurements.

Another important event in the history of goniometers and their role in crystallography took place in 1912 when von Laue showed that diffraction of X rays occurred when they were passed through a suitably oriented crystal. At the same time, he effectively established that X rays have very short wavelengths and that crystals have their atoms arranged in a regular structure. This was followed shortly by the experiments of Bragg in 1912, which gave further important insights into the nature of X rays. These pioneering experiments led to the development of the X-ray goniometer, which used X rays instead of visible light

as the source and an ionization chamber instead of a telescope as the detector and which was designed to measure angles between the atomic planes within crystals.

Reflectivity measurements of different kinds of surfaces have been made for various purposes in the field of astronomy from the turn of the nineteenth century using goniometric systems. The earlier studies mainly concerned the overall photometric behavior of materials. These measurements are still important today because they were made usually in a more systematic way than more recent ones.

As a result of satellite navigation systems, the role of one type of goniometer—the sextant—has lost some importance, although it still remains an essential marine navigational instrument. Thanks to earth-observation satellites, a new type of goniometer has been developed for remote sensing applications during the last few decades because of the need for a more complete characterization on the angular reflectance characteristics of natural and man-made surfaces. This is partially achieved today with the help of custom-built and generally large-scale goniometers capable of observing relatively large target areas, covering one to several tens of square meters.

APPLICATION FIELDS

Today, goniometers are found in a large number of traditional and new application fields. In navigation they are intensively used for direction finding. The direction of a radio transmitter can be rapidly and precisely obtained by measuring the angle of maximum intensity of the transmitted signal with respect to a fixed point. This technique is used in the electronic goniometer or direction-finding receiver, which can be mounted in a fixed or mobile station and is used to measure the direction of propagation of electromagnetic waves. The electronic goniometer can take a variety of forms but is a basic element in most direction-finding antenna array systems in navigation.

In surveying, the theodolite, a very accurate goniometric instrument, allows the precise measurement of angles, both vertically and horizontally. It is usually mounted on an accurately graduated horizontal circle and consists of a telescope and an alidade with a graduated vertical circle. The optical theodolite uses horizontal and vertical circles graduated on glass. Another version of the instrument used in terrestrial photogrammetry is the photo-theodolite consisting of a combined theodolite and camera. The electronic tachymeter or tacheometer is another example of a goniometer mainly used in large-scale topographic surveys and designed for the rapid determination from a single observation of horizontal angles and zenith distances. Trigonometric heighting, electronic distance measurement, traversing, contouring, and detail surveys are just some of the applications of electronic tacheometry.

Goniometers have also many medical uses. In anthropology and surgery, goniometric devices are used to measure the various angles of the skull and other bones. These are often in the form of hand-held goniometers used to measure the range of motion and the angle of joints in the human body (e.g., to monitor knee joint angle or movements of fingers about knuckle joints). Measurements of these angles also provide important information for analysis of changes in posture and movement of the human body during space flight. In certain eye diseases, it is necessary to measure the anterior chamber angle formed by the iris with respect to the rear face of the cornea.

This is performed by means of another goniometric device—the gonioscope.

In industry, goniometers have innumerable applications. To analyze the light distributions of different compact fluorescent lamps, for example, goniophotometers with automated control and data acquisition systems are used. These measure the angular distribution of light emanating from various lighting fixtures. Goniometers have been used in projection photometry where the angular characteristics of search lights and projectors (e.g., cone angles) are measured and analyzed and for the photometric calibration of light sources including the infrared region (2). Goniophotometric measurements are used to quantify light interactions with coated surfaces (e.g., to study the reflectance of metallic finishes in the automobile industry and to improve colorant formulation and shading). In the paper industry, goniometers have been used to provide the necessary scattering characteristics of cellulose fibers and pigments used in papers. Goniometers are also used in the test and calibration of angular dividing equipment. In general, they have helped to develop a better physical understanding of the complex problem of color and the appearance of coatings in industry.

In military use, goniometric compasses have been extensively employed for topographic measurements and for aiming artillery fire.

In science, goniometer systems play a vital role in a large number of disciplines. Three representative examples (advanced material science, astronomy, and remote sensing) are presented in some detail here.

ADVANCED MATERIAL SCIENCE

X-ray diffraction is probably the best known and most appropriate method of determining the crystalline structure of solid materials. It is used for a variety of structures, and even though it is particularly simple in the case of well-crystallized inorganic solids as a “fingerprinting” method for phase analysis, it can also be applied to a wide variety of more complicated organic molecules and to nanocrystalline or amorphous materials. In addition to the crystalline structure, X-ray diffraction can reveal information on stress, crystallite size distribution, preferred orientation, and texture.

Highly accurate goniometers are an essential part of X-ray diffraction systems (Fig. 1). The most common experimental

configuration for polycrystalline materials consists of an X ray beam passing precisely over the axis of a perfectly concentric dual goniometer system, each goniometer having a precision of better than 0.001° , and in many cases, especially in research, some orders of magnitude better than this. The sample is then moved by an angle θ , and the detector by precisely 2θ . The Bragg condition for diffraction peaks to appear is

$$n\lambda = 2d \sin \theta$$

where n is an integer and d is an interplanar spacing present in the material. The intensity of the peaks depends on several factors and in some cases may be zero. The diffracted intensity plotted against 2θ in a *diffractogram* may then be analyzed either by reference to a database of fingerprint spectra or, in the case of new crystalline structures, by comparison to simulated spectra generated by appropriate computer programs using postulated structures. Of course, there are many other experimental arrangements (e.g., for single-crystal analysis or for stress determination) and methods of analyzing or treating diffraction spectra to extract the appropriate information.

At the Institute for Advanced Materials of the Joint Research Centre, Ispra, Italy, a special system has been constructed for the structural analysis of very thin films. Instead of both goniometers being scanned during a measurement, the incident beam impinges on the sample at a very low incident angle, typically from 0.2° to 1.0° . Only the detector (2θ) goniometer is scanned, and special optics ensure that a good angular resolution is maintained. This glancing angle geometry renders the diffraction measurement much more sensitive to the surface of the specimen under examination. Additionally, in order to optimize the signal-to-noise ratio, the distances from x ray source to sample and from sample to detector have been minimized, a high-precision variable slit system to define the incident beam dimensions is used, and a solid-state detector is employed to isolate the desired wavelength without the use of filters or a monochromator and to reduce the background count rate to a minimum. The system may also be used for x ray reflectivity measurements, which can provide highly accurate information about the thickness of thin films, as well as their density and interfacial roughness, by monitoring the specularly reflected X ray intensity as a function of angle of incidence (3).

Many versions of horizontal and vertical goniometers with microprocessor control units have been designed for X ray powder diffraction measurements, phase analysis, and stress and texture analysis. Stand-alone operation of the diffractometers is often possible, thereby reducing the risk of human error. Automated sample spinning is also implemented in some systems to compensate for surface effects and nonrandom crystal orientations.

ASTRONOMY

Various kinds of goniometers have been devised by astronomers mainly for the measurement of the angular reflectance of planetary surfaces. Reflectance measurements of natural surfaces were not started in earnest until the 1960s with the advent of space exploration. Toward the end of the 19th century, however, Kononoviz (4) had started to measure the *surface reflectance* (i.e., the ratio of the light scattered from a surface to the light incident on the surface) of various surfaces on the Earth using



Figure 1. An X-ray diffraction goniometer at Joint Research Centre, Ispra, Italy.

a large range of angles. Ångström (5) also made a series of measurements on natural surfaces with various angles of incidence and reflectance. The measured surface materials were mainly different surface types of water, sand, and vegetation.

Albedo measurements were also accomplished later by Kulebakin (6) and Kimball and Hand (7). The main purpose of the studies was to establish a measure (usually the reflectivity) for different kinds of scattering media. Kulebakin made systematic measurements of natural and artificial surfaces using a rudimentary form of goniometer. His study focused on the simulation of the overall roughness of natural surfaces such as snow, sand, and vegetation, and some man-made surfaces, which were made from cardboard. Kulebakin's work was the first attempt to accomplish the parametrization of the scattering phenomena of surface materials from laboratory measurements. In addition, Kimball and Hand demonstrated for the first time the importance of surface roughness on the backscattering of light properties. They also included in their study natural surfaces such as water and snow. Measurements were made with the help of an airplane and the sun was used as the illumination source.

The most important studies of light scattering by surface materials were accomplished by Oetking (8) and Hapke and van Horn (9), who made extensive goniometric measurements of snow, rocks, and different kinds of powders. These works were the basis of the theoretical explanation of the scattering properties of the surface of the Moon. The study of Hapke and van Horn concentrated mainly on the Moon's photometric properties, whereas Oetking made systematic photometry of various kinds of powders with a selection of particle sizes. Oetking focused on the comparison of laboratory measurements and the reflectance of different parts of the Moon. His goniometer was capable of measuring phase angles of less than 1° , but operated in the principal plane only. The device was table mounted, with a constantly positioned 75 W Zenon arc lamp as light source and the detector, a photomultiplier, at the end of a rod. At the other end was a corotating sample tray with a mirror above it inclined at 45° . Because the detector was placed under another semitransparent mirror at 45° , it was also possible to measure samples at zero phase angle. Oetking measured extensively phase curves of different types of rocks, aluminum oxide, magnesium oxide, magnesium carbonate, small spheres, and even granulated sugar. The particle sizes were accurately measured, making this set of measurements an excellent example of a controlled and repeatable experiment.

The measurements of snow surfaces by Knowles Middleton and Mungall (10) have been important in explaining the photometric results of the bright and icy objects of the solar system. The aim of this study was to investigate the specular reflection of different kinds of snow surfaces. The smallest detectable angle was 5° . Thus the opposition spike could not be detected. They concluded that below an angle of incidence of 45° diffuse reflection dominates for most of the snow surface. With angles greater than that, the specular reflection is more profound.

Instruments for these types of goniometric measurements were usually made according to the purpose of the study so there was no standard way to build the instrumentation. This is still the case, as we shall see later in the section dealing with the remote sensing applications of goniometers. Sometimes, the instruments were portable [e.g., Knowles Middleton and Mungall, (10)], whereas others were fixed in the

laboratory. The original aim of the studies may have been rather restricted, and so the instruments are far from multipurpose. Also for most instruments, the financing seems to have been rather low, and therefore they were generally suited only for restricted purposes in a limited time.

Goniometric measurements by van Diggelen (11) were of great importance in understanding the reflection of the particulate surface of the Moon. The aim of the measurements was to compare the reflectivity of the Moon to terrestrial volcanic ash samples.

Egan (12) made extensive goniometric measurements of porous materials. The main purpose of his study was to understand the polarimetric properties of materials, particularly from Mars. The importance of his work is in establishing the albedo-polarization rule of the dark particulate surfaces, which is a useful tool in determining the radiometric sizes of atmosphereless objects. The method is based on the dependence of a negative polarization of the surface porosity of the object. This relationship is one of the major methods to determine the albedos of asteroids. The goniometer designed by Egan used two fixed photometers at viewing angles of 0° and 60° . The collimated light source could be rotated. The phase angles were limited to a range of 40° to 130° . Egan used tilted mirrors to produce polarization of light.

Other researchers have measured the polarization of a variety of targets in the principal plane using a goniometer with 1.2 m arm and a rotating Glan-prism to have an effective way to measure the degree of polarization (13). These basic studies of planetary materials have been successfully continued using a goniometer designed at Deutsche Forschungsanstalt für Luft- und Raumfahrt (DLR) of Berlin, Germany (14), which was used mainly to support space missions such as NASA's *Mars Pathfinder*.

Another interesting device used to measure the backscattering range of small phase angles is a type of microscopic goniometer (15). The measurements were used as a reference during the introduction of the coherent backscattering mechanism. The goniometer applied can reach all possible angles including the backscattering direction. This device has been deployed for the comparison of the scattering of bright materials with the observations of bright (icy) solar system moons. Recently, Piironen et al. (16) have published a series of goniometric measurements of snow at very small phase angles. The results show that the degree of backscattering depends on the type and amount of impurities in the snow.

Small particle goniometers have been widely used for basic research in the Free University of Amsterdam (17). Their instrument is based in the flowing particle stream and can measure the full Muller matrix (18). The Muller matrix explains the conditions of light scattering including polarization and directional information and is an extension of the Stokes four-element vector to a 4×4 matrix. Stokes parameters are used to describe the polarization of electromagnetic radiation. The activity has been recently concentrated on the study of micron- and submicron-sized particles of silt and algae in water. Another ingenious goniometer for small particle scattering measurements is designed by Sasse in DLR, Stuttgart, Germany (Fig. 2). The device uses fiber optics to measure the full half circle of scattering angles at the same time of a single particle inside an electrostatic levitator (19). This instrument has been used for scattering measurements of fluidized bed particles for improved solar power plants and meteoritic

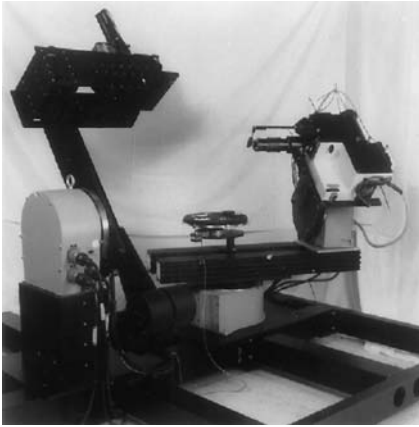


Figure 2. A small particle goniometer at DLR, Stuttgart, Germany.

particles in the size range from a few tens to a few hundreds of micrometers.

REMOTE SENSING

Most of the earth's surfaces reflect varying amounts of electromagnetic radiation in different directions. Unless a surface is reflecting isotropically, any change of the viewing or illumination directions has an impact on the spectral reflectance signature in remote sensing imagery. Varying illumination angles are caused by changes in the latitudinal coordinates in a global remote sensing data set or by seasonal and diurnal effects in a multitemporal approach. Viewing angle geometry is influenced by topographical effects, by the extent of the sensor's field-of-view, or by actively changing the view direction in pointable sensors. To make correct measurements of the total reflected flux, the earth's surface must be observed from many directions using air- or satellite-borne detectors. The directional reflectance characteristics or the amount of solar radiation scattered from a surface in a given direction provides important information about the surface characteristics. Many earth-observation satellite instruments today look only vertically downward at the earth's surface, but in the near future a highly increased amount of data from sensors with off-nadir capability will become available primarily from NASA's spaceborne sensor MISR (20). In order to apply remote sensing data to land use change and ecologically relevant studies within NASA's Earth Science Enterprise and other programs, bidirectional ground reference data must be widely available. A wide variety of bidirectional reflectance models have already been designed to use the multidirectional information effectively with these remote sensing data. However, there is still a lack of bidirectional ground reference data to adequately validate the remotely sensed data and the various models.

Reflectance, BRF, and BRDF

Bidirectional reflectance factor (BRF) and bidirectional reflectance distribution function (BRDF) are the key parameters addressed in goniometric measurements of reflected radiance. The bidirectional reflectance factor R is the ratio of radiant flux $d\Phi_r$ reflected from a target to the flux $d\Phi_{rid}$ reflected from a lossless isotropic (Lambertian) surface for a given

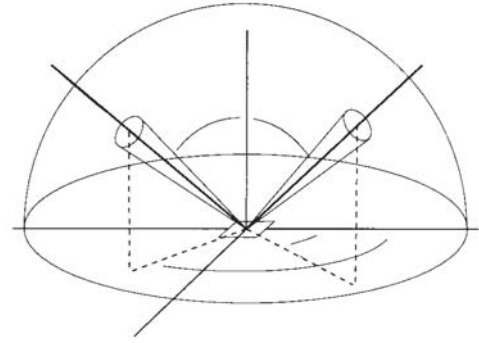


Figure 3. Concept and parameters of BRDF.

illumination-viewing geometry $(\theta_i, \phi_i; \theta_r, \phi_r)$ and wavelength (λ) :

$$R(\theta_i, \phi_i; \theta_r, \phi_r; \lambda) = d\Phi_r / d\Phi_{rid} \quad (\text{dimensionless})$$

The bidirectional reflectance distribution function f_r as defined by Nicodemus et al. (21) is the fraction of the radiance $L[W \cdot m^{-2} \cdot sr^{-1} \cdot nm^{-1}]$ of the incident irradiance $E_i[W \cdot m^{-2} \cdot nm^{-1}]$ from direction θ_i, ϕ_i reflected into a specific direction θ_r, ϕ_r (Fig. 3):

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r; \lambda) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r; \lambda)}{dE_i(\theta_i, \phi_i; \lambda)} \quad (sr^{-1})$$

Assuming isotropic irradiance and BRDF within designated solid angles, R and f_r are interrelated by

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r; \lambda) = \frac{R(\theta_i, \phi_i; \theta_r, \phi_r; \lambda)}{\pi}$$

It might be confusing that BRDF is not defined as a ratio of equal units although dL_r and dE_i are both clearly directional quantities. The reason is found in the diffuse reflection that causes the small, but finite incident radiant flux to be reflected in infinitesimally small parts in all directions over the hemisphere. It seems appropriate therefore to produce a dimensionless quantity only when all "parts" of dL_r reflected over the hemisphere have been integrated before taking the ratio of dL_r to dE_i . This however contradicts the bidirectional concept of the BRDF.

Both R and f_r are intrinsic properties of the reflecting surface and are mainly driven by the reflection properties of the material, the surface structure, multiple scattering, and mutual shadowing effects. Because of the directionality, BRF and BRDF expose values between 0 and infinity. Values over 1 are achieved in peak reflectance directions such as the specular spot of a metallic surface, where the reflected flux in a single direction is much higher than the flux reflected from a corresponding ideal diffuse (Lambertian) reflector. Hemispherical reflectance ρ , however, only exposes values between 0 and 1, as a result of the conservation of energy. ρ is defined as the ratio of total hemispherical incident Φ_i and reflected Φ_r radiant flux:

$$\rho = \frac{\Phi_r}{\Phi_i}$$

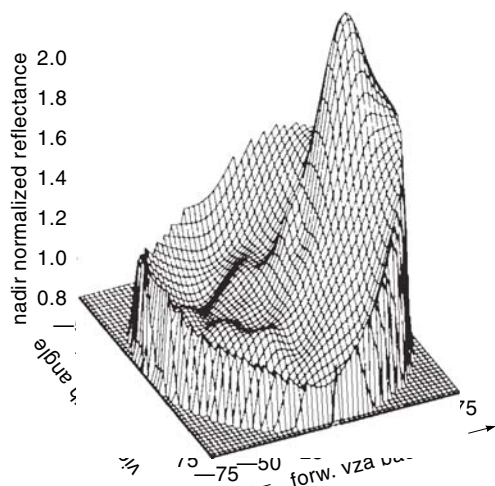


Figure 4. BRDF of a grass lawn at 675 nm wavelength acquired with the FIGOS field goniometer under 35° sun zenith angle.

For most natural and man-made surfaces bidirectional reflectance factors vary significantly for different view and illumination angles. Only very few, highly homogeneous and fine-structured objects like gypsum sand or reference panel material expose nearly lambertian (i.e., completely diffuse reflectance characteristics). Vegetation surfaces for example show a strong backscatter component with a peak reflectance in the direction of illumination, called hot spot (Fig. 4). Mirroring surfaces like metals and some natural surfaces, such as ice and weathered snow, expose distinctive high-reflectance values in the forward scatter direction related to the specular spot. Other materials like concrete exhibit a mixture of forward and backscattering components (Fig. 5). In general, BRDF effects are most prominent in the principal plane where the source of illumination, sensor, and target surface are aligned in one plane. In the plane perpendicular to the principal plane, bidirectional effects are least pronounced.

BRF and BRDF can only be approximated and not actually measured because by definition an infinitesimally small sensor

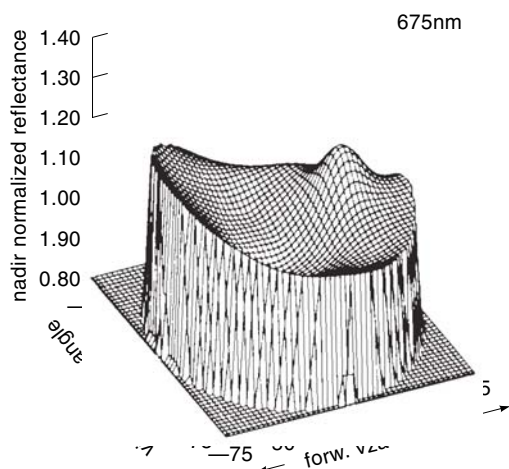


Figure 5. BRDF of a concrete slab at 675 nm wavelength acquired with the EGO laboratory goniometer under 35° illumination zenith angle.

field-of-view would be required. In practice, BRDF values are derived from goniometric measurements by dividing radiances L measured with a small aperture radiometer by the hemispherical irradiance E_i determined from calibrated reference panels. BRF values are likewise obtained by the ratio of fluxes $d\Phi_r$ and $d\Phi_{id}$ measured in small solid angles with a goniometer-mounted radiometer. For a full characterization of a BRDF, an infinite number of measurements would be required. For practical reasons, only a limited number of sample points are measured over the hemisphere, which in most cases are regularly spaced with the help of a goniometer. The BRDF is then derived from these sample points using various modeling techniques.

Spectral Ranges

Most of the goniometric measurements in the optical domain have been performed in the reflecting part of the spectrum between 300 and 2500 nm (e.g., EGO, FIGOS, and PARABOLA, which are described later). In the thermal infrared region relatively few data have been acquired using goniometric devices for remote-sensing applications. Thermal infrared studies revealed a rather high dependence of measured infrared temperatures on view zenith angles. Sun-illuminated parts of a plant canopy were found to be considerably warmer than the shaded components. Kimes et al. (22) found temperature differences as great as 13°C when changing the view zenith angle from 0° to 80° which were a function of canopy structure and vertical temperature distribution. Leaf temperature measurements by infrared emission at different detection angles have also been made using the FEBO spectrogoniometer (2). Goniometers have also been used in the past with Light Detection and Ranging (LIDAR) systems at the Joint Research Centre, Ispra, for the remote sensing of air pollution and in particular for the mapping of smoke plumes (23). Another unique goniometer system installed at the Joint Research Centre, Ispra, is the EMSL facility, which operates in the Radio Detection and Ranging (RADAR) domain and is described in more detail later.

Goniospectroscopy

With the tremendous increase of spectral bands in remote sensing devices in recent years, a need for hyperspectral ground reference data has arisen. Most of the BRDF data sets available today however lack a high spectral resolution. Only recently have hyperspectral BRDF data, acquired with the EGO and FIGOS goniometers, been analyzed (24). In these studies, a strong wavelength dependence of BRDF effects was found for vegetation canopies, mainly caused by multiple scattering effects as a function of leaf optical properties. The wavelength dependency is best observed in an erectophile (i.e., vertically structured vegetation surface such as a grass lawn), which exposes strong BRDF effects resulting from the gap structure of the canopy. Multiple scattering is particularly strong in the highly reflecting near-infrared range, and to a certain degree in the green part of the spectrum, and equalizes effectively the hard contrast between shaded and sunlit canopy particles. In the blue and red chlorophyll absorption bands, however, multiple scattering is low, and the canopy heterogeneity becomes dominant. As a consequence, the BRDF characteristics of a vegetation canopy are more isotropic in the near-infrared and the green range than in the blue and red range. Similar effects can also be observed in the mid-infrared range.

EXAMPLES OF GONIOMETERS USED IN REMOTE SENSING

Both laboratory (25) and field goniometers (26) have been developed and deployed for remote sensing applications, mainly addressing the reflecting range of the electromagnetic spectrum between 300 nm and 2500 nm. Out of the many goniometric systems recently developed, four significant state-of-the-art examples (EGO, FIGOS, PARABOLA, and EMSL) are described here.

EGO

The European Goniometric Facility (EGO) is operated by the Space Applications Institute of the Joint Research Centre, Ispra, Italy. It was constructed and assembled under custom design in a specialized workshop equipped with machinery to build and work large circular rails and arcs with a precision of 0.1 mm (27). The laboratory-based system allows the independent positioning of a light source and a detector anywhere on a 2 m radius hemisphere centered on a target, allowing bidirectional reflectance measurements under controlled laboratory conditions. To reduce light scattering, the goniometer is painted black, and accommodated in a special black laboratory featuring light-absorbing rubber floor material. The main support structure is made from a strong 10 cm double T angle iron. It consists of two horizontal circular rails of 4.4 m and 4.8 m diameter on which two vertical arcs, mounted on motorized sleds, rotate (Fig. 6). The outer arc supports the light source, and the inner arc holds the detector. In its most recent configuration, both arcs cover 90° on the vertical plane. The system is mechanically centered on the ceiling and the floor of the laboratory and weighs about 700 kg. Each arc supports a mobile sled that can displace a detector and a light source of up to 15 kg weight each. The vertical arcs are mounted 240 mm out of axis in such a way that both the source and the detector can reach the zenithal position and that a minimal distance between the source and the detector can be achieved. Another important characteristic is that the origin of the arcs, corresponding to the horizontal plane crossing the center of the hemisphere, is 280 mm above the mechanical plate on which the target is placed, thus allowing measurements at grazing angles. Microswitches and user-set software limits prevent rotation of the arcs beyond certain points, thus reducing the risk of mechanical damage resulting from collision.



Figure 6. The EGO goniometer installed at JRC, Ispra, Italy.

The EGO goniometer is equipped with an automated high-precision target tilt and rotation platform that can tilt and rotate targets with a precision better than 0.001°. The target can be rotated on a full circle, whereas the tilting is limited to an arc of 60° at a velocity of ~0.2°/s. This capability, although exceeding most remote sensing requirements, can be used for instance for targets with a row structure or for those where the specular reflecting component is of particular interest. The targets, which can be up to 1 m × 1 m in dimensions and can weigh up to 7 kg to 8 kg, are precisely positioned by means of small HeNe lasers. The four angular movements of the detector and the light source are realized by precision stepping motors with a resolution of 0.01° and an angular positioning accuracy of ±0.1°. Each stepping motor is equipped with an encoder. The angular velocity on the azimuth rail is 1°/s and on the zenith arcs 0.5°/s. All movements of detector, light source, and platform are controlled by a PC-based custom designed EGO monitoring system software, which is able to handle all experimental and measurement data. The control unit can be operated in manual or batch modes, allowing automated performance of bidirectional reflectance measurements. Two color charge-coupled device (CCD) TV cameras help to capture the experiment set up for future reference. Some of the currently available spectroradiometers used as detectors are the GER IRIS, Spectron Engineering SE590, and the ASD FieldSpec-FR. The system is also equipped with a high-resolution CCD camera. A series of calibrated polytetrafluorethylene (PTFE) panels are available as reference standards being very close to the ideal white diffuse lambertian reflector. Depending on the experiment purpose, various lasers and voltage-stabilized halogen lamps can be used as light sources. It is also planned to use natural light as an illumination source in the future thus enhancing the potential of the EGO goniometer. This versatile facility is currently being used by several European research groups involved in remote sensing applications.

FIGOS

The RSL Field-Goniometer System (FIGOS) was built by Willy Sandmeier at Fa. Lehner & Co AG, Gränichen, Switzerland, in cooperation with the Remote Sensing Laboratories (RSL) at the University of Zurich, Switzerland (28). The planning and construction required about 700 working hours. FIGOS is a transportable field goniometer that is operated with a PC-controlled GER-3700 spectroradiometer covering the spectrum between 400 nm and 2500 nm in 704 bands with a resolution of 1.5 nm (400 nm to 1050 nm) and 8.4 nm (1050 nm to 2500 nm), respectively. FIGOS consists of three major parts: a zenith arc, an azimuth rail of 2 m radius each, and a motorized sled with the spectroradiometer mounted (Fig. 7). FIGOS allows the acquisition of hyperspectral BRDF data under natural light conditions for a wide field of applications such as providing ground reference data for pointable remote sensing devices and BRDF model validation efforts. The technical design and the dimensions are similar to the EGO goniometer, but all parts are made of black-coated aluminum resulting in a total weight of only 230 kg. The complete goniometer system is stored and transported on a trailer with a specifically designed interior, allowing fast and convenient access to a field site. The construction of the zenith arc follows the technique used for cranes providing a high stability in spite of the rather small weight of 48 kg.



Figure 7. The FIGOS field goniometer in action in Zurich, Switzerland.

If transported, the zenith arc is separated into two parts. The azimuth rail is mounted on sockets forming a circle. It weighs about 150 kg altogether. For transportation, the rail and its base are split into four quarters. Mounting of the zenith arc is provided by sleds interconnected on the azimuth rail and allowing a full 360° rotation. The ball-bearing of the wagons embrace the azimuth rail in a way that the zenith arc is tightly fixed even on sloped terrain. A support linking the center of the zenith arc with the azimuth rail serves as a further stabilization and helps to guide the cables. Similar to the EGO goniometer, the zenith arc is mounted eccentrically on the azimuth rail to prevent it from shadowing the target when measuring in the solar principal plane. Freely placable labels on the zenith arc allow for an automated positioning of the spectroradiometer.

The sled with the spectroradiometer mounted is driven by a 24 V dc braking motor, and a precision chain serves as a guide-way for the $\frac{3}{8}$ in. cogwheel. The motor velocity is set to 2.5°/s. By default the labels are set every 15° resulting in 11 measurements with zenith angles ranging from -75° to +75°. It is also possible to drive the sled-motor manually from a remote control unit to any desired position on the zenith arc. The positioning precision on the zenith arc is within $\pm 0.2^\circ$. The geometric precision of the zenith arc is referenced with the help of a laser moving over the zenith arc on plane ground. The deviation of the laser spot, representing the center of the sensor's field-of-view, shows values within ± 3.5 cm. It is introduced by mechanical problems in bending the aluminum profiles. The roundness of the zenith arc is nearly perfect showing deviations of the laser spot from the center within ± 1 cm between -60° and +60°. The azimuth view angle is given by a scale engraved in the azimuth basement. In its current configuration, the zenith arc is positioned manually with the help of a pointer and a brake fixing the position of the zenith arc. The azimuth arc is almost perfectly round. A laser spot pointing vertically from the center of the zenith arc on the ground moves less than ± 1 cm when the zenith arc is rotated. By default, an increment of 30° is set on the azimuth arc resulting in 6 measurement profiles, each containing 11 measurements on the zenith arc. Thus to cover the full hemisphere, 66 measurements of the target surface are needed. In addition, a Spectralon reference panel is used to determine irradiance conditions during data acquisition. A full hemisphere is covered in approximately 15 min. About 90 min are needed for the set-up of the goniometer with a team of two people.

PARABOLA

The Portable Apparatus for Rapid Acquisition of Bidirectional Observations of Land and Atmosphere (PARABOLA) is a rotating head radiometer consisting of three primary units—the sensor head, data recording unit, and a power pack. It was designed and constructed by NASA to enable fast and effective in situ measurements of bidirectional reflectance (29). The original instrument features three spectral bands (0.65 μm to 0.67 μm , 0.81 μm to 0.84 μm , 1.62 μm to 1.69 μm), and an upgraded, commercially available version (Sensit Technologies, Portland, ND) consists of seven channels in the visible and near-infrared range. The sensor elements are mounted within a motorized two-axis rotating head, which scans almost the complete sky- and ground hemispheres in 15° instantaneous field-of-view sectors (respectively 5° for the upgraded PARABOLA version) in only 11 s (Fig. 8). To document the target observed, a nadir-looking camera with a wide field-of-view lens is mounted next to the radiometer head. The roll axis scan rate provides contiguous swaths at the nadir/zenith positions, and progressively increasing overlap at other elevation angles away from the nadir/zenith position. A timing wheel optical sensor triggers concurrent electronic sampling of the voltage outputs from the detectors along the roll axis. The angular positioning accuracy is estimated to be $\pm 0.5^\circ$ as a result of wind and other factors. Unlike EGO and FIGOS, targets are scanned in a continuous helical pattern rather than focusing on a specific spot. To reduce effects of spatial inhomogeneity, the system can be moved within a site, acquiring multiple replicates which are subsequently averaged. The height of the radiometer above the canopy surface is adjusted to accommodate the heterogeneity of the target and provide representative sampling. The primary mounting device for the 4 kg radiometer head is a lightweight, collapsible boom apparatus consisting of an aluminum triangular truss that decouples as four 2 m long sections. It can be deployed using a variety of mounting platforms including tripods, large van booms,



Figure 8. The PARABOLA instrument deployed in the field.

pick-up trucks, and even tower trams and hot-air balloons. All operations of the PARABOLA system are remotely controlled by PC-based software. Data have been collected over a wide variety of surfaces and have been intensively used in the BRDF community.

EMSL

In the microwave range of the electromagnetic spectrum, the European Microwave Signature Laboratory (EMSL), also installed at the Joint Research Centre at Ispra, Italy, is another example of a state-of-the-art goniometric facility that provides unique opportunities in terms of measurement capabilities and data processing (30). The laboratory is mainly devoted to polarimetric radar measurements aimed at complementing air- and spaceborne remote sensing experiments by providing stable and reproducible environmental conditions and flexible operational modes for well-controlled experiments. Although designed to serve researchers and users in the field of land-oriented remote sensing tasks, the laboratory can be efficiently used in many different research fields as well as for industrial applications. The overall structure is formed by the conjunction of a hemispherical and a cylindrical part, both with a radius of 10 m. The height of the chamber is 15 m so that the center of the hemisphere is located 5 m above the chamber floor. In the gap between the two parts, a circular rail is mounted where two sleds carrying the radar antennas can move independently. The object under test is transported inside the chamber through a large door (5 m wide and 8 m high) by means of a target positioner moving on a linear rail. The same system allows the rotational positioning of the object inside the chamber before and during the microwave measurements with a precision of $\pm 0.05^\circ$. Both the electromechanical components and the microwave measurement system are remotely controlled by a computer.

Field Versus Laboratory Measurements

Laboratory goniometers such as EGO and field instruments such as FIGOS and PARABOLA have nearly complementary advantages and disadvantages. Field measurements suffer from an instable irradiance resulting from changing atmospheric conditions and sun positions. But they allow us to measure targets in situ and in vivo under natural light conditions and are therefore generally better suited for remote-sensing applications than indoor measurements. In a laboratory, targets are either man-made or separated from their natural habitat and can suffer from water and temperature stress introduced by the intense laboratory irradiance. Additionally, the light intensity is usually lower than in the field leading to lower signal-to-noise ratios than in the field. Compared to sun light, laboratory irradiance is often highly heterogeneous and non-parallel and may suffer from voltage variations too. Furthermore, the spectral characteristics of the irradiance differ significantly from the solar spectrum. Major advantages of laboratory measurements, however, are the control over the light source position, the nearly complete lack of diffuse irradiance, and the ability to produce data sets in very high angular resolutions. Because of their application in an indoor environment, they are constructed much more sturdily and therefore are more precise than the lightweight field instruments. Thus, calibration procedures, sensitivity analysis, and controlled model validation efforts are generally easier to perform in a laboratory, and

remote-sensing ground reference data are more adequately acquired in goniometric field campaigns. The complete calibration of a large-scale goniometer is usually a very time consuming and tedious task however, and the results are rarely entirely satisfactory (31).

CONCLUSIONS

Geometrical optics have been of great interest during the last two millennia since visual phenomena have been man's most important contact with the physical world. Goniometers, in various forms, have played a very important role in the study of these phenomena. This has recently expanded with the development of sophisticated sensors operating in other regions of the electromagnetic spectrum and by the addition of powerful computerized control systems. Intelligent robots with goniometric positioning capabilities may become commonplace in the near future and replace many of the traditional forms of goniometers. In remote sensing, goniometric measurements and modeling of bidirectional reflectance data will become increasingly important with the advent of NASA's Earth Observing System (EOS) platforms and other satellites with multidirectional viewing capability. The potential of multangular data to derive biogeophysical parameters from remotely sensed imagery is still under study. Spectral libraries of bidirectional reflectance measurements of natural and man-made surfaces are only recently being set up and will become of significant importance for remote sensing and other applications. There is considerable hope that goniometric measurements from ground, air and space will contribute to a better understanding of the earth's biosphere and give insight into global change issues such as the global carbon balance issue. In astronomy, goniometric measurements of particles and particulate surfaces also have been essential in developing models and theories of light scattering from the planetary bodies and will continue to play an important role in the years to come.

BIBLIOGRAPHY

1. A. Carangeot, Goniomètre ou mesure-angle, *Observations sur la physique, sur l'histoire naturelle et sur les arts*, No. 22, March 1783, pp. 193–197.
2. P. Mazzinghi et al., A new narrow angle spectrogoniometer for the measurement of bidirectional reflectance, *Proc. IGARSS'95*, Florence, Italy, 1995, pp. 2075–2077.
3. P. Gibson, Thin films: X-ray characterization, R. W. Cahn (ed.), *Encyclopedia Materials Science and Engineering, Supplementary Vol. 2*, New York: Pergamon Press, 1990, pp. 1328–1332.
4. A. Kononoviz, Determination of the albedo of white cardboard independent of Lamberts' photometric calculations, *Zapiski Matematizhesko Otdela Novorossiskaya Obschestva Estestvoispytatelei*, No. 2, 1882.
5. A. Ångström, The albedo of various surfaces of ground, *Geografiska Annalen*, 1925, pp. 323–342.
6. V. Kulebakin, Reflection of light from terrestrial surfaces, *Trudy Gosudarstvennogo Eksperimentalnogo Elektrotekhnicheskogo Instituta*, 17: 5–29, 1926.
7. H. Kimball and I. Hand, Reflectivity of different kinds of surfaces, *Monthly Weather Rev.*, 58: 280–282, 1930.
8. P. Oetking, Photometric studies of diffusely reflecting surfaces with

- applications to the brightness of the moon, *J. Geophys. Res.*, **71**: 2505–2513, 1966.
9. B. Hapke and H. van Horn, Photometric studies of complex surfaces with application to the moon, *J. Geophys. Res.*, **68**: 4545–4570, 1963.
 10. W. Knowles Middleton and A. Mungall, The luminous directional reflectance of snow, *J. Opt. Soc. Am.*, **42**: 572–579, 1952.
 11. J. van Diggelen, *Photometric Properties of Lunar Crater Floors*, Ph.D. Thesis NASA technical translation to English, NASA F-209, Utrecht Observatory, 1963.
 12. W. Egan, Polarimetric and photometric simulation of the martian surface, *Icarus*, **10**: 223–227, 1969.
 13. B. Zellner et al., The asteroid albedo scale. I. Laboratory polarimetry of meteorites, *Proc. Lunar Sci. Conf.*, **8**: 1091–1110, 1977.
 14. A. Oehler, *Experimentelle und theoretische Untersuchung der goniospektrometrischen Eigenschaften regolithartiger Materialien in den Spektralbereichen UV, VIS, und NIR*, Ph.D. Thesis, Deutsche Forschungsanstalt für Luft- und Raumfahrt, Köln, Germany, 1996.
 15. B. Buratti et al., Spectrogoniometer for measuring planetary surface materials at small phase angles, *Appl. Opt.*, **27**: 161–165, 1988.
 16. J. Piironen et al., Backscattering of light by snow: Field measurements, *Proc. ENAMORS Workshop in Tuusula*, Sept. 1997, (in press), pp. 17–19.
 17. F. Kuik, P. Stammes, and J. Hovenier, Experimental determination of scattering matrices of water droplets and quartz particles, *Appl. Opt.*, **30**: 4872–4881, 1991.
 18. A. Hunt and D. Huffman, A new polarization-modulated light scattering instrument, *Rev. Sci. Instrum.*, **44** (12): 1753, 1973.
 19. C. Sasse and J. Peltoniemi, Angular scattering measurements and calculations of rough spherically shaped carbon particles, *Opt. Sci. and Sci. Instr.*, SPIE's 40th Annual Meeting, San Diego, July 9–14, 1995.
 20. D. Diner et al., A Multi-angle imaging spectro-radiometer for terrestrial remote sensing from the earth observing system, *Int. J. Imaging Syst. Technol.*, **3**: 92–107, 1991.
 21. F. Nicodemus et al., Geometrical considerations and nomenclature for reflectance, *National Bur. Standards Monograph*, vol. 160, 1977.
 22. D. Kimes et al., View angle effects in the radiometric measurement of plant canopy temperatures, *Remote Sens. Environ.*, **10**: 273–284, 1980.
 23. R. Benoit, A. Pedrini, and L. Stanchi, The electronics associated with a LIDAR system for remote sensing of air pollution, *Proc. 2nd Ispra Nuclear Electronics Symposium*, 1975, p. 237.
 24. St. Sandmeier et al., Physical mechanisms in hyperspectral BRDF-data of grass and watercress, *Remote Sens. Environ.*, **66**: 222–233, 1998.
 25. D. Gibbs et al., Automated measurement of polarized bidirectional reflectance, *Remote Sens. Environ.*, **43**: 97–114, 1993.
 26. R. Jackson et al., Bidirectional measurements of surface reflectance for view angle corrections of oblique imagery, *Remote Sens. Environ.*, **32**: 189–202, 1990.
 27. C. Koehler et al., The European Optical Goniometric Facility: Technical description and first experiments on spectral unmixing, *Proc. IGARSS'94*, Pasadena, 1994, pp. 2375–2377.
 28. St. Sandmeier and K. I. Itten, A field goniometer system (FIGOS) for acquisition of hyperspectral BRDF data, *IEEE Trans. Geosci. Remote Sens.*, **37**: 978–986, 1999.
 29. D. Deering, Field measurements of bidirectional reflectance, in G. Asrar (ed.), *Theory and Applications of Optical Remote Sensing*, New York: Wiley, 1989, pp. 14–65.
 30. A. Sieber, The European Microwave Signature Laboratory, *EARSeL Adv. Remote Sens.*, **2** (1): 195–204, 1993.
 31. I. Solheim et al., Calibration and Characterization of Data from the European Goniometric Facility (EGO), Report EUR 17268 EN, 1997.

Reading List

J. C. Stover, *Optical Scattering*, New York: McGraw-Hill, 1990.

B. HOSGOOD
Oklahoma State University
J. PIIRONEN
National Physical Laboratory
G. ANDREOLI
C. KOECHLER
Space Applications Institute
St. Sandmeier
University of Zurich

GYROSCOPES

A gyroscope, sometimes called a *gyro*, is conventionally a rigid body or wheel, spinning around an axis of rotation mounted in a movable frame. This movable frame permits the spinning wheel to tilt freely in any direction and rotate about any axis (Fig. 1).

One of the basic modes of operation and functionality of such a gyroscope can be understood by considering an airplane in which the gyroscope is mounted. In this mode of operation, the gyroscope is used as an instrument that measures the tilt and orientation of the plane. Associated with any spinning object is a quantity known as the *angular momentum*. The angular momentum is the product of the spinning body's *angular velocity* and *rotational inertia*. The angular velocity is measured in radians per second. The rotational inertia, also known as the moment of inertia, depends on the mass and geometry of the body. Conservation of angular momentum is one of the basic principles of classical mechanics. According to this principle, any body upon which no net external torque is applied maintains its angular momentum. For a rigid body under appropriate conditions, this means not only that the angular velocity remains constant, but also that the axis of rotation does not change. The function of the movable frame is to suspend the rotor in such a way that no external net torque acts on it. Thus,

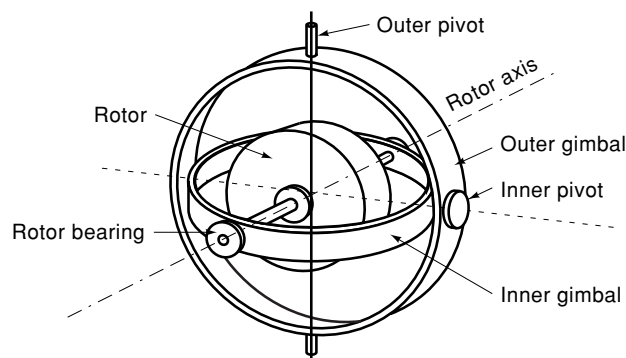


Figure 1. Parts of a two-axis flywheel gyroscope. The rigid body or wheel is referred to as the *rotor*. The two rings are referred to as *gimbals*, and constitute the movable frame.

if the airplane changes its orientation by tilting in one or more possible ways, the rotor will nevertheless keep spinning with the same velocity and with its axis pointing in the same direction. The frame, which is secured to the body of the airplane, will move around the rotor freely. By measuring this motion, it is possible to determine how much the plane has tilted with respect to its original orientation.

One of the most familiar example of a gyroscope is a toy spinning top. Such toy tops are usually set in motion by wrapping a string around the axle and pulling the string. They can perform amazing tricks such as balancing at an angle on the end of a finger or pencil even if the finger or pencil is moved randomly. If a spinning top is pushed gently in one direction, it seems to prefer another, and it moves obstinately in a direction perpendicular to the original force. As another familiar example of the gyroscopic principle in action, we might mention footballs and bullets, which are given a spin while setting them in motion, to keep them stable in flight and send them straight to the target. The same principles that govern the action of gyroscopes are also observed in nature. For example, the earth behaves as a giant gyroscope as it revolves around the sun.

Most modern gyroscopes can be broadly classified into two kinds:

- Mechanical gyroscopes
- Optical gyroscopes

Mechanical gyroscopes are based on the conservation of angular or linear momentum (1). The best-known example of a mechanical gyroscope is the *flywheel gyroscope* discussed above. Some more recent mechanical gyroscopes are based on vibrating, rather than rotating, structures. These vibrating structures are in the form of a tuning fork or a membrane or some other geometry.

Although optical gyroscopes do not contain rotating or vibrating masses, they serve the same purpose as mechanical gyroscopes. Optical gyroscopes have been under development as replacements for mechanical ones for over three decades. The operation of optical gyroscopes is often based on analyzing the interference pattern of two beams of light counterpropagating around a closed path. The interference pattern is an indicator of the direction of rotation. Having very few or no moving parts, these devices are easier to maintain. They also have no gravitational sensitivity, eliminating the need for gimbals.

Whether optical or mechanical in nature, gyroscopes can be further classified into two kinds as (1) *rate gyros*, which provide a voltage or frequency output signal proportional to the turning rate, and (2) *rate-integrating gyros*, which provide the actual turn angle (2). Fundamentally, gyroscopes provide angular rate information. It is important to note that rate-integrating gyros only detect relative angular position, and not absolute angular position like a magnetic compass. Thus, they must be initially referenced to a known orientation by some other means.

Although the rate information can be reliable over long periods of time, when integrated to provide orientation output, even very small rate errors can cause an unbounded growth in the error of integrated measurements. As a consequence, a gyroscope by itself is characterized by *drift* or position errors that grow with time. One way of overcoming this problem is to periodically reset the gyro output with other absolute location-sensing mechanisms and so eliminate the accumulated error.

Gyroscopes have a wide spectrum of applications. The most important use of gyroscopes is in navigation and stabilization instruments for aircraft, spacecraft, guided satellites, missiles, large ships, submarines and other underwater vehicles, cars and other land vehicles, and robots. Some of the specific gyroscopic instruments used are high-performance attitude and heading reference systems, compasses, stabilizers (antiroll equipment), autopilots, and inertial navigation systems (INSs).

An attitude and heading reference system determines the tilts of the vehicle in different directions, so as to aid stabilization, maneuvering, and navigation. Many other gyroscopic instruments, such as the rate-of-turn indicators and the gyro horizons, help aircraft pilots know the position and motion of their plane in the air. Ordinary magnetic compasses are affected by magnetic forces caused by the rolling and pitching of a ship. Gyroscopes are used to regulate such compasses so that navigators know more accurately in what direction their craft are headed. Gyrostabilizers compensate for the rolling of ships at sea, helping restore them to an erect position. Autopilots guide vehicles towards a destination with minimum or no human intervention. The purpose of INSs is to provide position and attitude information.

Gyros find other applications in mining and surveying, the automotive industry, medicine, motion measurement systems, and pointing technology for computers.

Gyroscopes with different accuracies are demanded in different application areas. High-accuracy gyros are required for aircraft, ships, and land vehicles. Medium-accuracy gyros are suitable for less-demanding applications such as vehicles with short flight times (e.g. some missiles) and land navigation systems. Accuracy is often measured in degrees per hour (deg/h). An accuracy of 1 deg/h means that the system makes an angular drift error of one degree over one hour of operation (3). Medium accuracy is defined as a bearing drift in the range 0.1 to 1.0 deg/h; high accuracy, a drift less than 0.01 deg/h (4). For instance, a drift of 1 deg/h would be acceptable in a warhead seeker or flight control system, but would only be tolerable for a short time for standalone navigation.

The *scale factor* of a gyroscope is defined as the ratio of the desired angle or rate information to the physical output (e.g. a voltage or a frequency) of the gyro. The scale factor can be asymmetric for positive and negative rates in mechanical gyroscopes and can be a source of error. This is much less of a problem for certain optical gyroscopes.

Gyro bias is the deviation between the ideal output of the gyro and its actual output. *Gyro bias error* is defined as the difference between the true low-frequency gyro bias (with period greater than the mission time) and the calibrated gyro bias value loaded in the computer to compensate for this error. As long as this quantity remains stable and the calibrated value is subtracted from the gyro output, the compensated gyro output will be correct. One-year stability of this error to better than 0.004 deg/h is achievable (5).

Thermal gradients in the environment affect all types of gyros. Unless adequate error models are built and incorporated in the system model, such errors will also drift the output. Current systems are able to compensate for thermal drift so that the residual thermal drift rate remains under 0.004 deg/h over a wide range of temperatures (5).

Magnetic sensitivity to the earth's magnetic field or to fields created by other instrumentation nearby is another source of

error. Proper shielding enables reduction of magnetic sensitivity by a factor of 60 (5).

HISTORY

The first written record of a gyroscope is found in Gilbert's *Annalen* of 1818. In that report, a flywheel gyroscope constructed in 1810 by a German scientist named G. C. Bohnenberger is described. Instead of a wheel at the center, it had a heavy rotor that was almost spherical, supported by three gimbals. The nineteenth-century French physicist J. B. L. Foucault first used the word gyroscope in 1852 to describe a device he had built to further confirm the rotation of the earth (6). Foucault had earlier demonstrated the rotation of the earth with the pendulum known by his name. However, his gyroscope failed due to the limited technical capabilities of his time. The device was basically a wheel, or rotor, mounted on a long axle within a framework composed of gimbal rings suspended on a ligament. The Foucault gyroscope with the framework suspended on a ligament is considered the original form of the north-seeking gyro. Foucault demonstrated that the spinning wheel maintained its original orientation in space regardless of the earth's rotation. He named this instrument a *gyroscope*, from the Greek words *gyros* (revolution) and *skopein* (to view). Thus, gyroscope means *to view a rotating body*.

A Scot, R. Whitehead, first used the gyroscope in military equipment by perfecting a gyroscopically controlled torpedo in 1896. Although the ability of a gyroscope to maintain its orientation suggested its use as a direction indicator, practical applications of the gyroscope were few before the twentieth century. This was because the equipment needed to keep the rotor of large gyroscopes spinning smoothly, such as electric motors and ball bearings, had not yet been developed. At that time, also, machining techniques were not sufficiently advanced to produce precision instruments.

About 1900, more ships were being built of steel instead of wood. The magnetic compasses that had been satisfactory in wooden ships were less reliable in steel hulls. To overcome this problem, the first gyrocompass for ships was invented and patented by a German scientist and engineer H. Anschütz-Kämpfe in 1903 (6). His colleague Schuler solved the problem of filtering external disturbing motions for the first time.

One of the leading figures in the development of the gyroscope was a US scientist and inventor, Elmer A. Sperry (1860–1930), who became interested in the instrument after seeing Foucault's historic gyroscope in France. In 1909, Sperry built the first automatic pilot using the direction-keeping properties of a gyroscope to keep an aircraft on route. In 1911, he successfully demonstrated a gyrocompass on a US battleship named Delaware. The same year, he patented and marketed a gyrocompass in the United States, and one was produced in Britain soon after. The Germany navy had also introduced gyrocompasses into its fleets by 1911. Sperry continued to extend the range of instruments based on the gyroscope.

The first autopilot for ships was produced by the Anschütz Company in Kiel, Germany, and installed in a Danish passenger ship in 1916. A three-frame gyroscope was used in the same year in the design of the first artificial horizon for aircraft. This instrument indicates roll (side to side) and pitch (fore and aft) attitude to the pilot and is especially useful in the absence of a visible horizon.

In 1912, Sperry's son Lawrence Sperry invented and flight-tested an automatic pilot that used four gyroscopes to stabilize an airplane. He competed with 53 others in a competition in Paris in 1914 and won a prize for the most stable airplane. In 1918, he developed the gyroscopically controlled turn indicator for airplanes.

The British fleet made use of a Sperry gyrocompass in World War I. The Sperry Gyroscope Company (now Sperry Marine, Inc.) devised a quite effective gyrostabilizer that reduced the rolling of ships in 1915. This not only minimized damage to cargo and increased passenger comfort, it also reduced stress in the hull of the ship. This gyrostabilizer fell out of favor because of its large weight, size, and cost. It was replaced by an underwater fin-type ship stabilizer developed by Japanese shipbuilders in 1925.

The directional gyroscope and the gyro horizon, which enable aircraft to fly safely at night and in bad weather, were developed by the Sperry Gyroscope Company in 1928. The same company developed its first gyropilot flight control for modern planes in 1932. This automatic pilot was installed on Wiley Post's airplane, the Winnie Mae. The automatic pilot helped Post make the first solo flight around the world in 1933 in a little more than 7 days, 18 hours. Sperry Gyroscope Company also led the development of several other gyroscopic instruments.

The principle of operation of the optical gyroscope was first discussed by Sagnac in 1913 (7). Sagnac's interferometer experiment produced a sensitivity of 2 rad/s. However, several technological developments had to be made before it could be put into practical use. Two years after the demonstration of the helium–neon laser at Bell Laboratories in 1960, the first operational ring-laser gyroscope (RLG) was developed by Warren Macek of Sperry Corporation (8). It took about two decades to achieve accuracies of 0.01 deg/h. In the early 1980s, smaller, lighter, more reliable RLGs quickly replaced the traditional spinning-wheel gyroscopes for applications such as commercial aircraft or automobile navigation. Navigational-quality RLGs (Fig. 2) have been employed in INSS for Boeing 757 and 767 airplanes since then.

As a result of the advances in the telecommunications industry, in particular optical fiber technology, fiber-optic gyroscopes (FOGs) have emerged as a low-cost alternative to RLGs (9). A FOG was first demonstrated by Stanford University researchers Vaili and Shorthill (10) in 1976. As with RLGs, it took about two decades to develop the technology in order to achieve better accuracies (11). More mature FOG technology emerged in the early 1990s. More than half a million navigation systems utilizing FOGs have been installed in Japanese automobiles since 1987 (12).

Mechanical Gyroscopes

Gyroscopic Forces. It is well known that spinning changes the properties of a body such as a cone-shaped toy top. For instance, while the top cannot stand on its pointed end when not spinning, it can easily be balanced on its pointed end when it is rapidly spinning. Why does the spinning motion give the system stability? Most simply, because the top is "confused" about which way to tip. If it starts to tip in one way, the rotation quickly reorients the tipping motion, and a new tipping process begins. The net result of this continuing process of tipping and reorientation is that the axis of the spinning top moves



Figure 2. MK-49 ring laser gyro navigator by Sperry Marine Inc. This is a navigational quality gyroscope. (Photo courtesy of Sperry Marine Inc., Charlottesville, VA.)

uniformly about a vertical line; this is the motion known as *precession*.

The operation of a bicycle depends on gyroscopic effects. The rider must control the gyroscopic forces in order to ride the bicycle successfully. To keep the bicycle standing upright, the wheels must be kept spinning. If the rider leans slightly to one side, the bicycle does not fall over, but turns in the same direction. Bicycles show two gyroscopic effects: (1) gyroscopic inertia, and (2) precession.

Gyroscopic inertia is the ability of the spinning axle of the gyroscope to maintain the same direction, no matter how the support of the gyroscope changes. The spin axis possesses inertia in the sense that it will maintain its direction as long as the gyroscope is undisturbed and continues to spin. The inertia of a body is its resistance to any change in its state of motion. Gyroscopic inertia plays an important role in determining a gyroscope's behavior. It is gyroscopic inertia that keeps the bicycle upright as long as the wheels keep spinning. If the speed of the wheel decreases, the gyroscopic inertia gradually disappears; the axle begins to wobble and ultimately becomes unstable.

Rotors with a high speed and a concentration of mass toward the rim of the wheel display the strongest gyroscopic inertia.

The consequence of gyroscopic inertia is that if a motor-driven gyroscope is observed for several days, its axis will appear to change its direction slowly, returning to its original position every 24 hours. For instance, if the spin axis of a rotating gyroscope is pointed at the sun, the end of the axis will seem to follow the sun as it crosses the sky. This is because the gyroscope holds its original position in an inertial frame of reference while the earth turns under it, causing the apparent motion. One exception is when the spin axis points toward the polar star. Then, there is no movement of the spin axis with respect to the observer's surroundings, as the axis is parallel to the earth's axis and points towards the poles.

The second interesting property of the gyroscope is its precession capability. Precession is the tendency of the gyroscope to move at right angles to the direction of any perpendicular force applied to it. If a force is applied to the gyroscope which has a component perpendicular to its rotational axis, the gyroscope does not move in the direction of the force. It starts to rotate, or *precess*, in a direction at right angles to the applied force. When the applied force is removed, the gyroscope stops its precessing motion. Precession makes the bicycle turn a corner when the rider leans to one side. This property can be illustrated using a bicycle wheel, mounted on an axle on which it is free to rotate. At first, the axle is supported at both ends while the wheel is made to rotate rapidly. If one support is removed, the wheel and the axle will tend to fall down. In addition, a reaction will be created which opposes the falling motion of the wheel, and instead causes the axle to appear to move horizontally about the remaining support. The removal of one support actually applies a torque to the axle. If a torque is applied to a rotating body that is not along the body's axis of rotation, then the rotational axis moves in a direction at right angles to the direction of the applied force. Hence, when a downward force is applied at one end of the rotational axis of a spinning gyroscope lying horizontally, the resulting torque will cause the gyroscope to precess. During precession, the rotational axis will be moving horizontally about the point of support. The rate of precession, Ω , is proportional to the applied torque τ and perpendicular in direction (13):

$$\tau = \Omega \times I\omega \quad (1)$$

where I is the rotational inertia of the rotor and ω is the rotor spin rate. Gyroscopic precession is the key factor in the operation of the north-seeking gyrocompass discussed below.

When the spinning top is acted upon by a gravity-induced torque, the original rotational motion tries to align itself with the added rotational motion of the torque. In other words, the spin vector tries to align itself with the torque vector by chasing the torque vector.

Flywheel Gyroscopes. The main part of the flywheel gyroscope is the wheel, or *rotor*, illustrated in Fig. 1. Typical rotors employed in aircraft instruments may be smaller than 5 cm and make 35,000 rpm, and are usually made of metal or fluid. In the latter case, the fluid is enclosed within a hollow sphere inside the gyroscope and is rotated at high speed to produce the gyroscopic action.

An *axle* passes at right angles through the center of the rotor. Usually, the rotor is heavily weighted around the rim to

increase its rotational inertia. One of the most important aspects of constructing a gyroscope is to mount the rotating wheel so that it is free of all unwanted torques. If this can be achieved, the spin vector of the gyroscope does not change from its initial value and always points in the same direction with respect to an inertial frame. This way, the gyroscope can function as a good reference direction system.

If the instrument is to be used in the gravitational fields of other bodies, the torques caused by the weight of the gyroscope must be eliminated. A gyroscope used on the earth, for example, is subject to the forces and torques due to gravity. To eliminate the torques, it is necessary to hold the gyroscope with a force applied at its center of mass, which is usually somewhere near the geometrical center of its rotor. It is very difficult to mount the gyroscope at this point. Thus, in order to be free of gravity, gyroscopes are placed in a *Cardan mount*, illustrated in Fig. 1. In the Cardan mount, the ends of the axle are mounted on ball bearings in a movable frame, or ring, called an *inner gimbal*. This inner frame is supported by bearings in an *outer gimbal*. The outer gimbal is mounted on bearings in a supporting frame, or *yoke*, by bearings along a diameter at right angles to the axis of the inner gimbal. Each of these three axes passes through the center of gravity of the entire system.

The number of gimbals determines the number of degrees of freedom of the gyroscope: If two gimbals are used, the gyroscope rotor is free to move in any direction without having to move the frame. If only one gimbal is used, the motion of the rotor is restricted, and it cannot assume certain positions without moving the frame. The center point of the rotor always remains at a fixed position within the gimbals no matter what the orientation of the axle is. The difference between two and three degrees of freedom can be observed in Fig. 3.

Three-Frame Gyroscopes. Unrestrained three-frame gyroscopes have little practical use, since their spin axes are subject to tilting and drifting due to the rotation of the earth. Three-frame gyroscopes are used in the *controlled state*, where the spin axis, by small continuous or intermittent applications of torque, is made to precess so that it oscillates around a mark fixed in relation to coordinates on the earth rather than in relation to space.

While the rotor of a three-frame gyroscope is spinning, if a slight vertical downward or upward pressure is applied to the horizontal gimbal ring at the two ends of the axle, the rotor axle will move at right angles in a horizontal plane. No movement will take place in the vertical plane. Similarly, if a sideways pressure is applied at the same point, the rotor axis will tilt upward or downward. A precession or angular velocity in the horizontal plane is caused by the application of a *couple* in the vertical plane perpendicular to that of the rotor wheel. A couple is a pair of equal and opposite parallel forces.

Controlled gyroscopes fall into three categories: north-seeking (meridian), directional, and gyrovertical (horizon indicator). The north-seeking gyroscope is used for marine gyrocompasses. In the settling (or normal) position, the spin axis is kept horizontal in the plane of a meridian as described later in the section on gyrocompasses. The directional gyroscope is used in aircraft and is sometimes called a self-leveling free gyroscope corrected for drift. Its spin axis is horizontal with directional properties, but the gyroscope does not seek the meridian plane automatically. The gyrovertical has its spin axis vertical and is used to detect and measure angles of roll and pitch. All these three-frame gyroscopes are *displacement gyroscopes*

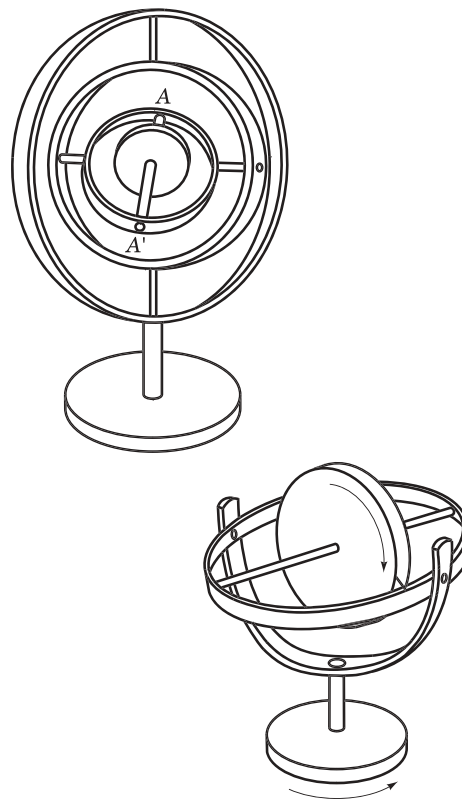


Figure 3. Three-frame gyroscope (top) and two-frame gyroscope (bottom). The number of frames determines the number of degrees of freedom of the gyroscope.

in that they can measure angular displacements between the framework in which they are mounted and a fixed reference direction, which is the rotor axis.

Two-Frame Gyroscopes. Suppose that with the rotor spinning with the spin axis in a horizontal plane, the base of the gyroscope is rotated uniformly in the horizontal plane (Fig. 3, bottom). A resistance due to the gyroscopic inertia will be felt. At the same time, the spin axis will begin to precess in the vertical plane and will continue to do so until the axis is vertical and all gyroscopic inertia disappears. If the same experiment is repeated, except that while the base is being turned in the horizontal plane, the precessional movement of the spin axis is stopped by the application of a force on the end of the shaft where it joins the gimbal ring, then the resistance to the turning motion of the hand due to gyroscopic inertia will cease to exist. The faster the base is turned, the greater the vertical downward force that must be exerted on the shaft to stop the precession. This force can be exerted by a spring arrangement [Fig. 4 (top)] or a U-shaped tube containing mercury fastened to the axis supports. This gyroscope measures the rate of change of azimuth and is used in aircraft and ships as a rate-of-turn indicator. Angular rate of roll in ships can be measured by applying the same principle. In this case, the spin axis is positioned at right angles to the fore-and-aft line and the rate of roll is measured about this line. This is illustrated in Fig. 4 (bottom).

These are velocity or *rate gyroscopes*, which must be distinguished from *displacement gyroscopes* described above. The sensitive or input axis of a rate gyroscope is at right angles to

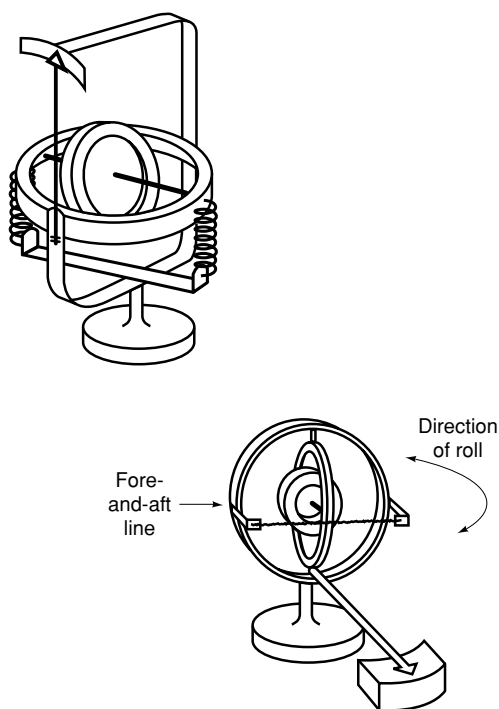


Figure 4. Rate gyroscopes for measuring rate of turn (top) and rate of roll (bottom). In the top figure, a spring arrangement is used to exert a force on the shaft to stop the precession. The amount of force exerted is a measure of rate of turn. In the bottom figure, the spin axis is positioned at right angles to the fore-and-aft line and the rate of roll is measured about this line.

its spin axis, while with a displacement gyroscope the spin axis is directly along the input axis. For example, a north rate gyroscope and a north displacement gyroscope have their spin axes perpendicular to each other.

If the bearing mounts of the gyroscope were frictionless, no external torque would be transmitted to the rotating wheel, irrespective of the orientation in space. In practice, the bearings on which the gyroscope rotates are made with care to minimize friction. They must be assembled in windowless, air-conditioned rooms so as to eliminate dust inside the bearings, which can cause a gyroscope to fail. As an alternative to ball bearings, rotors can also be supported (floated) by a fluid or electrostatic or magnetic fields in so-called *flotation gyroscopes*. In the first case, the airtight inner gimbal is suspended in an inert fluid. Gimbal bearing friction is reduced because the buoyancy of the inner gimbal is neutral in the fluid. Flotation also increases resistance to shock and vibration. There is a diaphragm that seals the product and allows for fluid expansion as the outside temperature and/or pressure changes. Alternatively, the rotor is suspended with the aid of electrostatic fields in a vacuum to prevent any mechanical contact between the spinning ball and the outside case. In some cases, a beam of light reflected from reference marks on the surface of the rotor measures changes in orientation. Moving charges, however, produce magnetic fields and currents that interact with each other and with the earth's magnetic field, thus producing torques just as frictional forces do. Many missiles are guided by electrically suspended gyroscopes.

Since very little friction is produced in the bearings, only a small amount of power is required to keep the rotor turning.

In most instrumental gyroscopes, the rotor is driven by a small self-contained electric motor or a small stream or jet of air that blows on the rotor.

The friction between the spinning wheel and the surrounding atmosphere also produces unwanted torques on the system. Such gaseous friction can be reduced by placing the gyroscope in hydrogen or in a vacuum. Over long periods of time, these combined torques, no matter how small, change the angular momentum of the gyroscope, causing it to drift.

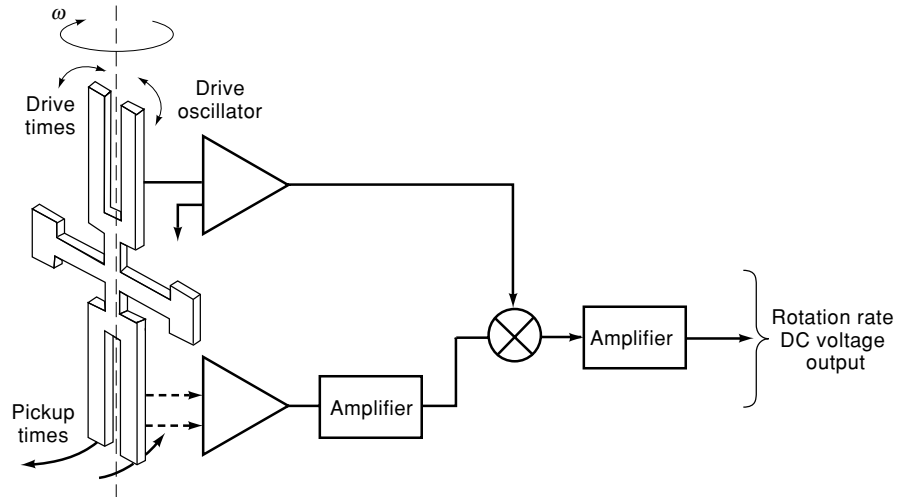
After frictional torques have been minimized, a gyroscope can be used to measure applied torques resulting from forced angular motion. The applied torque, acting over a time interval, affects the original angular momentum by changing either the direction of the spin axis or the magnitude of the spin. Measurement of the change in angular momentum thereby provides information about the applied torque.

Modern Spinning Gyroscopes. A number of precision-machined gyroscopes are available on the market, costing between \$10,000 and \$100,000, depending on the accuracy. There have been recent developments and updates in mechanical gyroscope technology with the advent of solid-state electronics. A miniature gyro named GyroEngine (Gyrations, Inc.) uses a conventional spin gyroscope in which motion is sensed around two independent axes using an optical sensor technique (14). The use of this technology greatly reduces the gyro's size and weight. An injection-molded, clear polycarbonate plastic is used for the housing and structural parts. Polycarbonate was selected because of its lower cost, high strength, ability to withstand a wide range of temperatures, and very good optical properties. Optical properties are important, since the optics needed for the sensor system are molded into the structure to reduce cost.

GyroEngine is a flotation-type, free-spinning, low-cost, low-power, two-degree-of-freedom gyroscope, about the size of a 35 mm film roll weighing only 40 g, that provides data in serial packets. The device is available as either a vertical gyroscope for measuring pitch and roll or a directional gyroscope for measuring yaw.

In a conventional gyroscope, rate data signals are passed in and out of the device through a series of precision slip rings. In the GyroEngine, a light-emitting diode is mounted inside the gimbal assembly and shines through a ring on the inner gimbal, which has a precision grating pattern mounted on it. A set of clear decals with printed optical diffraction gratings (moiré patterns) are mounted at four different places throughout the gimbal system. As the inner gimbal rotates, the light beam passing through the pattern on the ring is modulated in such a way that the motion of the gimbal is detected and tracked. The modulated light beam passes through the center of the gimbal bearing, where it is detected by a photoelectric diode sensor that is outside the gimbal system. The outer gimbal employs a similar optical sensing mechanism. The device reports gimbal position data digitally, eliminating the need for analog-to-digital conversion (ADC) and simplifying the electronics. The open collector outputs can readily be interfaced with digital circuits. The resulting digital signals are transmitted to a microcontroller for processing and output. The control electronics is included in the plastic housing rather than being external. The motor electronics is within the inner gimbal. Since much less signal processing is necessary, the output can be reported without delay (15).

Figure 5. The block diagram of a quartz rate sensor. An oscillating tuning fork senses angular velocity by using the Coriolis effect. The linear motion of the tines is translated into an oscillating torque. After demodulation, a dc voltage is produced, which is proportional to the rate of rotation. (Courtesy of Systron Donner Inertial Division, Concord, CA.)



The resolution of the device is 0.1 deg/rev. A typical drift rate is about 0.15 deg/s. The device costs a few hundred US dollars. Some of its advantages include reduction of the number of slip rings from six or more to two, output being digital rather than analog, data providing absolute position and direction of gimbals, and ability to report motion in real time.

Vibrating Gyroscopes

Quartz Gyroscopes. The concept of using a vibrating element to measure rotational velocity by employing the Coriolis principle has been used for more than 50 years. In fact, the idea developed long ago from the observation that a certain species of fly uses a pair of vibrating antennas to stabilize its flight.

Recently, vibrating quartz tuning fork technology has emerged for the production of microminiature, solid-state gyroscopes on a chip (16). These are similar to the mechanisms used in digital wristwatches to provide the frequency or time reference.

A basic quartz rate sensor (QRS) has essentially two components: drive and pickup as shown in the block diagram in Fig. 5. The drive portion functions exactly like a simple tuning fork: Exploiting the piezoelectric properties of quartz, an electrical signal applied to the tuning fork causes it to vibrate at a fixed amplitude. Drive tines are the active portion of the sensor, and are driven by an oscillator circuit. Each fork tine has a mass and an instantaneous radial velocity that changes sinusoidally as the tine moves back and forth. As long as the fork's base is stationary, only the drive fork vibrates as it responds to signals from an oscillator. The momenta of the two fork tines exactly cancel each other and there is no net energy transfer from the tines to the base. The amount of power required to keep the fork ringing is only about 6 μ W (17).

Another similar fork is employed as a pickup element to produce the output signal. The passive pickup tines of this fork are the sensing portion of the sensor that vibrate only when the device rotates. When the tuning fork is rotated around its axis of symmetry, the Coriolis effect causes the pickup fork to vibrate. According to the Coriolis principle, a linear motion within a rotating framework will have some component of velocity that is perpendicular to that linear motion. Thus, each tine will

generate a force perpendicular to the instantaneous radial velocity of each of the other tines according to the following equation:

$$\mathbf{F}_c = 2m\mathbf{w} \times \mathbf{v}_r \quad (2)$$

where m is the tine mass, $\mathbf{w}$ is the rotation rate, and $\mathbf{v}_r$ is the radial velocity of the tines. Since the radial velocity of the tines is sinusoidal, the resultant force on each tine is also sinusoidal, equal and opposite in direction and in phase with $\mathbf{v}_r$. Hence, the oscillating torque created at the base of the drive tine fork is directly proportional to the input angular rate. The sinusoidal torque variation causes the pickup tines to begin moving tangentially to the rotation and at the same frequency as the drive vibration. The output reverses sign with the reversal of the input rate since the oscillating torque produced by the Coriolis effect reverses phase when the direction of rotation reverses. If there is no rotation, the pickup tines will not move, indicating a zero-rotation input.

The resulting vibration (pickup signal) from this second fork can be electronically analyzed to measure the angle and intensity of movement. The signal is first amplified and then demodulated using the drive frequency as a reference. After further amplification and signal shaping, a dc signal output is produced that is directly proportional to the input angular rate. Signal-processing electronics are fairly simple and usually custom developed on chip, and included within the same package as the sensing element.

QRSs are fabricated chemically from a wafer of single-crystal, synthetically grown piezoelectric quartz material using photolithographic and chemical etching processes characteristic of the microelectronics industry and micromachining techniques (18). Hence, several thousand of these items can be mass-produced in a single batch. Dual tuning forks on a QRS are pure crystal and nearly transparent. An example quartz gyro, Gyrochip II by Systron and Donner Inertial Division, is illustrated in Fig. 6. In another device by the same manufacturer, called MotionPak (see Fig. 7), three QRSs are used in conjunction with three linear accelerometers in order to realize a six degree of freedom inertial sensor cluster.



Figure 6. The GyroChip II solid-state gyro which employs a quartz tuning fork. Several thousand of these items can be mass-produced in a single batch. (Photo courtesy of Systron Donner Inertial Division, Concord, CA.)

The main advantage of quartz gyros is expected to be lower cost of manufacturing and maintenance, which is one-third to one-half the cost of fiber-optical equivalents. Other advantages include superior reliability, design simplicity, low power consumption, small size, low weight, ruggedness, and long operating life. A conventional flywheel rate gyroscope may have 100 to 300 separate precision parts, and a fiber-optic gyro (FOG) may have 10 to 20 parts and many meters of fiber, but a quartz gyroscope has only a single part in its sensing portion. These can withstand high shocks and accelerations of up to 10,000 times the force of gravity resulting from gun or projectile launching. A typical quartz gyroscope has a 2 cm diameter, whereas a comparable fiber-optic gyro would measure 4 to 8 cm.

The typical accuracy of quartz gyroscopes is around 1 to 2 deg/min, which is comparable to that of conventional medium-accuracy rate or rate-integrating gyros. The target markets for FOGs and QRSs differ according to performance capability. FOGs have much higher accuracy than QRSs, in the range of 1 to 10 deg/h. As microelectronic gyros develop further, they will be challenging FOGs for lower-accuracy applications,



Figure 7. The MotionPak inertial sensor, comprising three quartz rate sensors and three linear accelerometers with a total of six degrees of freedom. (Photo courtesy of Systron Donner Inertial Division, Concord, CA.)

especially in single-axis configurations. However, three-axis FOGs will remain cost-competitive because in multiplexed FOG configurations, one set of electronics is shared by all three axes. On the other hand, in three-axis quartz gyros, three sets of electronics are used, so a three-axis quartz gyroscope costs as much as three times what a single axis gyro costs.

QRSs have several disadvantages compared to FOGs. Key performance parameters such as bias drift and scale factor are unstable over the long term (19). The scale factor is a measure of how accurate and stable a signal is over time. Secondly, the turn-on to turn-on bias may be large. A large turn-on to turn-on bias means that gyroscope drift has different values on one day and the next. The vibration and shock performance is usually uncertain. Part of the ongoing research is focused on improving the accuracy of quartz gyroscopes.

Hemispherical Resonator Gyroscope. The solid-state hemispherical resonator gyro (HRG) is based on the rotation-sensing properties of a ringing wine glass, first noticed by the British physicist G. H. Bryan in 1890. The modern implementation of this principle involves a gyro comprising three fused-quartz parts, a wine-glass-shaped resonator, an external forcer housing, and a pickoff housing on the unit's base. These are joined by indium solder and employ metalized thin-film electrodes and conductors. When voltage is applied to the resonator, it flexes up to 0.00025 cm and creates a low-amplitude wave that can be sensed by the pickoffs. The precession of this wave is correlated with the rotation angle. There are no bearings or surfaces in the HRG subject to wear, and the interior of the device is maintained in a high vacuum. A typical service time is expected to be 20 years with mean time between failures (MTBF) of more than 150,000 h. Precision manufacturing is critical in the operation of this gyroscope.

Murata Gyrostar. Another type of vibrating gyroscope is the Gyrostar, which is a small relatively inexpensive single-axis piezoelectric rate gyroscope originally developed for the automobile market and active suspension systems by Murata Manufacturing Company in 1991 (20). A picture of the device is provided in Fig. 8.

The principle of operation is again based on the fact that a proportionate Coriolis force results if angular velocity is applied to a vibrating object. The sensor element unit comprises a 40 mm long triangular prism made of a special substance called *elinvar* (elastic invariable metal), which is a nickel-chromium-steel alloy with a very small coefficient of thermal expansion to ensure good dimensional stability (21).

Both the excitation and the detection units employ piezoelectric ceramics. On each of the three vertical faces of the bar, a PZT-6 ceramic piezoelectric transducer is placed as illustrated in Fig. 9. The flexural resonance frequency of a triangular prism is given by (22):

$$f_n = \frac{ka}{4\pi l^2} \sqrt{\frac{E}{6\rho}} \quad (3)$$

where k is a constant, a and l are the width and length of the bar respectively, E is Young's modulus of elasticity, and ρ is the density of the bar material.

Excitation of the prism by the left and right transducers at the resonant frequency $f_n = 7.85$ kHz of the bar, perpendicular to its face, causes vibrations to be picked up by the third transducer, which provides feedback to the drive oscillator (22). The



Figure 8. The Gyrostar manufactured by Murata Electronics. This is a small, relatively inexpensive, single-axis piezoelectric rate gyroscope originally developed for the automobile market and active suspension systems. (See B. Barshan and H. F. Durrant-Whyte, *IEEE Trans. Robot. Autom.* **11** (3), 328–342, June 1995. Copyright IEEE, 1995.)

equilateral triangular prism allows the driving transducers to be configured in the direction of the compound vibration mode. The same elements can be used for both excitation of the bar and detection of the resulting Coriolis forces (23,24).

The gyroscope detects the angular rotation by measuring the differential output of the left and right transducers. If the sensor remains still, or moves in a straight line, the signals produced by the pickup transducers are exactly equal. If the prism is rotated around its principal axis, a Coriolis force proportional to the rate of rotation w about the z axis is created according to the equation

$$F_c = 2mw \times v_y \quad (4)$$

where F_c is the Coriolis force, m is the equivalent mass of the prism, and v_y is the rate of change of position of the element in the y direction. The actual rotation rate w can be determined by measuring the amplitude of the vibration at the pickup

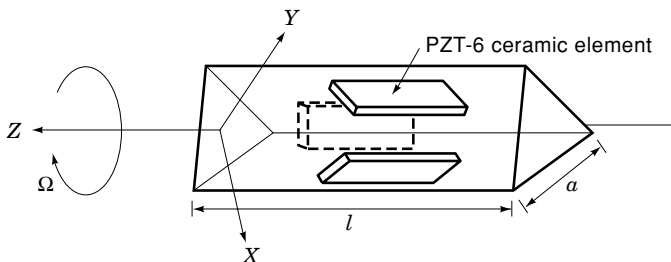


Figure 9. The triangular prism of Gyrostar with a piezoelectric ceramic transducer placed on each of the three faces of the bar. Excitation of the prism at its resonant frequency by two of the transducers causes vibrations to be picked up by the third transducer.

transducer. As the bar distorts due to the Coriolis forces, one detector output increases while the other decreases. The differential output is amplified to yield an analog voltage proportional to the angular velocity of the vehicle around the principal axis of the device. A block diagram of the device is provided in Fig. 10. The maximum rate that can be measured with Gyrostar is ± 90 deg/s within its linear range, with a response time of 0.02 s.

The unique geometry of the Gyrostar provides better performance than tuning-fork gyros at a significantly lower price (23). On the other hand, there is still thermally induced drift due to thermal gradients within the element and the mismatches in the material thermal expansion rates at the elinvar–ceramic interfaces. An evaluation of this gyroscope and modeling of its drift error is provided in Ref. 25.

The main application of the Gyrostar has been directional control in car navigation systems by keeping track of turns for short durations when the vehicle is out of contact with reference points derived from the additional sensors. Other target applications include video camera stabilizers, position and posture control of moving objects, robotics, and controlling the direction of satellite antennas on moving objects. The cost of the device is several thousand Japanese yen.

Optical Gyroscopes

Active Ring-Laser Gyroscopes. The operating principle of the active ring-laser gyro (ARLG), which is an angular-rate sensor, is based on the Sagnac (interferometer) effect. The basic device consists of two laser beams traveling in opposite directions around a closed-loop path. In 1966, Schulz-DuBois idealized the RLG as a hollow doughnut-shaped mirror in which the closed-loop path has circular shape (26).

Conventional RLGs include a source of lasing and optical mirrors. A circulant cavity or waveguide is made out of a low-expansion, quartzlike glass–ceramic material, and is filled with a gas such as helium–neon, which acts as the lasing medium. When the gas is excited, photons are emitted and begin to circle around in the cavity in both directions. In effect, this creates two counterrotating beams of coherent laser light, which create a stationary standing wave with intensity nulls and peaks as shown in Fig. 11, regardless of whether the gyroscope is rotating or not. The interference pattern tends to stay fixed in inertial space, and the path of the light varies with the rotational motion of the gyroscope. If the gyroscope cavity rotates in the counterclockwise (CCW) direction, then the CCW-propagating beam will traverse a slightly longer path than under stationary conditions. The path of the clockwise (CW) traveling beam will be shortened by the same amount. Consequently, a phase shift results between the two laser beams, which can be monitored. The magnitude of the change in the path length ΔL is given by

$$\Delta L = \frac{4\pi r^2 w}{c} \quad (5)$$

where r is the radius of the circular beam path, w is the angular velocity of rotation, and c is the speed of light in the medium. Since the change in path length is directly proportional to w , rotational rate measurement relies on accurate measurement of the change in the path length. The invention of the laser provided the means of accomplishing this measurement. A major portion of the light impinging upon the surfaces

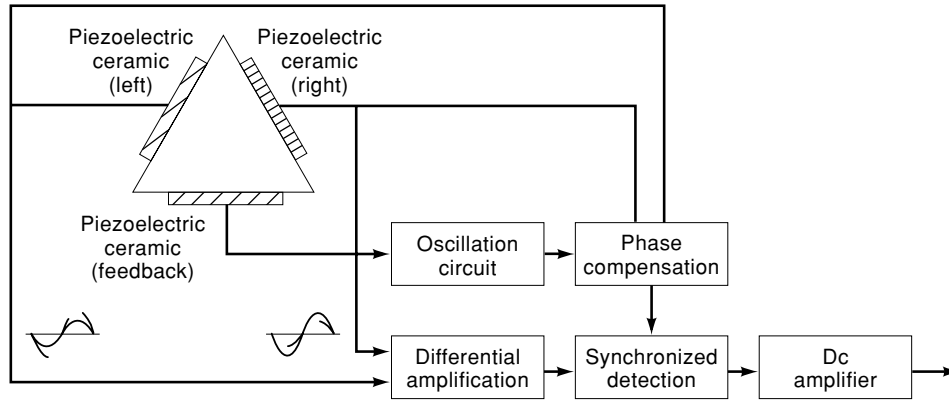


Figure 10. Block diagram of the Gyrostar. The differential output is amplified and phase compensated to yield an analog dc voltage proportional to the rate of rotation around the principal axis of the device.

of the mirrors is reflected by the mirrors, and a minor portion is transmitted through at least one of the mirrors. The light transmitted through is measured by a system that includes optical sensors and a data processor to permit the device to detect changes in rotational motion.

In order for lasing to occur, the round-trip beam path must be precisely equal in length to an integral number of wavelengths at the resonant frequency. Thus, the wavelengths of the two counterrotating beams must change when rotation occurs. The resulting frequency difference or beat frequency Δf , between the two beams is given by (26)

$$\Delta f = \frac{2f r w}{c} = \frac{2r w}{\lambda} \quad (6)$$

where f is the frequency and $\lambda = c/f$ is the wavelength of the beam.

A doughnut-shaped resonator cavity would be practically difficult to realize. For an arbitrary cavity, the equation becomes

$$\Delta f = \frac{4Aw}{P\lambda} \quad (7)$$

where A is the area enclosed by the closed-loop beam path and P is the perimeter of the beam path. For single-axis gyros, the closed-loop path is most often formed by aligning three highly reflective mirrors to create a triangular path as shown

in Fig. 12. Systems similar to Macek's early prototype employ four mirrors to create a square path. The mirrors are usually mounted on a monolithic glass-ceramic block with machined ports for the cavity bores and holes (27).

A modern triaxial design employs six mirrors, centrally mounted on the faces of a cube. Within the enclosed volume of the glass-ceramic block, three mutually orthogonal and independent RLGs are placed such that each employs four of the mirrors. Hence, each mirror is shared by two of the axes (28). A picture of an example triaxial gyroscope is given in Fig. 13. To avoid magnetic sensitivities, the most stable systems employ linearly polarized light and minimize circular polarization (8).

Reliability and robustness are among the advantages of ARLGs. The main disadvantage of the ARLG is a phenomenon called *frequency lock-in*, which occurs at low angular rates when the two beams "lock" together in frequency (29). As in any physical system that sustains two modes of oscillation, problems occur when the two frequencies approach each other. Energy is exchanged between the two modes and the frequencies tend to lock and become one, violating Eq. (7). This trading of energy or coupling is mainly caused by periodic modulation of the gain medium or other effects due to a very small amount of backscattered radiation from the imperfect mirror surfaces (2). The result is a small deadband region for low rotational rates, within which no output signal is observed. Above the lock-in threshold, the output converges to the ideal linear response curve in a parabolic fashion.

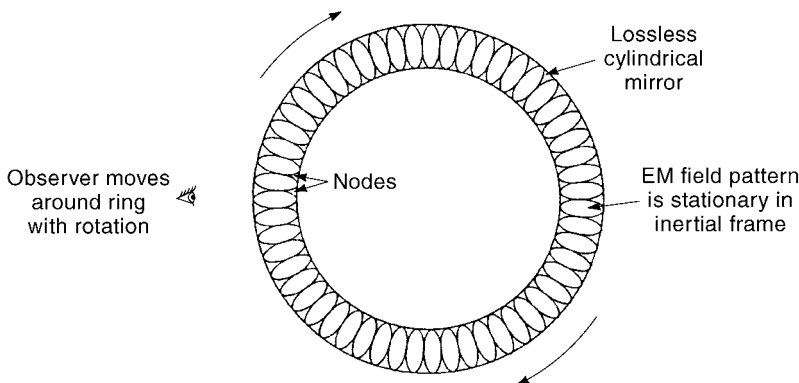


Figure 11. Standing wave pattern created by two counter-rotating laser beams in an idealized RLG. The interference pattern tends to stay fixed in inertial space, and the path length of the light varies with the rotational motion of the gyroscope.

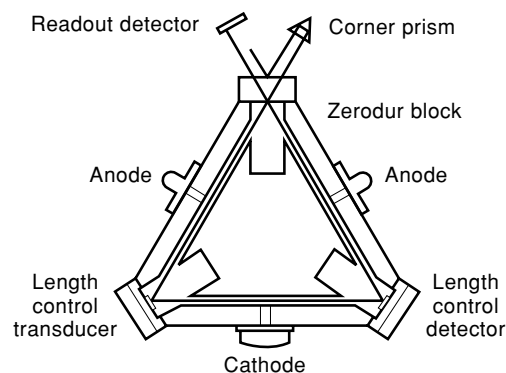


Figure 12. A triangular configuration for a single-axis RLG employing dual anodes. Three highly reflective mirrors have been used to create the triangular path. (Courtesy of John Wiley & Sons, Inc., New York; adapted from Ref. 2.)

One way to solve the frequency lock-in problem is by improving the quality of the mirrors in order to reduce backscattering. A more practical technique for reducing lock-in is to use a biasing scheme to shift the operating point away from the deadband zone. This is commonly accomplished by *mechanical dithering*, where the gyroscope assembly is oscillated back and forth about the principal axis (typically ± 100 arcseconds at 400 Hz)



Figure 13. An example triaxial monolithic RLG operating at 632.8 nm. (Photo courtesy of Kearfott Guidance & Navigation Corporation, Inc., Wayne, NJ.)

by using a stiff dither flexure suspension acting like a rotary spring. Piezoelectric transducers provide the force to rotate the laser cavity block. Disadvantages of this method include increased system complexity, increased failure rate due to moving parts, and crosstalk between axes. In addition, it introduces a random-walk error increasing as the square root of the time. After the application of randomized dither, residual lock-in falls just under 0.002 deg/h and random-walk error remains between 0.001 and 0.003 deg/h^{1/2}. Hence, dithering is not suitable for high-performance systems such as those used for flight control.

Other methods of reducing the frequency lock-in include the use of extremely short-duration laser pulses (30–33), the use of techniques based on nonlinear optics (8), or removing the lasing medium and using a passive resonator as described below.

The RLG is limited by the quantum noise level due to spontaneous emission in the gain medium (34). Other sources of error include stability of the mirror's optical axis, mirror surface erosion, outgassing of epoxy material within the laser cavity, precision of path length, current, and dither control, all of which affect the gyro bias error. Yet, the ARLG provides the highest sensitivity and is perhaps the most accurate implementation to date. The main problem with the ARLG is its high cost. If the cost can be lowered, the device will be more widely used.

Aronowitz (35), Menegozzi and Lamb (36), Chow et al. (27), Wilkinson (37), and Udd (2) discuss the theory of RLG and its fiber-optic derivatives in detail. Ezekiel and Arditty provide a tutorial review of the technologies, and an extensive bibliography on earlier work (34).

Passive Ring-Laser Gyroscopes. In the passive ring-laser gyroscope (PRLG), the laser source is external to the ring cavity,

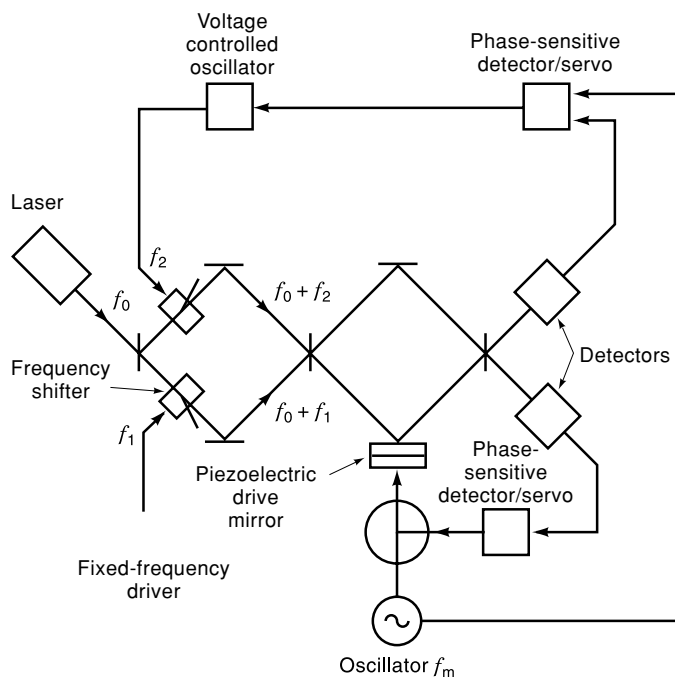


Figure 14. Block diagram of passive cavity ring-laser gyro. The laser source is external to the ring cavity, providing a solution to the frequency lock-in problem. (Courtesy of John Wiley & Sons, Inc., New York; adapted from Ref. 2.)

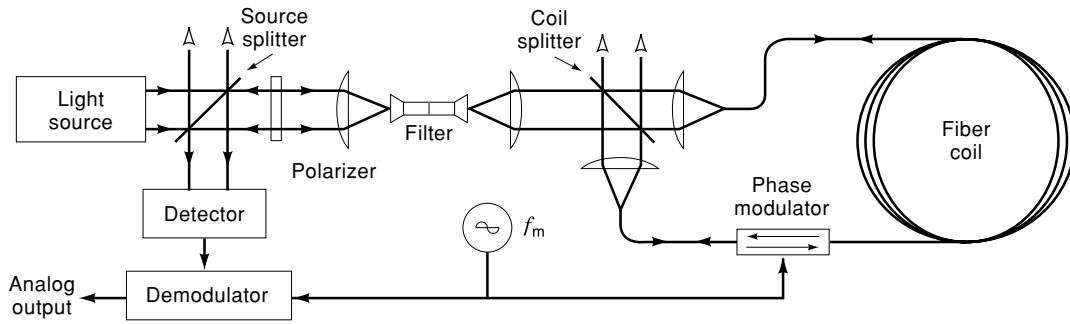


Figure 15. Simplified block diagram of an open-loop IFOG. A single beam of light is split, with the two parts propagating in opposite directions through the optical fiber coil. After demodulation, an analog output is produced which is proportional to the rate of rotation.

ity as in Fig. 14, providing a solution to the frequency lock-in problem. This configuration also eliminates the problems caused by changes in the optical path length within the interferometer due to variation in the index of refraction of the gain medium (27). One problem, however, is that the theoretical limit for the photon shot noise level is higher than that of the ARLG (34).

The main disadvantage of both active and passive RLGs is the bulky packaging compared to those gyroscopes based on fiber-optic technology. In addition, production of RLGs requires high-tolerance machining and clean-room assembly. As a result, the *resonant fiber-optic gyroscope* (RFOG) has emerged as the most popular of the resonator configurations (38).

Like RLGs, FOGs are angular-rate sensors based on the Sagnac effect. Basically, a long fiber-optic pathway is created and wound into coils. A single beam of light is split, with the two parts being sent in opposite directions through the optical fiber. A low-coherence source, such as a superluminescent diode, is typically employed to reduce the effects of noise (39). The primary source of noise is backscattering within the fiber and at any interfaces. Consequently, in addition to the primary mode, there are also a number of parasitic secondary modes that yield secondary interferometers (40). The limited temporal coherence of the broadband light source causes any interference due to backscattering to average out to zero, making the system sensitive only to the interference due to the primary mode (34,40). A simplified block diagram is provided in Fig. 15.

The glass fiber forms an internally reflective waveguide for the beams of light. In essence, it replaces the bulky doughnut-shaped cavity first suggested by Schulz-DuBois. A *step-index multimode fiber* comprises a core region with a high index of refraction surrounded by a protective cladding with a lower index of refraction to keep the light in the core region through total internal reflection (41). If the core diameter is much larger than the wavelength of the light, a number of rays following different-length paths can simultaneously propagate down the fiber. Such multimode operation is clearly not desirable in gyro applications. If the diameter of the core is sufficiently reduced to approach the operating wavelength, only a single mode can propagate (41). *Single-mode fiber* is employed to ensure that the two counterrotating beams follow identical paths when the gyro is stationary. The fiber is also chosen to be of the *polarization-maintaining* type, since light of different polarization states travels through an optical fiber at different speeds (9,41).

Mainly, two types of FOG exist: interferometric (IFOG) and resonator (RFOG). When the two counterrotating laser beams meet at the end of the pathway, the two beams are compared. If rotation has taken place, the two beams will differ in some way, and the difference can be detected. If this is done interferometrically, the device is called an IFOG. More than 95% of existing FOG applications employ the IFOG, although the RFOG has been becoming more widespread recently. IFOGs can be further classified as open-loop and closed-loop gyroscopes as described below.

Open-Loop Interferometric Fiber-Optic Gyroscopes. The number of fringes of phase shift introduced by gyro rotation is given by (2)

$$Z_R = \frac{LD}{\lambda c} w \quad (8)$$

where L is the length of optical fiber in the loop, D the diameter of the loop, λ the wavelength of light, c the speed of light in vacuum, and w the rotational rate. The stability of the scale factor relating the phase shift to rotational rate is dependent on the stability of L , D , and λ (34). Typically, an IFOG operates over $\pm\pi/2$ phase shift with a theoretical sensitivity of $1 \mu\text{rad}$ or less (42). Increasing L by the use of multiple turns of fiber enhances resolution by effectively multiplying the change in the path length due to the Sagnac effect by a factor N equal to the integer number of turns (2). The optimal length is of the order of several kilometers, after which the fiber attenuation (typically 1 dB/km) begins to degrade performance (34). This large amount of fiber required represents a significant percentage of the overall system cost.

For FOGs the main accuracy limit is not optics but noise in the electronic control loops. The accuracy of FOGs depends on fiber length and coil diameter used as implied by Eq. (8) above. FOGs typically operate at one of three wavelengths: 0.86 , 1.33 , or $1.55 \mu\text{m}$ (42). The shortest wavelength is optimum for a precision FOG, since it offers the greatest sensitivity. In contrast, longer wavelengths are preferable for telecommunication systems because of lower signal attenuation.

A minimum-configuration open-loop FOG consists of an optical source (such as a semiconductor diode light source), source-detector directional coupler, polarizer, coil directional coupler, and an optical-fiber coil (Sagnac ring). The source-detector coupler is not part of the Sagnac interferometer and serves mainly to provide isolation between the broadband optical source and

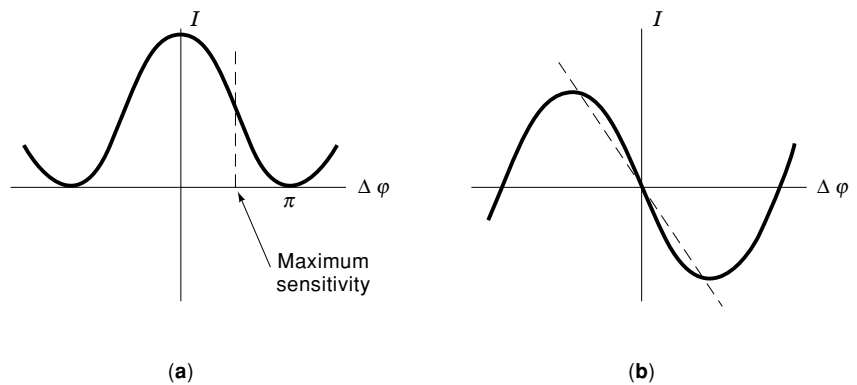


Figure 16. Detected intensity versus phase shift (a) and output of demodulator (b) for the analog open-loop IFOG. The symmetry of the original intensity pattern, shown in (a), does not allow distinguishing the direction of rotation. By introducing a phase shift of $\pi/2$ to the two beams, the operating point is shifted to the region of maximum sensitivity as shown in part (b).

the photodetector. When the two counterrotating laser beams meet at the end of the pathway, the fringe pattern created due to interference is inspected. Rotation of the path compresses the fringes in one direction and expands the fringes in the other direction, resulting in a changed pattern in the fiber. The intensity pattern caused by constructive and destructive interference, shown in Fig. 16, is in the form of a cosine, which is symmetric and cannot distinguish between CW and CCW rotation. The peak intensity occurs at zero rotation rate, where the sensitivity of the intensity to small changes in rotational rate is minimum. To overcome these problems, the two beams are phase-modulated to introduce nonreciprocal phase shifts (2). If a phase shift of $\pi/2$ is introduced, the operating point is shifted to the region of maximum sensitivity and the direction of output can be determined from the sign of the output (34). The return signal from the interferometer traverses the laser and is recovered by a photodetector behind the rear laser facet. This signal is distinguished from the laser signal by the bias modulation. This minimal configuration eliminates at least one directional coupler, the separate photodetector, and at least two fiber splices.

The disadvantages of this open-loop configuration are the nonlinearity of the input–output relationship, bias instability, the long length of optical fiber required relative to other FOG designs, limited dynamic range in comparison with ARLGs, and the sensitivity of the scale factor to analog component drifts/tolerances and light source intensity (43). On the other hand, it offers reduced manufacturing costs, high tolerance to shock and vibration, insensitivity to gravitational effects, quick startup, and fairly good sensitivity in terms of bias drift and random walk coefficient. The coil geometry is not critical. Therefore open-loop FOGs are more suitable for low-cost systems for low- to medium-accuracy applications such as gyrocompassing in automobile navigation, pitch and roll indicators, and attitude stabilization (44). Hitachi Cable, Ltd., Tokyo, has developed several IFOGs for a number of industrial and commercial uses such as automotive applications, mobile robotic systems, and agricultural helicopters (45,46).

A popular example of an open-loop IFOG is the Andrew AUTOGYRO®, which is a single-axis all-fiber rate sensor that includes an interface to the vehicle odometer and backup lights for use in dead-reckoning navigation (47) (Fig. 17). The device integrates angular rate data and odometry information over an interval, and combines these into a serial digital data output. The sensing element is a coil fabricated from elliptical-core, polarization-maintaining optical fiber which does not depend

on stress-inducing members and operates from unconditioned vehicle power. The signal-processing electronics are stable with time and temperature, and the unit does not need to be recalibrated. An internal temperature sensor output enables further calibration in more accuracy-demanding applications. The packaging is rugged, consisting of an aluminum housing with weather-resistant gaskets.



Figure 17. Picture of the AUTOGYRO® Navigator by Andrew Corporation. This is an example of an open-loop, all-fiber, single-axis IFOG. (Photo courtesy of Andrew Corporation, Orland Park, IL.)

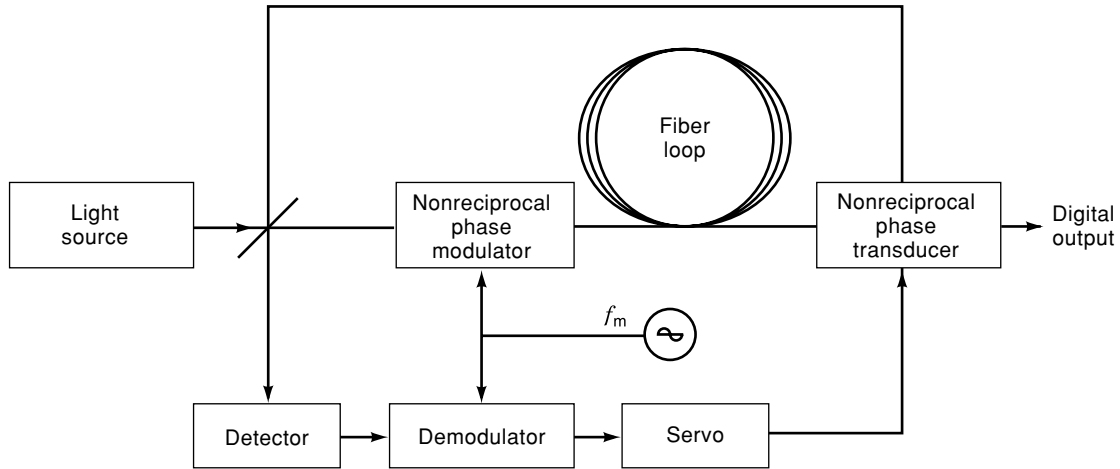


Figure 18. Simplified block diagram of a closed-loop phase-nulling IFOG. The output of the demodulator is passed to a servo amplifier that drives a nonreciprocal phase transducer which introduces a frequency difference between the two beams.

Analog and digital output versions are available. The analog signal is the angular rate output, whereas the digital output is the result of integration of the rate output, which can be interpreted as either the incremental angle change during the measuring interval or the average rate. A wide analog bandwidth is maintained while reducing the output data rate. All required voltages are generated internally from a single, wide-range input voltage that can be directly connected to vehicle power.

The lack of moving parts implies longer life. Among other advantages of the device are stable operation, lack of sensitivity to rotation about or acceleration along other axes, good resolution, threshold and dynamic range, resistance to shock and vibration, high reliability, ease of interfacing, and freedom from preventive maintenance.

The device can detect input rotation rates between ± 100 deg/s with a minimum detectable rotation rate of ± 0.05 deg/s.

The AUTOGYRO® is used in a continuous positioning system (CPS) costing about 3000 US dollars, comprising the AUTOGYRO® and a GPS receiver. The CPS is used in automated bus stop announcement systems and emergency vehicles such as ambulances, police cars, and fire trucks. The price of the AUTOGYRO® is under a thousand US dollars. A more detailed discussion of the AUTOGYRO® is provided in Refs. 48 and 49.

Closed-Loop Interferometric Fiber-Optic Gyroscopes. For applications demanding higher accuracy, the closed-loop FOG is more suitable, with drifts in the average 0.001 to 0.01 deg/h and scale-factor stabilities greater than 100 ppm (43). In closed-loop systems, an active torquing feedback loop into a frequency or phase-shifting element is employed to cancel the rotationally induced phase shift (2). Since the system is always operated around a zero phase shift, the gyro accuracy and drift are improved, and intensity variations in the light source and analog component tolerances have an insignificant effect (34). Closed-loop systems, however, rely on costly high-speed optical and electronic components.

A simplified block diagram of a closed-loop IFOG is illustrated in Fig. 18. The output of the demodulator is passed

to a servo amplifier that drives a nonreciprocal phase transducer (NRPT), which is typically an electro-optic frequency shifter placed within the fiber interferometer (34). The NRPT introduces a frequency difference Δf between the two beams, resulting in a fringe shift Z_F given by (2)

$$Z_F = -\frac{\Delta f L n}{c} \quad (9)$$

where n is the index of refraction and c is the speed of light. The linearity and stability of the gyro depend only on the NRPT (34).

To null out the phase shift at the detector, the fringe shift Z_R due to gyro rotation must be precisely canceled out by the fringe shift Z_F due to the relative frequency difference of the two beams:

$$Z_R + Z_F = 0 \quad (10)$$

Substituting Eqs. (8) and (9) for Z_R and Z_F and solving for Δf yields (2,34,44)

$$\Delta f = \frac{4AN}{n\lambda L} w = \frac{4A}{n\lambda P} w = \frac{D}{n\lambda} w \quad (11)$$

where A is the area of the fiber loop, N the number of turns in the loop, and P and D the loop perimeter and diameter respectively.

The gyro output Δf is thus inherently digital, as opposed to an analog dc voltage level, and also linear. However, closed-loop digital signal processing is considerably more complex than the analog signal processing employed in open-loop IFOGs.

Resonant Fiber-Optic Gyroscopes. The resonant fiber-optic gyro (RFOG) has evolved as a solid-state derivative of the PRLG described above. A block diagram is provided in Fig. 19. A passive resonant cavity is formed from a multiturn closed-loop of optical fiber. Frequency-modulated (FM) light is coupled from a laser source into the resonant loop in both the CW and CCW directions. For the case of no motion, maximum coupling occurs when the frequency of the laser during the FM sweep is

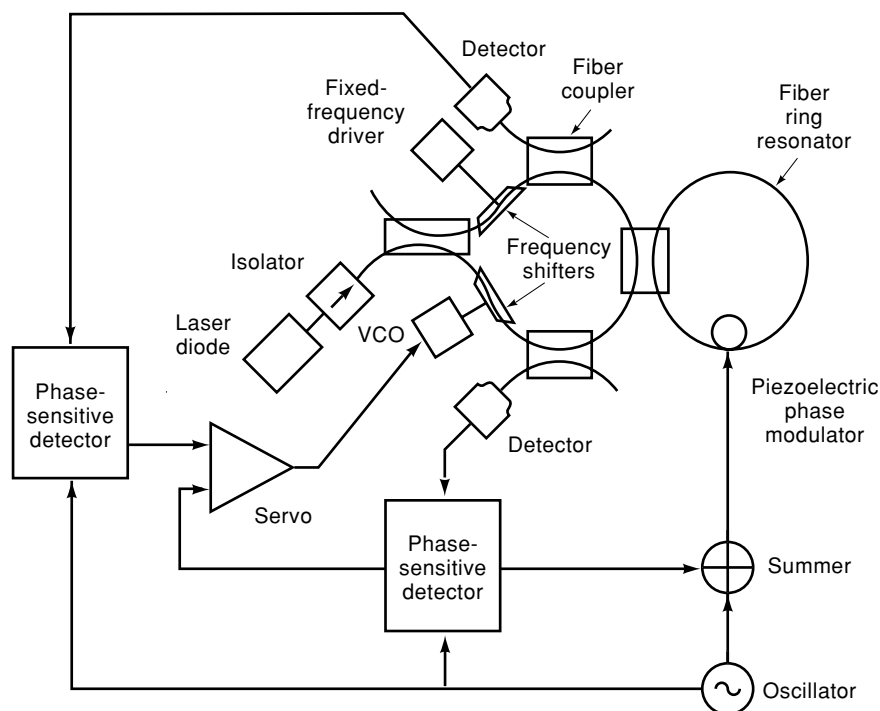


Figure 19. Block diagram of a fiber-optic ring resonator. Frequency-modulated light is coupled from a laser source into the resonant loop in both directions. The output coupler samples and detects the intensity of the energy in the loop. (Courtesy of John Wiley & Sons, Inc., New York; adapted from Ref. 2.)

such that the perimeter of the loop is an integral multiple of the wavelength (38). If the loop is rotated, the path lengths for the two beams will no longer be equal and the resonant frequencies shift accordingly. The output coupler samples and detects the intensity of the energy in the loop. The demodulated output at the detectors will show resonance peaks separated by a frequency difference of Δf given by (38)

$$\Delta f = \frac{D}{\lambda n} w \quad (12)$$

In practice, the frequency of the laser is adjusted to maintain resonance in one direction, and an electro-optical frequency shifter is employed to drive the other direction into resonance (44). This results in a frequency shift twice that induced by the Sagnac effect.

Advantages of RFOG are high reliability, long life, quick startup, and light weight. It requires 10 to 100 times less fiber in the sensing coil than the IFOG configuration. Since the optical fiber can be as short as 50 to 100 m, the size of these gyros is comparable to that of a spool of thread. RFOGs are limited by the same shot-noise level as IFOGs (38). According to Sanders, this is due to the fact that light traverses the loop multiple times, as opposed to once in the IFOG. Two disadvantages are the need for a highly coherent source and the need for extremely low-loss fiber components (43).

Fiber-Optic Gyroscopes. Currently, RLGs are used for navigation applications requiring high accuracy, whereas FOGs are used for medium-accuracy applications such as in motion sensors. FOGs have certain advantages over RLGs. The latter rely on an active source of lasing for their operation and are very expensive to manufacture. FOGs do not require a coherent source, and their operation does not depend on a high-performance cav-

ity, significantly reducing manufacturing costs (50). The mass production price of FOGs is estimated to be one-third that of comparable RLGs. Utilization of optical fiber in FOGs provides considerably greater ruggedness. Among other advantages of FOGs are light weight, low power consumption, small size, potential for mass production, little or no lock-in, no plasma flow problems, no critical mirror fabrication or aging problems, smaller number of parts, absence of mechanical moving parts (which makes them durable), and high resistance to environmental influences. These gyros are immune to electromagnetic interference and can withstand large accelerations (42). An excellent treatment of the features, advantages, and disadvantages of RLGs versus FOGs is provided by Udd (2).

Typical drift rates are 0.01 deg/h (3). One similarity between RLGs and FOGs is that in both it has been relatively easy to achieve low performance, but much research has been needed to improve the performance to 0.01 deg/h levels. FOGs have been replacing RLGs in many civil and military applications, especially those with less demanding drift rates of approximately 1 deg/h. Accuracy has been the main disadvantage of FOGs when compared with mechanical gyroscopes and RLGs. Within a few years, however, FOGs are expected to achieve comparable drift rates to RLGs. Target values for precision FOGs are 0.001 deg/h drift rate, 0.0005 deg/h^{1/2} random walk, 5 ppm scale factor, and 500,000 h stability (51).

Miniaturized Gyroscopes on Chip. A number of products combining integrated optics and fiber optics have been developed recently. The cost of medium-performance FOGs of accuracy range 1 to 10 deg/h can be reduced by using an integrated optoelectronics module as the key component. The processes for fabricating the integrated optics components are similar to the batch processes used by the semiconductor industry to mass-produce integrated circuits. In addition, this technology allows

the use of inexpensive telecommunications optical fiber for the fiber coils. Such systems comprise a light source, its driver, a detector, a filter, an ADC, and an integrated optics coupler. Typically, all of the optical components are integrated optics, except for the fiber coil, light source, and optical detector–preamplifier and depolarizers.

The coils are mounted on a chassis along with the other components. The light source is typically a superluminescent diode or an edge-emitting LED, similar to those used in short-range telecommunications. These light sources have short coherence lengths and can be efficiently coupled into the core of a single-mode fiber. The low coherence is important for minimizing optical reflections and other disturbances that can affect performance and reliability. The fiber-optic couplers (normally 2.5 to 5 cm) and a phase modulator (normally 0.6 to 2.5 cm), when integrated, fit on a 2.5×2.5 cm integrated optics chip.

These optoelectronic chips are available in single-, dual-, and three-axis configurations for medium- and high-accuracy applications with bias performance in the 0.01 to 10 deg/h range. If the gyro is multiaxial, many of the optical and electrical components can be time-shared between the axes, resulting in significant reductions in cost, size, weight, volume, and number of parts. For example, a single chip is produced that allows all gyros to share a single laser diode, eliminating the need for three separate laser diodes. The reduction in the number of parts has the additional benefit of operation without mechanical gyro errors. An all-fiber gyro is less accurate than a fiber–integrated-optics gyro.

Many such devices incorporate intelligence on chip, and employ application-specific integrated circuits (ASIC) for operation, eliminating the need for a digital-to-analog converter (DAC). On-board processors take care of data handling, temperature compensation, scaling, and filtering. Such a chip can be designed with an operating wavelength between 630 and 3200 nm.

Optoelectronic gyroscopes can give rate outputs for angular accelerations up to 300,000 deg/s² and can withstand 30g shocks from six directions. Typical packaging volume is 100 cm³, and typical power consumption is 2 W.

In 1993, Hitachi Cable developed a business-card-size FOG, which integrates an optical coupler, polarizer, detector, and other components into an optical integrated circuit chip (52). The integrated circuit comprises a Y-branch glass optical waveguide and two phase-shift modulators. It connects a signal-processing circuit, sensing coils, and a fiber coupler to make up the phase-shift modulation gyro. The device is mainly used in car navigation systems and costs less than 10,000 Japanese yen.

Another device by Hitachi, named GyroAce V, which measures $80 \times 80 \times 35$ mm, is an analog-output gyro that uses a polarization surface protection optical fiber, which is not affected by temperature and vibration. The response speed of the device is 1 ms, the operating temperature ranges from -30 to 75 °C, and the price is about 200,000 Japanese yen.

The trend of the technology is towards greater integration, resulting in a gyro on chip in which the sensing element and the electronics to analyze it will be included in a single piece of silicon of size 1 cm², mass-produced with very low cost. One problem is the development of pigtail techniques (procedures to connect fiber to the integrated chip). This is challenging due to the alignment required. It is particularly important for higher-

accuracy (0.01 deg/h) FOGs. When polarization-maintaining fibers are employed, the major axes of these fibers need to be precision-aligned with the polarization axis of the waveguide and the light source.

Nuclear Magnetic Resonance Gyroscopes

Certain particles such as electrons, protons, neutrons, and nuclei intrinsically possess angular momentum. Employing this angular momentum as the basis for a gyroscope has long been considered, and such instruments have been suggested (53). All approaches suggested so far are based on nuclear magnetic resonance (NMR), where a net nuclear magnetization is established by some means (54).

The precession of the net nuclear angular momentum is observed in an applied magnetic field H_0 , from which rotation information is deduced. The frequency of precession about H_0 (the Larmor frequency) is proportional to the magnitude of H_0 : $w_0 = \gamma H_0$. The constant γ characterizes the particle used. If the frame is rotating at an angular rate w_r , then the observed frequency will be shifted: $w = \gamma H_0 - w_r$. For typical values of w_r for a practical navigational gyro, use of this equation would require very precise knowledge of H_0 .

Using two particles with different spins in the same magnetic field, two different Larmor frequencies w_1 and w_2 can be observed. The two unknowns, the magnetic field and the rate of rotation, can be found by solving the following pair of equations:

$$w_1 = \gamma_1 H_0 - w_r \quad (13)$$

$$w_2 = \gamma_2 H_0 - w_r \quad (14)$$

Any method based on these equations relies upon the constancy and knowledge of the gyromagnetic ratios γ_1/γ_2 . This dependence can be eliminated by the use of two magnetic fields in opposite directions and a resonance cell containing both kinds of nuclei in each magnetic field (5).

Factors characterizing the performance of an MRG are noise on the angle output, angle random walk, and bias effects. The drift rate is approximately a few hundredths of a degree per hour. Research on the development of this type of gyro is ongoing (55,56).

APPLICATIONS

Ships and Aircraft

Gyrocompass. A gyrocompass is an instrument that is used in ship or aircraft navigation to provide a fixed reference direction by pointing toward true or geographic north. The device is basically a special configuration of the rate-integrating fly-wheel gyroscope, employing a gravity reference for the north-seeking function. A compensated magnetic compass indicates magnetic north, which is different than true north. A gyrocompass, however, when properly adjusted, can be made to indicate true north. A gyrocompass is more reliable than a magnetic compass because it is not affected by magnetic disturbances such as those due to the steel hull of a ship. The gyrocompass is not dependent on the magnetic field of the earth and should not be confused with gyromagnetic compasses such as the Gyrosyn Compass described later.



Figure 20. MK 37 VT Digital Marine Gyrocompass by Sperry Marine Inc. (Photo courtesy of Sperry Marine Inc., Charlottesville, VA.)

Marine Gyrocompass. Almost every large ship carries at least one master gyrocompass installed in its own gyro room. An example is shown in Fig. 20. A transmission system links the master compass to *repeaters*, which are used on the ship for steering, positioning, and course recording. A marine gyroscope can take a number of different forms, all of which make use of the principle of the gyroscope and the pendulum. In its simplest form, the instrument consists of a rapidly spinning wheel, driven by a motor, and suspended so as to function as a pendulum and to respond to the earth's gravity. It is a three-frame gyroscope with its spin axis horizontal. As the earth rotates, the gyroscope finds an equilibrium position with its axis of spin parallel to the local meridian—that is, the axis assumes a north–south orientation. The gyrocompass incorporates damping features to control its motion as it aligns itself with the local meridian. It also needs to be compensated for the rolling, pitching and other motions of the ship on which it is placed.

In another form of a gyrocompass, a gimbal ring is attached to a support rod extending from the diameter of the ring. The axis of the spinning wheel and the support rod are perpendicular to each other. In most pendulum suspensions, a knife-edge is used as the top support point, restricting the motion of the pendulum and placing unwanted torques when a twist occurs. In this case, a ball floating in mercury is employed so that pendulum motion can take place freely in any vertical plane. This way, the support point can twist freely about a vertical line without placing an unwanted torque on the gyroscope pendulum.

In order to understand the north-seeking (or meridian-settling) property of the gyrocompass, consider a gyroscope at the equator whose axis is lying along the east–west direction. Since the gyroscope is supported as a pendulum, the support point, the center of mass of the gyroscope, and the center of the earth tend to lie along the same vertical line. However, as the earth rotates, the vertical line sweeps through space along the

equator. The gyroscope cannot maintain its initial orientation due to gravitational forces. A torque, or twist, is impressed on the gyro, with the torque vector lying along the north–south direction. As soon as a tilt develops, the pendulum introduces torques that precess the spin axis towards the meridian and tries to align the gyroscope axis in the same direction as the torque. The gyroscope precesses from its original westward direction until its axis also points along a north–south direction. Depending on the amount of damping, it follows a spiral with an ever-decreasing radius. When stabilized, the spin axis is maintained in the meridian plane by a precession equal but opposite to the drift at the particular latitude. Once the alignment with the north–south direction is complete, the earth's rotation cannot impress any further torque on the gyro, because the torque cannot be transmitted through the bearings of the spinning wheel when in that direction. As the ship that carries the gyrocompass continues its motion, the torque redevelops and the gyroscope realigns itself with the local meridian. A gyrocompass functions at other latitudes besides the equator also, but only the motion of the component of the vertical line that is perpendicular to the earth's axis can impress a torque on the gyrocompass. When there is no tilting effect, the marine gyrocompass will lose its directional properties and become useless. This is the case at the poles of the earth, where the vertical line is parallel to the earth's axis and no torque can be transmitted to the instrument. Thus, the sensitivity of the gyrocompass is maximum at the equator and decreases to zero as the north or the south pole is approached. Also, when a vehicle moves to the west with a speed equal to the surface speed of the earth, the gyrocompass loses its sensitivity. This device cannot be used for air navigation because this condition can easily occur in the middle and upper latitudes.

Aircraft Gyrocompass. Aircraft gyrocompasses are almost always of the magnetic type, stabilized by a gyroscope. These are based on automatically monitored directional gyroscopes spinning on a horizontal axis, in which the monitoring device senses the direction of the meridian and ensures that the gyroscope axis is maintained in this direction, pointing towards the magnetic north pole.

Gyrosyn Compass is the tradename for a magnetic compass developed by Sperry Gyroscope Company and that came into use following World War II. Unlike an ordinary magnetic compass, it is driven electrically, and it adjusts more quickly to course changes of the aircraft. The monitoring device consists of a magnetic sensing unit, called the *flux valve*, which is mounted in the wing tip or some other remote location on the plane. Any relative change between the aircraft and the earth's magnetic field results in a voltage signal in the flux valve, making allowance for variation in the direction of the earth's magnetic field. This voltage signal is amplified electronically and applied to the gyroscope to keep it pointing towards magnetic north. The device can operate repeater compasses wherever they are needed in the aircraft.

Gyrostabilizer. When its rotor is very heavy, a gyroscope can function as an actuator: a device that directly controls a vehicle. The stiffness of the rotor is used to maintain the orientation of vehicles such as spacecraft, ocean ships, and automobiles. A gyrostabilizer is a large gyroscope that counteracts the rolling motion of a ship at sea and makes travel more comfortable. The device is mounted on pedestals bolted to the ship's framework. A self-contained electric motor drives the rotor. A number of

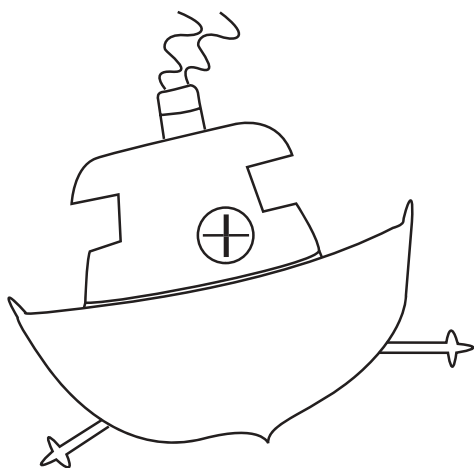


Figure 21. A ship stabilizer with fins controlled by a gyroscope. The gyroscope is unaffected by the rolling motion of the ship. It controls the pitch of the stabilizing vanes that counteract the rolling.

ships and yachts built between World War I and World War II used gyrostabilizers, which accounted for about $1\frac{1}{2}\%$ of the entire weight of a ship. Such large gyroscopes, by themselves, are no longer used to stabilize ships, because they would have to be very large and heavy in order to control the large ships of today. The US Navy uses them only aboard Polaris missile-firing submarines.

Modern stabilizers make use of a set of fins controlled by gyroscopes as illustrated in Fig. 21. The fins extend out from the ship's hull and are operated so that the forward motion of the ship produces a tilt in one direction on one fin, and in the opposite direction on the other fin. The gyroscopes sense the vertical angular displacement and the roll velocity and provide the proper control for the fins to oppose the rolling motion caused by waves beating against the ship.

Gyropilot. The gyropilot, also called an automatic pilot or autopilot, is a device which automatically steers ships or aircraft closer to a course than a human helmsman or pilot can. The gyropilot allows more accurate navigation and economical operation. If autopiloting can be done very accurately and reliably, unmanned aerial vehicles or ships can be realized. An autopilot uses the artificial horizon and the gyrocompass to operate the mechanisms controlling the rudder, elevators, and ailerons, thus allowing the craft to fly long distances without a human pilot.

In modern gyropilots, rate detection is the principal reference, and displacement detection plays a secondary role. Basically, the gyropilot consists of three devices. Each device detects deviations of the aircraft from its proper course and attitude in one plane and generates a corrective signal to correct for the disturbances. The corrective signals are voltage displacements, which are amplified and sent to servo units, which have small electric motors that move the aircraft's controls. The controls are: rudder control for azimuth and sudden change in heading disturbances (yaw), aileron control for roll disturbance, and elevator control for pitch disturbance. First, corrective rudder control is applied to the rudder servomotor. The roll disturbance is detected by a roll rate gyroscope and a roll angle pendulum, which senses displacement. The aileron servo applies corrective action. Pitch disturbance is detected by a pitch rate gyroscope

and a pitch pendulum. The elevator servo also applies corrective action.

The gyropilot for ships controls the rudder according to corrective signals from a gyrocompass. The first automatic pilot for ships was used in the early 1920s. The automatic pilot aboard ships is often called "iron mike" or the "iron quartermaster."

Other Aircraft Instruments. In aircraft, gyroscopes are used in several other instruments which are either part of the autopilot or are used for visual reference. The three primary gyroscopic instruments fitted to the flight panel are a rate-of-turn indicator (or turn-and-bank indicator), which is simply a rate gyro; a directional gyroscope; and an artificial horizon. These gyros may be driven either by electric motors or by air jets.

The directional gyroscope forms a standard reference for the pilot and navigator, indicating in what direction the aircraft is heading. It is a three-frame gyroscope with its spin axis on the horizontal plane. As soon as tilt develops, a switch is closed between the gyroscope housing and the vertical gimbal ring, and a motor introduces a torque in the horizontal plane that causes the gyroscope to precess back toward the horizontal.

The artificial horizon, or horizon indicator, indicates the orientation of the aircraft relative to the horizon and displays the rolling and pitching motion of the aircraft, without the pilot's having to look at the ground or horizon. It consists of a three-frame gyroscope with its spin axis vertical, and automatic correcting devices to counteract the apparent motion of the spin axis around the celestial pole and any other random precessions.

Inertial Navigation (Guidance) Systems. One of the most important uses of the gyroscope is in INSs for aircraft, spacecraft, or military equipment. These depend on the inertia of an extremely precise gyroscope to keep the craft traveling in exactly the right direction. Gyroscopes are the primary source of navigation and guidance information on spacecraft, since magnetic compasses are useless in outer space and there is no local horizon to sight on. Gyros are also used to accomplish the orientation control of the accelerometers, which are key components of INSs. INSs are nonradiating, nonjammable, dead-reckoning systems.

When INSs emerged for civil aviation in the 1970s, they were so large, heavy, expensive, and costly to maintain that they were used only on very large transoceanic transports. In the early 1980s, smaller, lighter, more reliable RLG-type gyroscopes were developed, and the civil market was expanded to a greater number of potential applications such as inertial navigation of larger business aircraft.

Modern INSs require a small platform very accurately stabilized by gyroscopes. This type of platform was perfected only in the 1950s after advances in the design of air-supported bearings and flotation gyroscopes. The inertial platform is extremely small and must be stabilized to an extraordinary degree of precision. The two types of self-contained INSs are the *gimbaled* INS and the *strapdown* INS. In a gimbaled system, the gyros and accelerometers are isolated from the rotations of the vehicle so that they can be maintained in a specific orientation with respect to the earth or inertial space. The stabilized INS physically represents the inertial reference frame. This greatly simplifies the position and velocity computations and reduces the dynamic-range requirements on the gyros. In a strapdown INS, the gyros measure the full rotation rates of the aircraft and

keep track of the instantaneous orientation of the accelerometers in order to properly integrate the accelerations into velocity and position. This is done by applying control torques to the gyros, causing them to precess at precisely the same rate as the combination of the earth's rate of rotation plus the angular rate of the vehicle's position with respect to the earth. This way, two of the accelerometers are locally leveled, and the third one is kept vertical at all times. Since computation of a reference attitude matrix is required, strapdown INSs were made possible only by the advent of small and speedy digital computers (3). Most of the modern aircraft and marine INSs today are of the strapdown type. The errors caused by vehicle maneuvers and accelerations can be larger in strapdown systems. For gimbal systems, the typical gyro drift rate is 0.01 deg/h; for strapdown systems, the error is 20 to 50% worse.

If single-degree-of-freedom gyros are to be used, three mutually orthogonal gyros are required for the north, east, and vertical coordinates. If two-degree-of-freedom gyros are to be used, only two gyros need be installed. The types of gyros currently found in INSs include floated rate-integrating gyros, electrostatically supported gyros, dynamically tuned rotor gyros, and RLGs (57).

Military

Gyroscopes have also been exploited for military applications such as guiding and controlling the automatic steering mechanisms of torpedoes, missiles, and projectiles, and smart and terminally guided or precision-guided munitions for large-caliber guns (or field artillery) on tanks and howitzers. Gyroscopes are also employed underwater to steer torpedoes toward the target regardless of underwater conditions.

Many of the most accurate long-range missiles are steered by INSs using gyroscopes. Conventional three-frame gyroscopes are used in ballistic missiles for automatic steering together with two-frame gyroscopes to correct turn and pitch motion. German engineers made significant advances in this field during the 1930s, and their knowledge was later used in the design of guidance systems for the V-1 flying bomb, a pilotless aircraft, and the V-2 rocket, an early ballistic missile. Most tactical-grade missile systems require an accuracy of 1 deg/h.

The ability of gyroscopes to define direction very accurately, when used in conjunction with sophisticated control mechanisms, led to the development of stabilized gunsights (or predictor sights), bombsights, antiaircraft systems, and platforms to carry guns and radar antennas on ships during World War II. The gyroscopic gunsight revolutionized aerial gunnery for jet fighter planes that automatically aim guns, rockets, and bombs. The method of stabilization used for gun platforms is essentially the same as the principle of stabilizing an inertial platform. The gyroscopes that detect platform displacement are not as accurate as the flotation type. The sight fitted on the gun contains a rate gyroscope capable of measuring angular velocities in two planes at right angles to each other. The gyroscope is a three-frame one constrained by horizontal and vertical springs to the inner and the outer gimbal, respectively. Sometimes, variable-strength magnetic fields are employed to constrain the rotor axle in azimuth and elevation instead of a mechanical spring arrangement. The field coils for producing the horizontal component of this magnetic field are coupled to the rangefinder. The current through the vertical coils is adjusted so that the field depends on the drop of the projectiles

due to gravity. The sensitivity of the gyroscope in the horizontal plane is a function of the sighting range; in the vertical, it is a function of the gravity drop. In operating the gunsight, the gunner holds the image of a central dot over the target while the gun is automatically aimed by the gyroscope at the predicted target position when the projectile motion is completed. This way, correct aiming to direct antiaircraft guns, guns of warships, and bombs to moving targets is provided.

Automotive

Among the automotive applications of gyroscopes are antiskid braking systems for vehicles and active suspension systems. In antiskid braking systems, a single-axis yaw-rate gyroscope senses any tendency of the vehicle to spin and sends signals to the braking and steering systems to prevent skidding (20). In active suspension systems, the springs and shock absorbers of conventional suspensions are replaced with computer-controlled hydraulic units for dynamic ride control. Gyroscopes find use in such systems as yaw-rate units to provide rate data, as well as instrumenting the vehicles for crash studies.

Robotics

Control of robotic tools in space by robot arms, machine control, development of autonomous warehouse and factory robotic systems and mobile platforms, as well as rough-terrain navigation, are some of the applications that immediately come to mind.

A number of robotic systems have been described which use some form of absolute sensing mechanisms for guidance (see Ref. 5 or 58 for surveys). Such systems typically rely on the availability of easy-to-see beacons or landmarks, using simple encoder information to predict vehicle location between sensing locations. This works well when the density of beacons or landmarks is high and the ground over which the vehicle travels is smooth. In cases where the beacon density is sparse or the ground is uneven, such systems can easily lose track of their position. This is particularly a problem for vehicles operating in outdoor environments. INSs can potentially overcome this problem. Inertial information can be used to generate estimates of position over significant periods of time independent of landmark visibility and the validity of encoder information. Clearly, positions derived from INS must occasionally be realigned using landmark information, but a system that combines both inertial and landmark sensing can cope with substantially lower landmark density and can also deal with terrain where encoder information has limited value.

INSs have been widely used in aerospace applications (5,59,60) and have been only recently exploited in robotics applications, where they have considerable potential. In work reported in Ref. 61, inertial sensors were used to estimate the attitude of a mobile robot. With a three-gyro, two-accelerometer configuration, experiments were performed to estimate the roll and pitch of the robot when one wheel climbs onto a plank using a small inclined plane.

One reason that INSs are widely used in aerospace applications but not as much in robotics is simply that high-quality aerospace inertial systems are too expensive for most robotics systems. However, low-cost solid-state INSs, motivated by the needs of the automotive industry, are increasingly being made available commercially. Although a considerable improvement on past systems, they clearly provide substantially less accu-



Figure 22. A low-cost INS comprising three gyroscopes, a triaxial accelerometer and two tilt sensors, developed for a mobile robot platform. (See B. Barshan and H. F. Durrant-Whyte, *IEEE Trans. Robot. Automa.* 11 (3), 328–342, June 1995. Copyright IEEE, 1995.)

rate position information than equivalent aerospace systems. However, such systems are at a point where, by developing reasonably detailed models of the inertial platform, they can provide valuable information in many robot positioning tasks (25,62). An inexpensive INS developed for a mobile robot platform is illustrated in Fig. 22.

In a robotics context, the primary motivation has been the need to develop a system capable of providing low-cost, high-precision, short-time-duration position information for large outdoor automated vehicles and mobile platforms. In particular, the interest has been in obtaining location information for short periods when the vehicle is not in contact with any beacon or landmark information. Rough terrain, variations in wheel radius, tire slip, and body deflection cause the encoder information to be unreliable for location estimation except over very short sample intervals. Inertial sensing offers a potential solution to this type of problem.

Positioning Systems

Another system that is potentially of great value for vehicle localization is the global positioning system (GPS) (63). It is a satellite-based radio navigation system that allows a properly equipped user access to useful and accurate positioning information anywhere on the globe. The fact that an absolute identification signal, rather than a direct measurement of range or bearing, is used to compute location means that measurements are largely independent of local distortion effects. The position accuracy that can be achieved with GPS is 5 m in the military band, and 50 m in the civilian band. However, using a technique known as differential GPS, in which a separate base receiver is employed, civilian accuracy may be improved to 5 m. Although this is not as good as can be achieved using high-frequency radar, it may still be adequate for some applications. It is also worth noting that the cost of GPS receivers is remarkably low (about \$1000). In Ref. 64, integration of GPS with INS is described for precision navigation in aerospace applications. The cost of these hybrid systems has been considerably reduced due to the dramatic drop in the cost of GPSs, the expected mass-

production cost being \$15,000, about one-third the price of a comparable system using RLGs.

A hybrid INS–GPS offers important benefits, especially for military applications. The accuracy required of an INS can be relaxed significantly when used in combination with a GPS receiver. Overall system performance can be improved by fusing INS data with GPS signals to correct for inertial drift errors. Although GPS offers higher accuracy, most military planners would not want to rely solely on GPS, since the GPS satellites can be attacked by the enemy and the received signals are vulnerable to short-term interference and jamming, unlike INS. The possibility of jamming increases the requirement for inertial accuracy. For weapons, in the event that the GPS were jammed after launch, the INS must be sufficiently accurate to complete the task by itself. Many hybrid INS–GPS systems need the INS for short signal blockages that occur during aircraft maneuvering, and for navigation through clutter or adverse weather conditions. As an example, an augmented GPS-based system outfitted with 0.1 deg/h FOGs can tolerate a loss of GPS signals for 10 min and still arrive within 30 m of its intended target (12).

Although GPS is excellent for navigation, it does not sense dynamic changes rapidly enough for air vehicle flight control systems. GPS provides location data but not in real time, since it relies on satellite transmission received at certain intervals. During the gaps between satellite updates, localization by GPS is impossible.

Development of such accurate positioning systems would allow unmanned autonomous transport vehicles, seekers, and intelligent-vehicle highway systems for land navigation (65). More widespread installation of hybrid INS–GPS systems on emergency vehicles such as police cars, ambulances, and fire engines is expected in the near future.

Industry and Mining

Gyros suitable for surveying tasks were developed in Germany between 1947 and 1949. Later, the instruments were improved for practical applications. Road mappers and surveyors find gyroscopes useful in recording curves and grades. Wellbore-logging systems in oil and gas exploration have used gyroscopic devices to control the orientation of the shaft to ensure that the drilling is performed in the correct direction (66). Survey gyros are mainly employed for geodesy, subterranean geometry, surveying tunnels and boreholes in underground mine exploration, and shield excavators used in tunnel construction. Frequently, the rate gyros or gyrocompasses are designed to allow the fitting of optical measuring instruments or to allow combination with a theodolite (gyro theodolite). In these applications, gyrocompasses are preferred, since magnetic compasses would be disturbed by metal deposits. Ligament-suspended surveying gyros attain accuracies of one minute of arc within about 2 to 5 min in the absence of disturbances (57).

Other Applications

Vertical three-frame gyros with pen recorder attachments are often used to analyze rolling and pitching movements of ships and rocking motions of trains and to record the conditions of railroad tracks. With a very accurate gyro, almost every fault in the level of tracks can be detected. Other motion measurement systems that employ gyroscopes are those used in sports

training or at remote buoys for long-term recording of wave motion.

Gyroscopes are utilized as steadiers, stabilizers, and jitter/motion compensators for equipment such as hand-held video camcorders, ground cameras, lenses in satellites, binoculars, and other optical instruments. On the larger scale, gyroscopes are used to stabilize and compensate for the random motion of millimeter-wave seekers in long-range rocket systems, communication or shipboard satellites, radar-tracking antennas, antennas on remotely piloted vehicles, and weapon platforms. They are also used for the stabilization of the line of sight in various instruments.

Geophysical applications include measurement of the earth's rotation, continental drift, and earthquakes. A recent example is a Zerodur glass RLG manufactured by Carl Zeiss (67). This gyro is capable of achieving a relative resolution of 0.1 ppm over long periods of time. This allows high accuracy in determining the fluctuations in the earth's rotation, which is of importance to geophysicists who use such fluctuations to study the earth's interior structure, continental drift, and the occurrence of earthquakes. A 1.2×1.2 m block of Zerodur glass with a thickness of 180 mm and a weight of approximately 600 kg comprises the body of the RLG. The use of Zerodur stabilizes the length of the laser beam path and guarantees the required measuring accuracy by virtue of the nearly zero thermal expansion coefficient of the material. Inside the Zerodur block, there are four 1 m long longitudinal bores for the laser beam. A closed square resonator is produced by four deflecting mirrors fitted at each of the corners of the block. The principle of operation is otherwise the same as other RLGs. The gyro is installed in a subterranean cave on the Banks Peninsula in New Zealand.

Another device for measuring the rotation of the earth, a superfluid gyroscope designed like an ac superconducting quantum interference device (SQUID), has been demonstrated by R. Packard and his colleagues at the University of California, Berkeley. In their demonstration, the metrologists measured earth's rotation with a precision of 0.5% (68). In a superfluid, fluid circulation is quantized, which can be exploited to measure tiny rotations. In the Berkeley experiment, the flow of superfluid helium through a ring-shaped vessel is interrupted by a very thin barrier containing a sub- μ m-sized pinhole. When the vessel is rotated, the helium must squirt back through the hole to maintain constant circulation in space. By monitoring the squirting, it is possible to measure the rotation of the earth.

Researchers at Stanford University and NASA have developed an experiment, known as Gravity Probe B, involving very precise gyroscopes for testing two unverified predictions of Albert Einstein's general theory of relativity (69–71). The experiment is based on measuring how space and time are affected by the presence of the earth ("the geodetic effect"), and how the earth's rotation drags space-time around with it ("frame dragging"). These effects are detected by precisely measuring tiny changes in the direction of spin of four gyroscopes. The gyroscopes will be contained in an earth satellite orbiting at an altitude of 400 miles directly over the poles. Since the gyroscopes are free from any disturbance, they provide an almost perfect space-time reference system.

Gyroscopes have also found some applications in medicine. The alignment and rotational motion-sensing properties of gyros can be used for diverse medical applications such as precise positioning of surgical instruments, analyzing back motions for

orthopedic diagnosis, orthotics, and prosthetics, and measurement of human movement and diagnosis of motion disorders. For example, measuring the tremors associated with Parkinson's disease aids in diagnosis and treatment of the disease.

Enhanced pointing technology for computers has been developed in the form of a three-dimensional mouse equipped with a very small gyroscope (the GyroEngine), which is no longer restricted to the desktop and can be operated in free space (72,73). This is useful for three-dimensional applications such as airplane design.

Other miscellaneous applications include flight simulators, flight training systems, measurement of wave motion, and development of smart tools such as carpenter's levels for instrumentation.

Ref. 74 provides a detailed literature and patent survey on gyroscopes and their applications.

FUTURE TRENDS

Current gyro technology is progressing along several fronts. On one front, traditional spinning gyros are made smaller, lighter, and more reliable. In some cases, optics is integrated with mechanics, resulting in optomechanical gyroscopes. However, new types of gyroscopes with no moving parts have greater potential importance. Conventional flywheel gyroscopes are being replaced by solid-state gyroscopes with no moving parts that are smaller, less expensive, more versatile, rugged, accurate, and reliable. Computer chips are replacing mechanical components, which wear out and are expensive to fabricate. Advances in optical fiber research and solid-state technology have a direct impact on the development of modern gyroscopes. It is expected that by the turn of the century, gyroscopes will be more widely used in the control systems of automobiles, trucks, and emergency vehicles, among many other applications.

BIBLIOGRAPHY

1. I. Cochin (ed.), *Analysis and Design of the Gyroscope for Inertial Guidance*, New York: Wiley, 1963.
2. E. Udd (ed.), *Fiber Optic Sensors Based on the Sagnac Interferometer and Passive Ring Resonator*, New York: Wiley, 1991, pp. 233–269.
3. C. Lin, *Modern Navigation, Guidance, and Control Processing*, Englewood Cliffs, NJ: Prentice Hall, 1991.
4. Anonymous, DARPA boosts interferometric fiber-optic gyros, *Opt. Mater. and Eng. News*, 1 (7): Mar. 1991.
5. M. M. Kuritsky and M. S. Goldstein (eds.), Inertial navigation, in I. J. Cox and G. T. Wilfong (ed.), *Autonomous Robot Vehicles*, New York: Springer-Verlag, 1990.
6. E. F. Carter (ed.), *Dictionary of Inventions and Discoveries*. New York: Crane, Russak, 1966.
7. G. M. Sagnac, L'éther lumineux démontré par l'effet du vent relatif d'éther dans un interféromètre en rotation uniforme, *C. R. Acad. Sci.* 95: 708–710, 1913.
8. G. J. Martin, Gyroscopes may cease spinning, *IEEE Spectrum*, February 1986, pp. 48–53.
9. Merton Allen Associates, Design, uses, and performance of fiber optic gyroscopes, technical report 1, Merton Allen Associates, Silver Spring, MD, January 1997.
10. V. Vaili and R. W. Shorthill, Fiber ring interferometer, *Appl. Opt.*, 15: 1099–1100, 1976.

11. P. J. Klass, Fiber-optic gyros now challenging laser gyros, *Aviation Week & Space Technol.*, **145** (1): 62, July 1996.
12. M. K. Reunert, Fiber optic gyroscopes: Principles and applications, *Sensors*, August 1993, pp. 37–38.
13. J. Fraden (ed.), *AIP Handbook of Modern Sensors*, New York: Amer. Inst. of Physics, 1993.
14. Anonymous, Injection-molded optical gyroscope offers cost, size, weight, savings, *Aviation Week & Space Technol.*, **135** (16): 100, October 1991.
15. Anonymous, New miniature gyroscope developed using optical sensors, *Opt. Materials and Eng. News*, **2** (3): November 1991.
16. B. D. Nordwall, Quartz fork technology may replace INS gyros, *Aviation Week & Space Technol.*, **140** (17): 50, April 1995.
17. S. D. Orlosky and H. D. Morris, A Quartz Rotational Rate Sensor, *Sensors*, February 1995.
18. Vital Information Publications, Sensor markets and technologies update: microminiature solid state quartz gyro is tuned into growth applications, *Sensor Bus. Dig.* **2** (12): September 1993.
19. Anonymous, New quartz IMUs may challenge fiber optic gyro market, *Mil. and Commerical Fiber Business*, **7** (19): September 1992.
20. T. Shelley and J. Barrett, Vibrating gyro to keep cars on route. Eureka on campus, *Eng. Materials and Design*, **4** (2): 17, Spring 1992.
21. B. Dance, Piezoelectric ceramic elements from compact gyroscope, *Design News*, September 20, 1993, pp. 113–115.
22. S. Fujishima, T. Nakamura, and K. Fujimoto, Piezoelectric vibratory gyroscope using flexural vibration of a triangular bar. In *Proc. Frequency Control Symp.*, 29 May 1991.
23. Murata, Gyrostar piezoelectric vibrating gyroscope (product literature), Smyrna, GA: Murata Electronics North America, 1994.
24. Murata, Gyrostar piezoelectric vibrating gyroscope: Test and reliability data (technical manual, catalog no. T-03-B), Smyrna, GA: Murata Electronics North America, 1994.
25. B. Barshan and H. F. Durrant-Whyte, Evaluation of a solid-state gyroscope for robotics applications, *IEEE Trans. Instrum. Meas.*, **44**: 61–67, 1995.
26. E. O. Schulz-DuBois, Alternative interpretation of rotation rate sensing by ring laser, *IEEE J. Quantum Electronics*, **QE-2**: 299–305, 1996.
27. W. W. Chow et al., The ring laser gyro, *Rev. Modern Phys.*, **57** (1): 61–104, January 1985.
28. J. G. Koper, A three-axis ring laser gyroscope, *Sensors*, March 1987, pp. 8–21.
29. S. Chao, W. L. Lim, and J. A. Hammond, Lock-in growth in a ring laser gyro, In *Proc. Physics and Optical Ring Gyros Conf.*, Snowbird, UT, January 1984, Proc. SPIE 487, pp. 50–57.
30. N. Buholz and M. Chodorow, Acoustic wave amplitude modulation of a multimode ring laser, *IEEE J. Quantum Electronics*, **QE-3**: 454–459, 1967.
31. J. Chesnoy, Picosecond gyrolaser, *Opt. Lett.*, **14** (18): 990–992, September 1989.
32. W. R. Christian and M. J. Rosker, Picosecond pulsed diode ring laser gyroscope, *Opt. Lett.*, **16** (20): 1587–1589, October 1991.
33. M. L. Dennis, J. M. Diels, and M. Lai, Femtosecond ring dye laser: A potential new laser gyro, *Opt. Lett.*, **16** (7): 529–531, April 1991.
34. Fiber optic rotation sensors and related technologies, in S. Ezekiel and H. J. Arditty (ed.), *Proc. First Int. Conf.*, MIT, Springer-Verlag: New York, 1982.
35. F. Aronowitz, *The Ring Laser Gyro*, Vol. 1, New York: Academic Press, 1971.
36. L. N. Menegozzi and W. E. Lamb, Theory of a ring laser, *Phys. Rev. A*, **1** (4): 2103–2125, October 1973.
37. J. R. Wilkinson, *Ring Lasers*, Vol. 11, Pergamon Press, Oxford, UK, 1987.
38. G. A. Sanders, Critical review of resonator fiber optic gyroscope technology, in E. Udd (ed.), CR44, SPIE Opt. Eng. Press, Bellingham, WA, September 1992.
39. S. Tai et al., All-fibre gyroscope using depolarized superluminescent diode, *Electron. Lett.*, **22**: 546, 1986.
40. H. C. Lefevre, The interferometric fiber-optic gyroscope, in E. Udd (ed.), CR44, SPIE Optical Engineering Press, Bellingham, WA, September 1992.
41. D. A. Nolan, P. E. Blaszyk, and E. Udd, Optical fibers, in Ref. 2, pp. 9–36.
42. P. J. Klass, New device cuts cost of fiber-optic gyros, *Aviation Week & Space Technol.*, **145** (20): 62, November 1996.
43. P. Adrian, Technical advances in fiber-optic sensors: Theory and applications, *Sensors*, September 1991, pp. 23–45.
44. H. R. Everett, *Sensors for Mobile Robots*, Wellesley, MA: AK Peters, 1995.
45. Hitachi, Fiber optic gyroscope (HOFG-1), specification SP 94-28-1005, Tokyo: Hitachi Cable, Ltd., August 1994.
46. Hitachi, Fiber optic gyroscope (HGA-D), specification SP 94-28-1006, Tokyo: Hitachi Cable, Ltd., August 1994.
47. Anonymous, Andrew launches fiber gyro-based land navigation system, *Sensor Business Digest*, **5** (8): June 1996.
48. S. Emge et al., Reduced Minimum Configuration Fiber Optic Gyro for Land Navigation Applications, *Proceedings of Fiber Optic Gyros: 20th anniversary conference* (SPIE): Denver, CO, August 5–6, 1996.
49. S. Bennett, S. R. Emge, and R. B. Dyott, Fiber optic gyros for robotics, *Service Robot: An Int. J.* **2** (4): 1996.
50. J. Blake et al., Design, development, and test of a 3-inch open loop all fiber gyro, in *Proc. 14th Biennial Guidance Test Symp.*, MSD-TR-89-21, Holloman AFB, NM, October 1989, 255–266.
51. Pasha Publications, Inc., Fiber optic gyro, *Defense and Aerospace Electron.*, **3** (44): November 1993.
52. Anonymous, Hitachi Cable develops business card-size optical fiber gyro, *Hitachi Weekly*, 9 (46), November 1993.
53. J. H. Simpson, *Electron. Lett.*, **2**: 42, October 1964.
54. B. Cagnac, *Ann. Phys.*, **6**: 1961.
55. P. Harle, G. Wackerle, and M. Mehnnng, A Nuclear-Spin Based Rotation Sensor using Optical Polarization and Detection Methods, *Appl. Magn. Res.*, **5** (2): 207–220, 1993.
56. S. Appelt, G. Wackerle, and M. Mehnnng, Deviation from Berry Adiabatic Geometric Phase in a XE-131 Nuclear Gyroscope, *Phys. Rev. Lett.*, **72** (25): 3921–3924, 1994.
57. R. S. Sindlinger, North-seeking gyros and their application for land vehicle navigation, lecture at the Carl Cranz Course, "Principles and Performance of Mechanical and Optical Inertial Sensors, Intertial and Satellite Navigation Systems," Braunschweig, Germany, October 1992.
58. J. J. Leonard and H. F. Durrant-Whyte, *Directed Sonar Navigation*, London: Kluwer Academic, 1992.
59. C. T. Leondes (ed.), *Theory and Applications of Kalman Filtering*, London: Technical Editing and Reproduction, 1970.
60. D. A. Mackenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*, Cambridge, MA: MIT Press, 1990.
61. J. Vaganay and M. J. Aldon, Attitude estimation for a vehicle using inertial sensors, in D. Charnley (ed.), *Preprints of the 1st IFAC International Workshop on Intelligent Autonomous Vehicles*, Southampton, Hampshire, U.K., Pergamon, 1993, pp. 89–94.
62. B. Barshan and H. F. Durrant-Whyte, Inertial navigation systems for mobile roots, *IEEE Trans. Robot. Autom.*, **11**: 328–342, 1995.

64. B. Tiemeyer and S. Vieweg, GPS/INS integration for precision navigation, technical report, Institute of Flight Guidance and Control, Technical University of Braunschweig, Hans-Sommer-Str. 66, D-3300 Braunschweig, Germany, 1992.
65. Anonymous, Vehicle navigation drives opportunities for fiber gyros, *Sensor Business Digest*, **5** (1): 1995.
66. Anonymous, Fiber optic gyros for wellbore logging, *Enhanced Energy Recovery & Refining News*, **18** (8): March 1995.
67. Anonymous, Carl Zeiss Makes Ring Laser Gyro for Measuring Continental Drift, *EuroPhotonics*, Laurin Publishing Co., Inc., February/March 1997, p. 15.
68. K. Schwab, N. Bruckner, and R. E. Packard, Detection of the earth's rotation using superfluid phase coherence, *Nature*, **386** (6625): 585–587, April 1997.
69. G. Taubs, The Gravity Probe, *Discover*, **18** (3): 62–71, March 1997.
70. B. Ianotta, Music of the Spheres, *New Scientist*, 28–31, August 1996.
71. A. Lawler, Earthly Politics Boosts Space Probe, *Science*, **267**: 1756–1758, March 1995.
72. Business Communications Company, Inc., New miniature gyroscope developed using optical sensors, *Optical Mater. Eng. News*, **2** (3): November 1991.
73. S. W. Bhargava, Putting a new spin on the gyroscope, *Business Week*, **1** (3228): 1401, November 1991.
74. F. W. Cousins, The anatomy of the gyroscope: a report in three parts comprising a literature and patent survey directed to the gyroscope and its applications, in J. L. Hollington, (ed.), Technical Report AGARDograph Series No: 313, NATO Advisory Group for Aerospace Research and Development, 1988.

BILLUR BARSHAN
Bilkent University

HYGROMETERS

INTRODUCTION

The presence or absence of water vapor in air or other gas influences a vast range of physical, chemical, and biological processes. Humidity measurements are important for a wide range of industries and to some they are a critical aspect of business costs, product quality, and health and safety.

There are many different techniques for measuring humidity. The subject is also complicated by the confusing variety of ways of expressing humidity. As for other fields of measurement, a consistent and harmonized approach to humidity measurement is increasingly important, as is a common understanding of humidity terms and definitions.

Humidity is a relatively difficult quantity to measure in practice, and the accuracy achievable is not as good as for many other areas of measurement. For example, the mass of objects can be found by weighing in a typical laboratory to within 1 part in 100,000. Similarly, atmospheric pressures can often be measured to within 5 parts in 10,000. Humidity, however, can commonly be measured to only about 3 parts in 100, that is, with an uncertainty in the result of $\pm 3\%$.

To make a reliable humidity measurement at any given level of accuracy, it is necessary to use an appropriate measuring technique in a sound way. To do so, one must be able to define the measurement task, select a suitable method of measurement underpinned by reliable calibration, carry out the

measurement correctly without introducing errors, and interpret the result of the measurement.

Water in Air or Gases

The word “humidity” denotes the presence of water vapor in air or other gas. Water vapor is the gaseous form of water and can be thought of much like any other kind of gas. It is normally transparent, and makes up about one hundredth (or 1%) of the air around us.

Humidity arises in practice because, in the same way that hot water gives off steam, so water at lower temperatures—including ice—also gives off water vapor. Wherever there is water or ice, there is evaporation (or its opposite, condensation). The extent to which this happens depends on a number of factors, the most important of which is temperature. Similarly, other liquid or solid materials—most of which have some water content—will give off (or sometimes soak up) water vapor. Of course, water vapor can also be found in places where there is no liquid or solid nearby, for example, in remote parts of the Earth's atmosphere.

Air (or empty space, or any other gas) has a given capacity to absorb water vapor. This capacity depends mainly on temperature. Generally speaking, the hotter the air, the more water vapor it can hold. At any particular temperature, air that contains its full capacity of water vapor, is said to be “saturated.” The “relative humidity” of the air expresses how fully saturated it is with water vapor. A variety of other (“absolute”) measures are used to express how much water vapor is actually present (regardless of temperature or level of saturation). Definitions and explanations of the most important of these terms are given in the next Section.

Humidity Terms and Definitions. Some vocabulary specific to humidity (and other common words with specialized meanings in this context) are as follows:

Absorption (of water vapor)—retention (of water vapor) by penetration into bulk of a material

Adsorption (of water vapor)—retention (of water vapor) as a surface layer on a material

Condensate—condensed material, such as liquid water or ice

Desorption—release of adsorbed or absorbed substance

Humidity—the presence of water vapor in air or other gas. Some people use the word *humidity* to mean relative humidity only. Strictly speaking, “humidity” also refers to all kind of absolute indications of humidity. For very low humidities, other more specific terms tend to be used.

Hygrometer—any instrument for measuring humidity

Hygroscopic—tending to absorb water vapor

Absolute humidity—The mass of water vapor present in unit volume of moist air of a given temperature and pressure. SI (metric) units are grams of water per cubic meter of air (g m^{-3}). Older references may be in terms of pounds per million cubic feet (lb ft^{-6}) or in grains per cubic foot (gr ft^{-6}) or in grains per cubic foot (gr ft^{-1}). (One grain ≈ 0.0648 gram.)

Dew point (or dew-point temperature)—The temperature at which dew, or condensation, forms on cooling a gas. This is, in effect, the temperature at which air becomes

saturated in equilibrium with water. Expressed in degrees Celsius ($^{\circ}\text{C}$) or occasionally in other units of temperature (e.g., in degrees Fahrenheit ($^{\circ}\text{F}$) in the United States).

Relative humidity—The ratio of the actual vapor pressure to the saturation vapor pressure over a plane liquid water surface at the same temperature, expressed as a percentage. This is commonly understood when the term “X percent relative humidity” is used.

Saturation vapor pressure (of water)—maximum pressure of water vapor that can exist at a given temperature. Expressed in units of pressure, such as in pascals (Pa), or in non-SI units such as millibars (mbar) or millimeters of mercury (mm Hg).

Vapor pressure—that part of the total pressure contributed by the water vapor. Expressed in units of pressure, such as in pascals (Pa), or in nonmetric units such as millibars (mbar) or millimeters of mercury (mm Hg).

TYPES

In the past moisture in the air or other gas was measured by mechanical hygrometers. Today there are a number of hygrometer types available, the following being the mostly frequently used: mechanical, wet- and dry-bulb, adsorption, dew-point, impedance, piezo electric, electrolytic and optical fiber hygrometers.

MECHANICAL HYGROMETERS

In a simple mechanical type of hygrometer the sensing element is usually an organic material that expands and contracts with changes in the moisture in the surrounding air or gas. The material used most often is human hair. Other materials may be paper, animal tissues, and wood. As illustrated in Fig. 1, a bundle of hair is held under slight tension by a spring, and a magnifying linkage actuates a pointer. The usually designed operating range of this type of hygrometer is from about 4°C to 38°C . For very short periods these hygrometers may be used down to 0°C and up to 60°C . When used within the range of usual room conditions, these instruments show very little drift in their measurement over periods of a year or more, but if exposed for any length of time to extremes of humidity or temperature, a permanent change in the

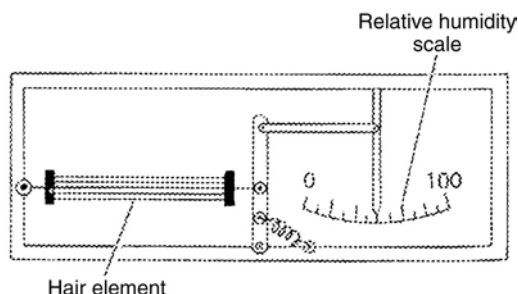


Figure 1. Hygrometer that uses hair as the sensing element.

sensitive material may result. The time required to respond fully to a sudden wide change of humidity may be as much as 5 min (1).

WET- AND DRY-BULB HYGROMETERS

A wet- and dry-bulb hygrometer consists of two matched temperature sensors over which humid air is drawn. One sensor is enclosed in a porous medium (wick or “wet sock”) which is maintained wet by capillary action from a reservoir of water. Water evaporates from the wick at a rate related to the humidity of the air. The evaporation causes the wet sensor to be chilled. The temperatures of the wet and dry sensors are used to calculate the humidity of the air. Relative humidity from vapor pressure is determined by Eq. (1):

$$\text{Relative humidity (in\%)} = \frac{p_d}{p_s} \times 100 \quad (1)$$

where p_d is the actual vapor pressure of water and p_s is the saturation vapor pressure of water at the prevailing temperature and barometric pressure (2). Water vapor pressure, p_d , in air is found from the actual or dry-bulb temperature t and the wet-bulb temperature t_w using the equation

$$p_d = p' - A \cdot P \cdot (t - t_w) \quad (2)$$

where p' is the saturation vapor pressure of water at temperature t_w , P is the total barometric pressure (in the same units), and A is a coefficient established for the psychrometer. Widely recognized values of A for t_w above 0°C are:

$A = 6.66 \times 10^{-4} \text{ }^{\circ}\text{C}$ for moving air as in the Assmann ventilated psychrometer

$A = 8.0 \times 10^{-4} \text{ }^{\circ}\text{C}$ in a Stevenson screen as used by the Meteorological Office

The values of A given above yield approximate values of humidity, to which corrections can be added or subtracted, on the basis of a traceable humidity calibration. Alternatively, on the basis of a calibration of a given psychrometer, an individual value or formula for A may be assigned for that particular instrument. The use of a particular value of A is only valid if the correct conditions of air flow are provided for the instrument in question. An aspirated psychrometer with a fan for adequate air flow over the thermometers is the recommended basis for this measurement. Other approaches, such as Mason’s (nonaspirated) psychrometers and whirling (hand aspirated) psychrometer, are less successful realizations of this principle and are prone to errors, as are “home-made” psychrometers (2).

Psychrometer

The psychrometer consists of the wet- and dry-bulb thermometers. Their temperature difference (psychrometric difference) enables the relative humidity to be calculated. The basis for

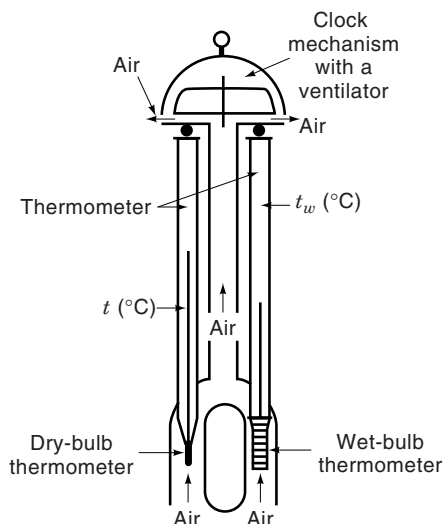


Figure 2. Assmann's aspirated psychrometer (4).

calculation is Sprung's psychrometric approximate formula (3).

$$p_d = p_f - k \cdot (t_d - t_f) \cdot p \text{ in mbar}$$

t_d = the temperature of the dry-bulb thermometer in °C

t_f = the temperature of the wet-bulb thermometer in °C

p_d = the partial vapor pressure of water in mbar

p_f = the vapor pressure at the temperature of the wet-bulb thermometer in mbar

$$k = \text{constant} = 0.61 \times 10^{-3} \text{ deg C}^{-1} \text{ for water/air} \\ = 0.57 \times 10^{-3} \text{ deg C}^{-1} \text{ for ice/air}$$

We can obtain a much better result if the measured air passes by the wet-bulb thermometer with a velocity of approximately 2 m/s. In case of nonventilated air the measurement is very inaccurate. For practical measurement the Assmann's aspirated psychrometer is the most widely used. Figure 2 shows its construction which uses a small clock mechanism with ventilator (nowadays the electric ventilator) to artificially create the air flow.

Relative humidity may be determined by comparing temperature readings of wet- and dry-bulb thermometers as in Fig. 3 (4). Table 1, extracted from more extensive US National Weather Service tables, gives the relative humidity as a function of air temperature t_d (dry bulb) and the difference $t_d - t_w$ between dry- and wet-bulb temperatures (t_w = wet-bulb temperature) (5). The data assume a pressure near normal atmospheric pressure and an instrumental configuration with forced ventilation.

The psychrometer is principally suited to large spaces and medium or high humidities. The uncertainty of measurement is around $\pm 2\%$ rh to $\pm 5\%$ rh, at best.

Following are advantages (+) and disadvantages(–) of wet- and dry-bulb hygrometers:

- + Psychrometers are relatively simple, cheap, reliable, and robust instruments.

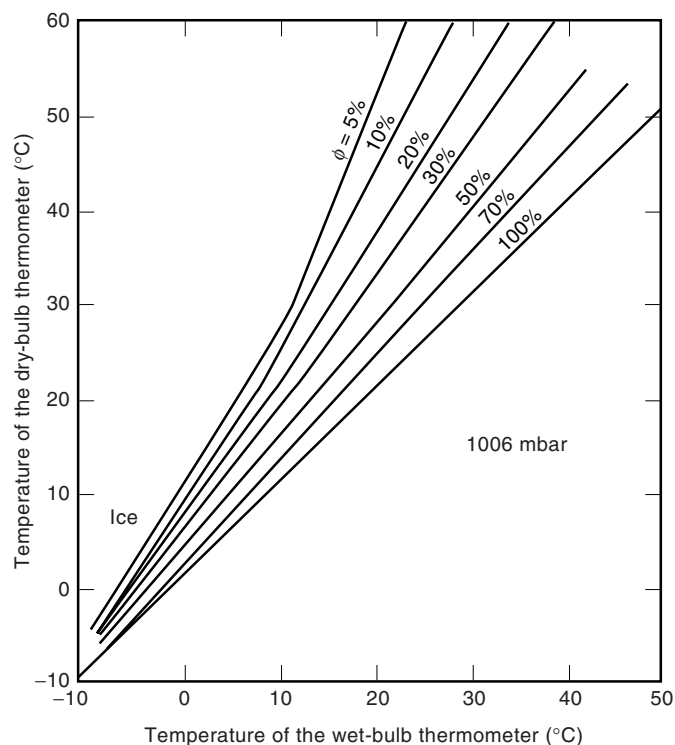


Figure 3. Psychrometric diagram for determining the relative humidity of the air from the reading of the temperature of wet- and dry-bulb thermometers.

- + They can have good stability.
- + They tolerate high temperatures and condensation.
- Some skill is usually required to use and to maintain the instrument.
- Results may have to be calculated from temperature readings (although some are automatic).
- A large air sample is required for the measurement.
- The sample will be humidified by wet-sock evaporation.
- Airborne particles or water impurity may contaminate the wick; this can and should be remedied by regular cleaning or replacement.
- Measurement can be complicated below about 10°C because of the uncertainty regarding the presence of water or ice on the wick when the temperature falls below 0°C.

ADSORPTION-TYPE HYGROMETERS

Adsorption-type humidity measurement is based on the phenomenon that a hygroscopic material gives off water according to absorption–desorption hysteresis as shown in Fig. 4 (6).

Centrifugal Method

This method is based on a well-known desorption method of relative humidity determination in alkaline solutions. In the case of cellulose, for instance, the cellulose fiber is first plunged into hydrochloric acid. With the help of centrifugation

Table 1. Determination of Relative Humidity from Wet- and Dry-Bulb Temperatures (5)

$t_d - t_w$ (°C)												
t_d (°C)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
-10	83	67	51	35	19							
-8	86	71	57	43	29	15						
-6	88	74	61	49	37	25	8					
-4	89	77	66	55	44	33	23	12				
-2	90	79	69	60	50	40	31	22	12			
0	91	81	72	64	55	46	38	29	21	13	5	
2	91	84	76	68	60	52	44	37	29	22	14	7
4	92	85	78	71	63	57	49	43	36	29	22	16
6	93	86	79	73	66	60	54	48	41	35	29	24
8	93	87	81	75	69	63	57	51	46	40	35	29
10	94	88	82	77	71	66	60	55	50	44	39	34
12	94	89	83	78	73	68	63	58	53	48	43	39
14	95	90	85	79	75	70	65	60	56	51	47	42
16	95	90	85	81	76	71	67	63	58	54	50	46
18	95	91	86	82	77	73	69	65	61	57	53	49
20	96	91	87	83	78	74	70	66	63	59	55	51
22	96	92	87	83	80	76	72	68	64	61	57	54
24	96	92	88	84	80	77	73	69	66	62	59	56
26	96	92	88	85	81	78	74	71	67	64	61	58
28	96	93	89	85	82	78	75	72	69	65	62	59
30	96	93	89	86	83	79	76	73	70	67	64	61
35	97	94	90	87	84	81	78	75	72	69	67	64
40	97	94	91	88	85	82	80	77	74	72	69	67

$t_d - t_w$ (°C)												
t_d (°C)	6.5	7.0	7.5	8.0	8.5	9.0	10.0	11.0	12.0	13.0	14.0	15.0
4	9											
6	17	11	5									
8	24	19	14	8								
10	29	24	20	15	10	6						
12	34	29	25	21	16	12	5					
14	38	34	30	26	22	18	10					
16	42	38	34	30	26	23	15	8				
18	45	41	38	34	30	27	20	14	7			
20	48	44	41	37	34	31	24	18	12	6		
22	50	47	44	40	37	34	28	22	17	11	6	
24	53	49	46	43	40	37	31	26	20	15	10	5
26	54	51	49	46	43	40	34	29	24	19	14	10
28	56	53	51	48	45	42	37	32	27	22	18	13
30	58	55	52	50	47	44	39	35	30	25	21	17
32	60	57	54	51	49	46	41	37	32	28	24	20
34	61	58	56	53	51	48	43	39	35	30	26	23
36	62	59	57	54	52	50	45	41	37	33	29	25
38	63	61	58	56	54	51	47	43	39	35	31	27
40	64	62	59	57	54	53	48	44	40	36	33	29

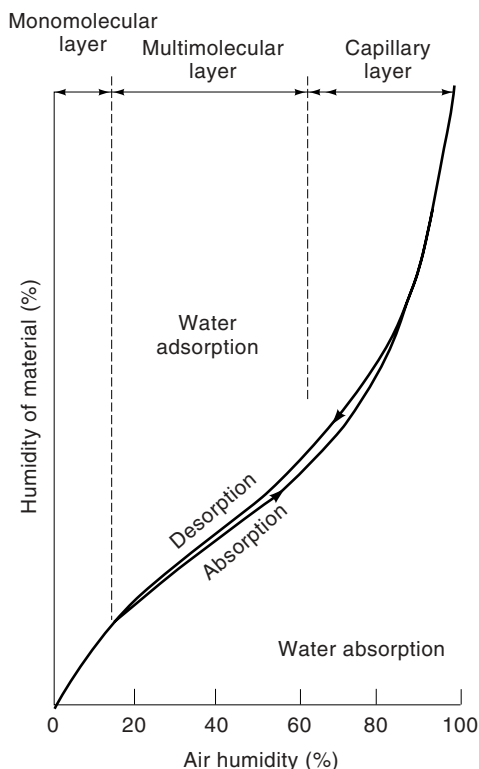


Figure 4. Absorption-desorption characteristics. The humidity of material exposed to increasing air humidity rises according to absorption curve. The humidity of material exposed to decreasing air humidity drops according to desorption curve.

(desorption method), humidity is adsorbed from fiber, as can be seen in Fig. 4. Thus the degree of humidity in cellulose fiber can be controlled with the rate of centrifugation (7,8).

Hygrometers Based on Adsorption of Ceramic

Adsorption of gases is an important property of solid surfaces. There are probably at least two different types of adsorption, namely, physisorption and chemisorption. Adsorption isobars often appear like those shown in Fig. 5 (9). Three or more regions are found for the adsorption of a particular gas on a particular solid.

At low temperatures of region a-b, physisorption occurs, where the surface coverage is at equilibrium and decreases with increasing temperature (Fig. 5).

At some temperatures in region b-c, irreversible chemisorption occurs. The activation energy increases with coverage, and the rate of adsorption decreases with coverage. If at each temperature the outgassed sample is permitted to adsorb gas for a given time, a curve shaped as indicated by the dashed line may be obtained. At this temperature, desorption will not occur by simply lowering the pressure. High-temperature outgassing is required (Fig. 5).

At temperatures high enough to overcome the activation energy (region c-d), reversible chemisorption is in principle reached, and the volume chemisorbed decreases with increasing temperature. However, behavior in this high-temperature region is often masked by chemical reactions (Fig. 5).

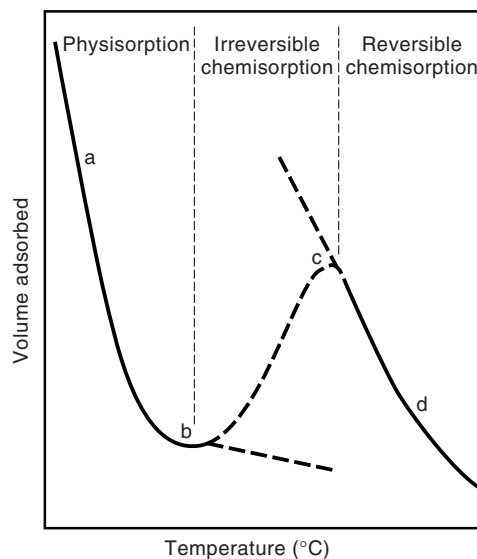


Figure 5. Adsorption isobar. The three adsorption regions of a particular gas on a particular solid (a-b physisorption, b-c irreversible chemisorption, c-d reversible chemisorption).

The surfaces of most metal oxides usually exhibit a complex structure comprising metal ions, oxygen ions, impurities, and imperfection. Thus there are many interactions between adsorbates and surface state. In atmospheres containing water vapor, for example, the surfaces of metal oxides are covered by hydroxyl groups on which water molecules adsorb through formation of hydrogen bonds. The surface hydroxyl groups can generally be removed by dehydration at higher temperatures. However, the surface structure of metal oxides in powder form is easily subject to permanent change by repeated heat-cleaning cycles at high temperatures. On the other hand, a ceramic body sintered at high temperature is essentially more stable physically, chemically, and thermally than the powder form. As with many porous systems, controlled-pore ceramics adsorb gases throughout the pore structure. The adsorption of gases generally tends to occur preferentially on the necks of the grain surfaces. The neck usually exhibits properties closely related to the crystal grain itself. In particular, a metal oxide ceramic semiconductor tends to exhibit changes in electrical conductivity due to the adsorption of gases on the necks.

DEW-POINT HYGROMETER

The relative humidity may be determined from the dew point at the temperature, t_d , at which the humidity in the air condenses when the mixture is cooled from an initial temperature t . A simple equation relating saturation vapor pressure and dew point is the Magnus formula (3). At a temperature t (°C), the saturation vapor pressure $e_w(t)$, in pascal, over liquid water, is

$$\ln e_w(t) = \ln 611.2 + \frac{17.62t}{243.12 + t} \quad (3)$$

The Magnus formula given above can be restated in alternative form, to derive dew point values from vapor pressure.

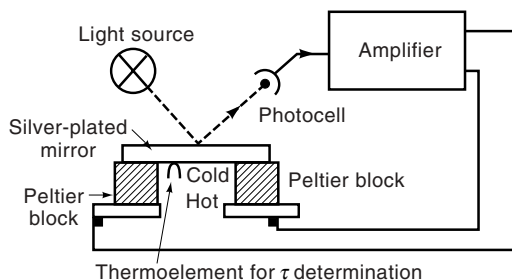


Figure 6. Dew-point hygrometer by a Peltier element. The appearance of dew on the metal mirror changes the reflected light flow.

For vapor pressure e_w (Pa) over water, the dew point t_d ($^{\circ}\text{C}$) is given by

$$t_d = \frac{243.12 \cdot \ln(e_w/611.2)}{17.62 - \ln(e_w/611.2)} \quad (4)$$

(100 Pa = 1 mbar). The uncertainty in values of t_d given by this equation is less than $\pm 0.04^{\circ}\text{C}$ in the range -45°C to $+60^{\circ}\text{C}$. For vapor pressure e_i over ice, the frost point t_i (in $^{\circ}\text{C}$) is given by

$$t_i = \frac{272.62 \cdot \ln(e_i/611.2)}{22.46 - \ln(e_i/611.2)} \quad (5)$$

The uncertainty in values of t_i given by this equation is less than $\pm 0.08^{\circ}\text{C}$, in the range -65°C to $+0.01^{\circ}\text{C}$ (10).

Dew-Point Hygrometer Using a Peltier Element

In modern dew-point hygrometers a small silver-plated mirror is placed in the measured air flow and cooled with the help of a Peltier element (Fig. 6). The latter, together with the amplifier and the thermoelement, enables the dew-point temperature to be measured. The appearance of dew on the metal mirror changes the reflected light flow. This change has influence on the amplifier and consequently on the dc current of the Peltier element. As soon as the dew point is formed, we can measure the dew-point temperature. The setting time of the measuring device is only a few seconds. A dew-point hygrometer using a Peltier element can also be used for aggressive and other condensing gases (11).

LiCl Hygrometer

The sensing medium, which is a hygroscopic salt (LiCl), absorbs water from the air. An electrical voltage is applied across the salt, and a current passes according to the amount of water vapor that has been absorbed. The insulating tube is covered with a glass wool and then coiled up with a wire in the shape of a double screw (see Fig. 7). The wire warms the thermometer placed inside the tube. The glass wool is moistened with LiCl, which also contains water molecules. The current running through the conducting moistened glass wool warms the coil. This causes the lowering of the glass wool conductivity (due to water evaporation), until the partial pressure of water vapor does not exceed the equilibrium point. The temperature of the thermometer, too, exceeds the equilibrium temperature. The reduction of water in LiCl salt reduces the glass wool conductivity and the thermometer warming. Consequently, the temperature of the thermometer falls, LiCl absorbs the humidity again, the

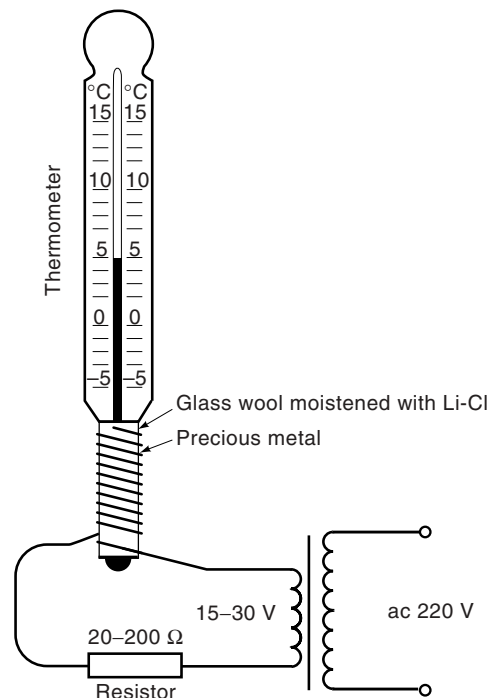


Figure 7. LiCl dew-point hygrometer.

conductivity increases, and as a result a higher current is once more obtained. This current again warms the coil and thus reduces the humidity in the LiCl salt. This process repeats until the equilibrium dew point is achieved. The temperature at which this occurs is related to the water vapor pressure. The measuring range is -20°C to $+60^{\circ}\text{C}$ (dew-point temperature). Long-term stability is within $\pm 0.5^{\circ}\text{C}$ to 1°C of the dew-point temperature. The electrical power used for this process is 1 W. The setting time is approximately 1 minute (1–3). The instrument is usually in probe form, with reading displayed in terms of dew point.

Conversion Between Dew Point and Relative Humidity

There is no simple direct formula for converting in either direction between dew point and relative humidity. Conversion between these two parameters must be carried out via the intermediate step of evaluating both the actual vapor pressure of water and the saturation vapor pressure at the prevailing temperature. Values of relative humidity or dew point can then be calculated using the appropriate formula above. Table 2 shows values of relative humidity at selected temperatures and dew points.

Table 2 gives relative humidity as a function of the dew-point depression t_d for several values of the dew point. This table is used to estimate relative humidities between the values listed; the uncertainty may be up to $\pm 1\%$ or $\pm 2\%$. Small differences in air pressure within the normal atmospheric range will affect these figures by less than 0.1% relative humidity (rh) (10,14).

IMPEDANCE-TYPE HYGROMETERS

This general type of sensor is fabricated from a hygroscopic material, and the electrical properties are altered as it absorbs

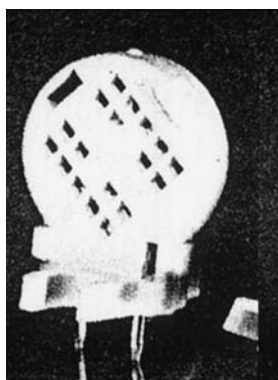
Table 2. Values of Relative Humidity at a Selection of Temperature and Dew Points (12,13)

Temperature: (°C)	0	5	10	15	20	25	30	35	40	50	60	70	80	90	100
Dew Point (°C)	Relative Humidity (%)														
0	100	70.1	49.8	35.8	26.1	19.3	14.4	10.9	8.3	4.9	3.1	2.0	1.3	<1	<1
5		100	71.1	51.1	37.3	27.5	20.5	15.5	11.8	7.1	4.4	2.8	1.8	1.2	<1
10			100	72.0	52.5	38.7	28.9	21.8	16.6	9.9	6.2	3.9	2.6	1.7	1.2
15				100	72.9	53.8	40.2	30.3	23.1	13.8	8.6	5.5	3.6	2.4	1.7
20					100	73.8	51.1	41.6	31.7	18.9	11.7	7.5	4.9	3.3	2.3
25						100	74.6	56.3	42.9	25.7	15.9	10.2	6.7	4.5	3.1
30							100	75.4	57.5	34.3	21.3	13.7	9.0	6.1	4.2
35								100	76.2	45.6	28.2	18.0	11.9	8.0	5.6
40									100	59.8	37.0	23.7	15.6	10.5	7.3
50										100	61.9	39.6	26.1	17.6	12.2
60											100	63.9	42.1	28.4	19.7
70												100	65.8	44.5	30.8
80													100	67.6	46.8
90														100	69.2

water molecules. Changes in humidity are measured as changes in the sensor's electrical capacitance or resistance or some combination of the two. The majority of modern impedance sensors use thin-film technology. Impedance hygrometers are usually also fitted with temperature sensors. Readings are displayed directly, sometimes with a choice of units (e.g., relative humidity or dew point), and output of an electrical signal may also be available (15). There are several distinct types of electrical sensors.

Capacitive Sensors

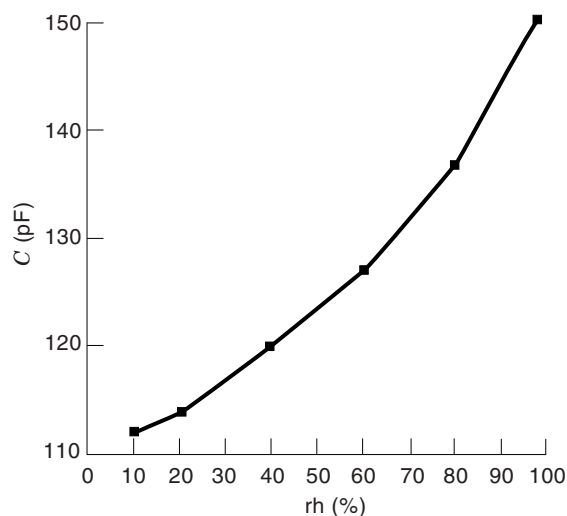
Capacitive sensors respond most closely to relative humidity, rather than dew point, with best linearity at low relative humidities. Figure 8 shows the capacitive sensor for measuring humidity. It consists of a perforated housing inside of which

**Figure 8.** Capacitive sensor for humidity measurement.

there is a thin hygroscopic film layer coated with gold. This thin film layer represents acts as a dielectric for the capacitor. Under the influence of humid air, the dielectric constant of the thin film layer and consequently the capacitance of the capacitor change, as can be seen in Fig. 9. With the help of a simple measuring circuit we obtain the change of the voltage, which depends on the relative humidity (16).

Resistive Sensors

Resistive sensors correspond most closely to relative humidity hygrometers, rather than to dew point hygrometers. The

**Figure 9.** The dependence of sensor capacitance on the relative humidity.

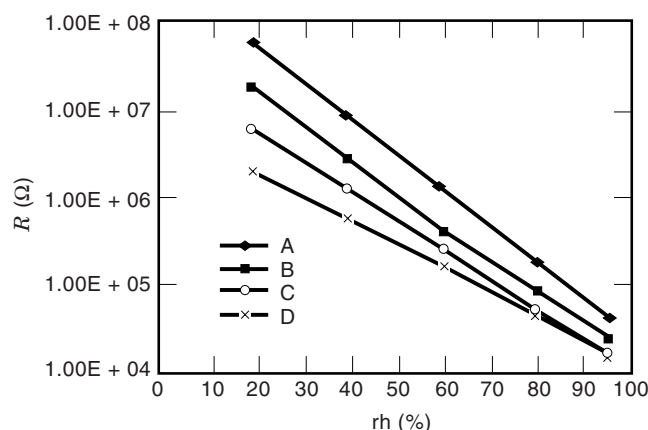


Figure 10. Humidity characteristics of the thick film resistive sensor for various molar ratios of $\text{TiO}_2/\text{SnO}_2$ (curves A, B, C, and D correspond to 2.9:1, 1.9:1, 1:1, and 0.5:1, respectively).

linearity of resistive sensors is best at high humidities. Most resistive sensors cannot tolerate condensation. However, some are “saturation guarded,” with automatic heating to prevent condensation. Figure 10 shows the humidity characteristic of devices at 40°C for various mole ratios of $\text{TiO}_2/\text{SnO}_2$. The mixture in mole ratio of $\text{TiO}_2:\text{SnO}_2 = 2.9:1$, $1.9:1$, $1:1$, or $0.5:1$ was sintered at 1200°C for 2 h with Ta_2O_5 (several mole %). The results indicate that the electrical resistance of the device in any mole ratio decreases exponentially with an increase in rh, and the increase in mole ratio of $\text{TiO}_2/\text{SnO}_2$ causes an increase in the whole device resistance and moisture-sensitivity (17).

Dew-Point Type Impedance Sensors

Dew-point type impedance sensors are a special case of impedance hygrometer, used to measure in absolute units rather than relative humidity. Following a similar general principle, the sensor may feature aluminum oxide or other metal oxides, or a silicon base for the active element. This type of sensor responds to the partial pressure of water vapor. Commonly the signal is converted into other absolute units, resulting in values displayed by the instrument in dew point or parts per million by volume. These sensors can have a wide range of measurement, including very dry gases. They are often used at high pressure, or to sample gas originating from the system at high pressure (15). Figure 11 shows the change of impedance in an Al_2O_3 sensor. The latter consists of an aluminum plate having a surface of 4 mm^2 onto which first a thin layer of aluminum

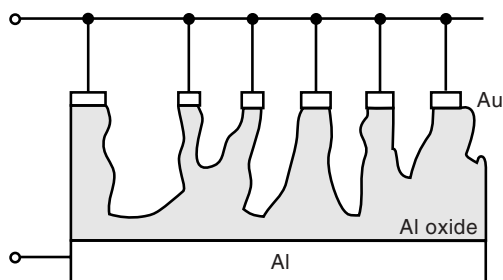


Figure 11. Cross section of an Al_2O_3 sensor.

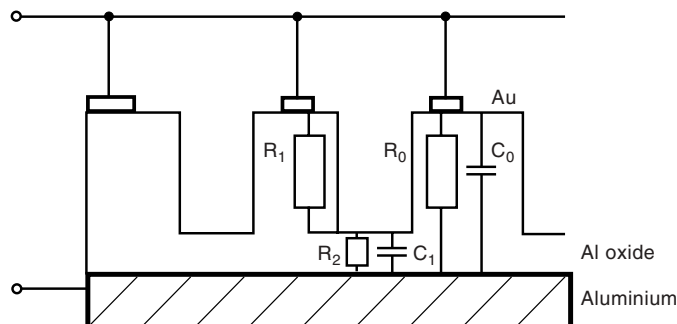


Figure 12. Schematic view of the electrical circuit equivalent to the Al_2O_3 sensor.

oxide and then a gold film are deposited. The absorption of water molecules which pass through the $0.3\text{ }\mu\text{m}$ thick gold layer to the pores of aluminum oxide depends on the partial pressure of water vapor. Consequently the resistance and the capacitance of the aluminum oxide change. Figure 12 shows the equivalent electrical circuit of an Al_2O_3 sensor shown in Fig. 11 (18).

Resistive and capacitive sensors are used for a variety of applications in ambient conditions, including control of air conditioning and other processes. Some are used in more extreme environments. “Dew-point” types are distinct from others in this group in that they cover much lower levels of humidity, in absolute units, and are used in the control and monitoring of relatively dry gases. They include aluminum oxides and other metal oxides, and silicon-based sensors. Relative humidity types can achieve uncertainties of 2% to 3% of value, at best, while dew-point types have uncertainties of around 2°C in dew point, at best, to 5°C at low dew points (15).

Advantages(+) and disadvantages(–) of these sensors include the following:

Capacitive sensors in particular

- + Usually tolerate condensation (though calibration may shift)

Resistive sensors in particular

- May be damaged if condensation occurs (though some sensors are protected)

Dew-point impedance sensors in particular

- + Usually tolerate condensation (though calibration may shift)
- + Can have wide measuring range
- May be slow to settle at very low (absolute) humidity
- Have some tendency to drift

PIEZOELECTRIC HYGROMETERS

Humidity-gas sensors have been developed using AT-cut quartz-crystal resonators vibrating in the thickness-shear mode. The sensing element is a radio-frequency quartz crystal coated with a hygroscopic polymer material. In a simple circuit using such a crystal, the frequency of oscillation decreases when the crystal gains weight due to water absorption by the

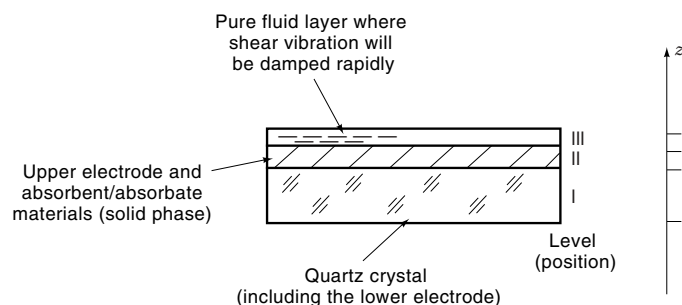


Figure 13. Three-layer model.

coating. Since frequency changes can be measured accurately and rapidly, this results in an accurate and simple moisture analyzer. In a SAW device a surface acoustic wave is electrically excited in a piezoelectric single-crystal plate substrate (e.g., quartz, lithium niobate, lithium tantalate) by use of a metallic (e.g., aluminum) interdigital transducer (IDT) structure. It is also possible to use the quartz crystal as a stable oscillation element whose substitutional electrical structure only is being changed through the variation of the serial capacitance. The sensor probe C_x is an open capacitor (having the form of a comb).

Coated Quartz-Crystal Humidity Sensor

While a coated quartz-crystal resonator is vibrating in ambient air (assuming only one surface is in contact with the ambient air), it can be simplified as a three-layer model. Figure 13 shows such a model, where layer I consists of the quartz crystal (including the lower electrode), layer II consist of the upper electrode and absorbent/absorbate materials that can be regarded as solid phase, and layer III is a pure fluid layer where shear vibration will be damped rapidly (19).

The principle is mainly based upon Sauerbrey's formula (20,21):

$$\Delta f = \frac{-2\Delta m n f_0^2}{A(gm_q \phi_q)^{1/2}} \quad (6)$$

where Δm = mass of substances absorbed on the surface, Δf = shift of frequency, ϕ_q = density, μ_q = shear modulus, A = surface, $m = 1$ (when the effect of generalized viscosity of layer III can be neglected).

When the resonant frequency f_0 is 3.5 MHz, the frequency change between 20% and 95% humidity is approximately 1000 Hz. Typical data obtained by an instrument using this principle of measurement are as follows: range: 0.1 to 33,000 ppm; speed: 99% dry-down in one minute; minimum detectable: ± 0.1 ppm; selectivity: hydrogen, hydrocarbons, olefins, alcohols, acid gases, and chlorinated solvent vapors cause only negligible interference.

SAW Resonator for Absolute Humidity Measurement

The frequency range covered starts at about 50 MHz and ends up at a few gigahertz, corresponding to a state-of-the-art interdigital finger linewidth below $0.5 \mu\text{m}$ (22). Because many parameters may contribute to a frequency change of the SAW device, the reverse determination of a parameter change from a measured frequency change is ambiguous. The most common solution of the problem is the use of two equal devices, where only one of the devices is coated with a layer that produces a known unique sensitivity to the measured gas (e.g., analytic concentration), while the other device is used as a reference element as in Fig. 14. The difference of the frequency changes of both devices is only sensitive to the measured gas, with the frequency changes caused by all other influences being compensated by this difference principle.

The SAW sensor is a 200 MHz two-port resonator fabricated on ST-cut quartz with aluminum metallization. The interdigital transducers were designed with split fingers apodized for narrow bandwidth. They are coated with a radio-frequency 20 nm thick sputtered with silicon dioxide layer as the sensing film. The coated SAW resonator and the reference resonator are fabricated on separate substrates. Responses of the coated SAW sensor to water vapor provide shifts in SAW resonator frequencies using modulation circuits. The frequency shift which is caused by the change of humidity from 1 to 6 g/m³ is approximately 500 Hz (23).

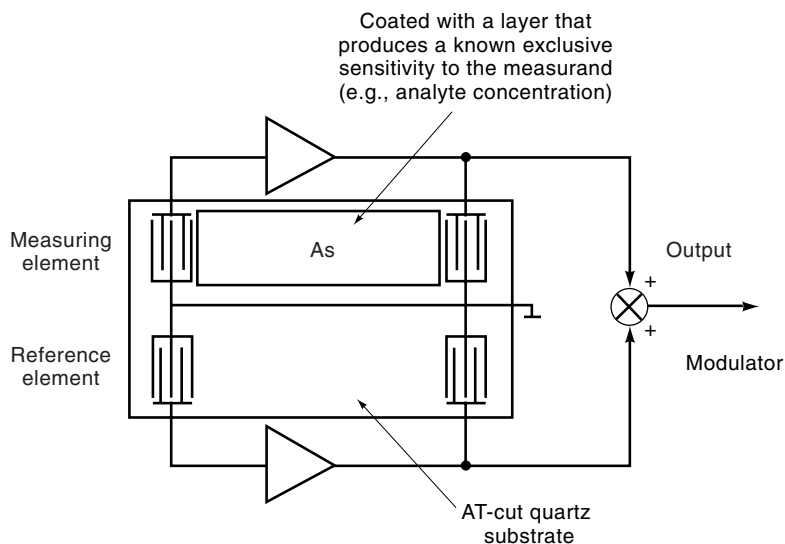


Figure 14. SAW sensor using the difference principle.

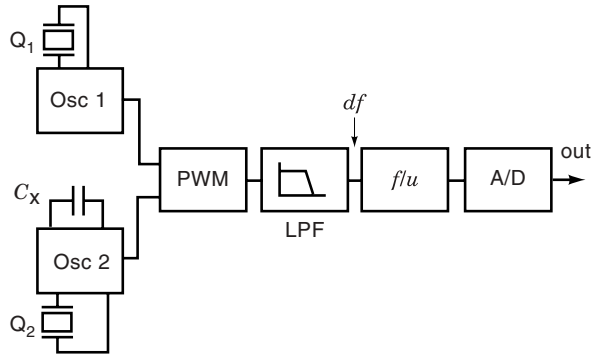


Figure 15. Hygrometer using the sensitive capacitive dependent crystals.

The absolute humidity of water-vapor density ρ_w in saturated moist air is given by the following equation (24):

$$\rho_w = \frac{18.0152 f(P, T) e_w(T)}{V_m P} \times 10^6 \text{ gm}^{-3} \quad (7)$$

where V_m is the molar volume of the moist air, P is the total pressure, $e_w(T)$ is the saturation water-vapor pressure at the absolute temperature T , and $f(P, T)$ is the enhancement factor. The molar volume, V_m , is obtained by solving the equation of state for moist air as described in Ref. 23. In the range of water-vapor density from 1 to 30 g m^{-3} , the estimated maximum relative uncertainty is 0.1%.

The Piezoelectric Hygrometer Using Sensitive Capacitive-Dependent Crystals

The frequencies of oscillator 1 and oscillator 2 of the sensor depicted in Fig. 15 are 18 MHz. The output of the pulse-width modulator (PWM) is a pulse-width signal which is compensated for temperature and voltage drift (Fig. 15) (24).

The sensor probe C_x is an open capacitor (having the form of a comb) (Fig. 16). The distance between the open capacitor plates (C_x) is 0.3 mm. The sensor probe is formed on an Al_2O_3 basis together with electronics and a temperature-sensing element. The crystal is used as a stable oscillation element whose electrical structure only is being changed through the variation of the serial capacitance C_x .

The humidity measurement is determined by Eq. (8). (K = constant depending on mechanical realization).

$$\Delta f = K^{\text{humidity}} \quad (8)$$

The sensor probe [open capacitor in the form of a comb which does not charge itself statically (Fig. 16)] has the surface $A1(C_{x1} = 9 \text{ pF})$, $A2(C_{x2} = 6 \text{ pF})$, $A3(C_{x3} = 3 \text{ pF})$ (Fig. 17).

The frequency measurement uncertainty is $\pm 0.1 \text{ Hz}$. The fact that the capacitor does not charge itself statically gives this

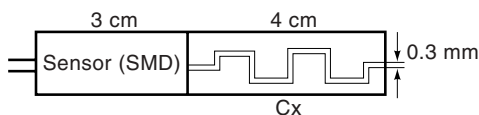


Figure 16. Sensor probe C_x .

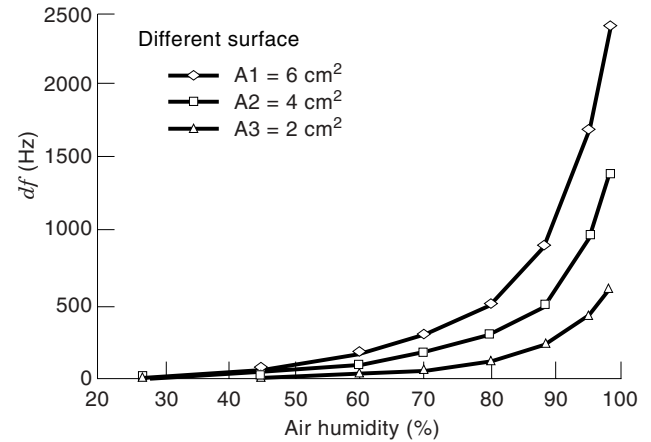


Figure 17. Air-humidity measurement.

method the advantage over the others. The typical uncertainty of high-air-humidity measurements is about $\pm 2.5\%$ ($T = 15$ to 25°C and humidity = 50 to 98%); however, this sensor presented has an improved uncertainty of $\pm 0.1\%$. The fact that the sensor is not made of absorption materials that can be affected by dust ensures long-term repeatability even after calibration. (Sensor probe is the only open capacitor.)

ELECTROLYTIC HYGROMETERS

There are two basic methods of electrolytic hygrometer: The first method is based on the measurement of conductivity and the second uses phosphorous pentoxide. The basis of the conductivity method represents the measurement of conductivity of electrolytic hygroscopic material (Fig. 18). The electric current flowing through the hygroscopic material causes the ionization of water molecules in the electrolyte which results in ionic conductivity. The second method uses phosphorous pentoxide, which absorbs humidity from the air.

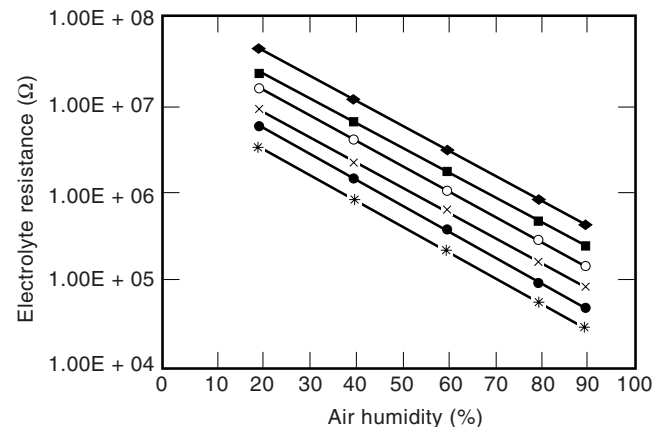


Figure 18. Temperature and resistance characteristics of an electrolytic hygrometer. The temperature rises from the lowest curve to the highest ones.

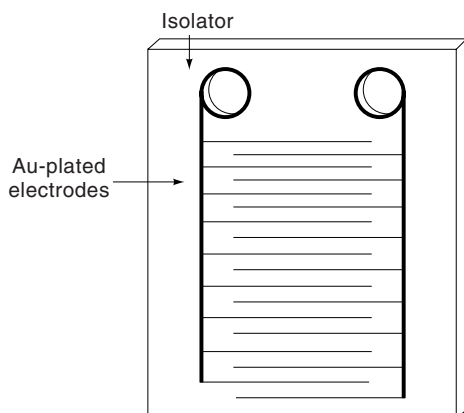


Figure 19. Electrolytic probe. Gold plated electrodes covered with either porous textile or cotton wool or glass wool or porous ceramic.

Method Based on the Measurement of Conductivity

The rate of humidity absorbed from the air affects the electric conductivity of hygroscopic material. The voltage applied is alternating due to polarization.

Figure 19 shows gold-plated electrodes covered with either porous textile or cotton wool or glass wool or porous ceramics. The uncertainty of the humidity measurement is $\pm 3\%$ (25,26) (Fig. 20).

Electrolytic Hygrometer Using Phosphorus Pentoxide

The sensor consists of a film of powerful desiccant, phosphorus pentoxide (P_2O_5), which strongly absorbs water vapor from the surrounding gas. A voltage is applied across the P_2O_5 and electrolysis takes place, dissociating the water into its constituents, namely, hydrogen and oxygen. The current that flows in this process is related (by Faraday's Law) (27) to the amount of water electrolyzed. Thus the magnitude of the current indicates the humidity of the gas being measured. These sensors are suitable for measuring very low humidities, though they require a steady (known) flow rate of gas. They are normally used in flow sampling configurations, rather than in probe form (27). In the electrolytic hygrometer the gaseous sample flows at a steady rate (conveniently 100 mL/min) through a polytetrafluorethylene (PTFE) tube about 76 cm long and with a 0.7 mm bore. Inside the tube are two interlaced but separate helical electrodes of fine platinum wire. The electrodes are covered with a thin film of partially hydrated phosphorus pentoxide, and a voltage of about 70 V dc is maintained between them. The sensitive element (which is flexible) is coiled and

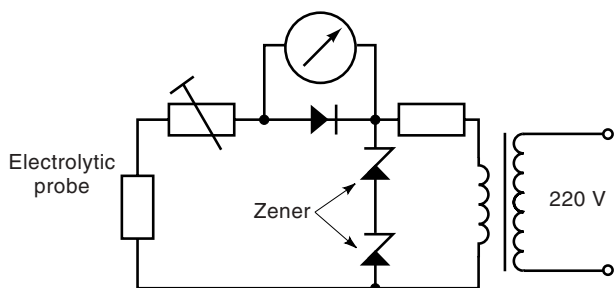


Figure 20. Humidity measurement based on conductivity method.

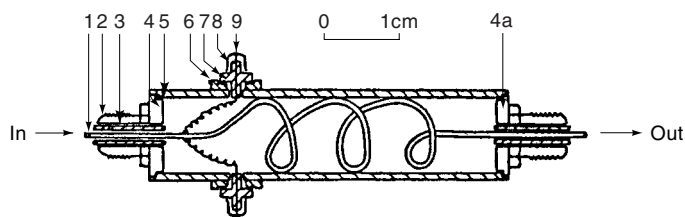


Figure 21. Cross section of detector. 1, Inlet end of sensitive element; 2, Fluo bush; 3, 6.35 mm male flare union brazed to part No. 4; 4 and 4a, brass end pieces; 5, brass tube filled with Seekay wax R93C; 6, boss brazed to brass tube; 7, Tufnol bush secured with Araldite; 8, metal cap secured with Araldite; 9, electrode soldered to metal cap.

potted in Seekay wax (chloronaphthalene wax) in a 1.905 cm bore brass tube 10.16 cm long (Fig. 21) and provided with connections for the gas supply with the terminals of the electrical circuits (28).

The principle of this method is that an equilibrium is set up (often within a few minutes) between the rate of absorption of water and the rate of electrolysis of the absorbed water. For a gas flowing at 100 mL/min at 20°C and 100 kPa (760 mm Hg), the equilibrium current is 13.4 μ A per part per million by volume of water vapor (28). Since the equilibrium current I is proportional to both the water content W and the sample flow rate R , the water content in ppm v/v at 20°C and 100 kPa is given by

$$W = \frac{100 I}{13.4 R} \quad (9)$$

where I is in microamperes and R is in milliliters per minute.

In some cases (e.g., Arcton 6) the film does not absorb all the water from the sample; the water content can then be obtained by the absorption efficiency of the detector for the particular gas being used. A method of determining the efficiency is described later. Equation (9), which is applicable to any gas, is used throughout this article. In the refrigeration industry, however, water contents are usually specified as a proportion by weight; the conversion factor for Arcton 6 is 1 ppm w/w = 6.7 ppm v/v .

This is a long established method for use in dry environments. The uncertainty is ± 3 to $\pm 10\%$ of reading at best (29). (+ advantages, – disadvantages)

- + Compatible with some corrosive gases
- + Can have good sensitivity
- Measuring cells have a limited life, after which they must be recoated (and recalibrated)
- Flow rate of gas is critical
- Damaged by exposure to normal ambient humidities and by periods of extremes of dryness (below 1 ppm)
- Slow response at lowest humidities
- Hydrocarbons, butadiene, ammonia, and some other contaminants prevent proper operation
- Adds traces of hydrogen and oxygen to the gas sample, which can recombine, leading to errors

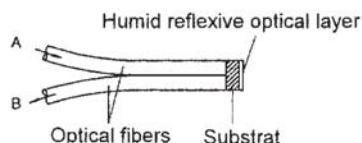


Figure 22. Optical fiber hygrometer.

OPTICAL FIBER HYGROMETERS

The optical fiber hygrometer is a sensor consisting of two optical fibers. One end of these two fibers is attached to the humid reflexive optical system as shown in Fig. 22. The other two ends are connected to the light source and attached to the optospectral measuring device and polychromator. Fibers are used only for the transmission of light (30).

The humidity sensing element is a multilayer system with good measuring reflective spectrum (Fig. 23). This is a Dünuschicht-Fabry-Perot interference filter, which consists of several consecutive high and low refractive dielectric $\lambda/4$ layers. The optical efficiency of the thickness of a single layer as a product of the geometric thickness and optical refraction of light (n_L for low and n_H for high refractive layers) needs to be set to the medium working light source wavelength λ . The polychromator has to be adjusted to that wavelength as well. The multilayer system consists of 8 dielectric $\lambda/4$ layers with low and 10 layers with high refractive index. The reflexive minimum is situated near 800 nm working wave range. All layers together are approximately $2 \mu\text{m}$ thick. SiO_2 with low refractive index and TiO_2 , ZrO_2 or HfO_2 with high refractive index are used as dielectric material. They are temperature stable and their chemical resistance is known. A more or less large structure of pores depends on the degree of evaporation in vacuum (30).

In humid environments water molecules can be reversibly sorbed. The optical refractive index of layers is changed and the reflexive spectrum is shifted towards longer wavelengths. The shift reversibly depends on the partial pressure of water vapor. The volume of different size pores and their probable distribution determine the functional course of the spectrum shift. To achieve the wanted porous structure, technological parameters of certain layers need to be preset. The final sensitivity depends on the optical resolution of the polychromator.

The Sensor's Measuring Range

The sensor's porous structure is designed to measure the dew-point temperature in the range from $+10$ up to -100°C . It is interesting to note that the temperature above 300°C does not destroy the sensor. In such case the sensor only needs a regeneration time and a new calibration. The dew also does not damage the sensor. Once the latter is dry again, it is like newly calibrated.

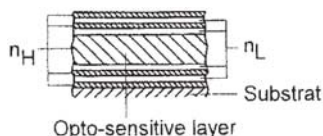


Figure 23. Multilayer structure.

The optical fiber sensor has the following advantages:

- it is very small ($\phi < 1 \text{ mm}$)
- it can be more than 100 m away from the measuring electronics
- the measuring results are not influenced by electromagnetic disturbances, capacitance and inductance between the conductors, parallel resistance
- it can work in an explosive environment
- it is temperature and corrosive stable
- it works up to 100°C
- it has a short reaction time (5 s)

CALIBRATION

Calibration is the process of comparing a measuring instrument with a fixed reference for the same type of measurement, to identify any bias or systematic error in the readings. The outcome of a calibration is normally a certificate listing any corrections that need to be applied to the values indicated by the instrument, together with an estimate of the uncertainty in the calibration, and other relevant information. For example, a calibration of a given instrument at, say, 50% rh, might show it to read too high by 1% rh. If so, a required correction of -1% rh would be shown on the certificate.

Humidity Generation

The main methods of humidity generation are as follows (12):

Dew-Point Generator. Gas is humidified by bubbling it through (or flowing it over) water or ice at a given temperature. Depending on the efficiency of the design, this will produce gas with a dew point more or less close to the temperature of the saturator.

Two-Temperature Generator. Gas is humidified at a given temperature, defining its dew point, or vapor pressure. The gas is then fed to a chamber at another (higher) temperature. By varying one or both temperatures, it is possible to obtain any chosen value of relative humidity or dew point.

Two-Pressure Generator. Gas is humidified at elevated pressure and then expanded to a lower pressure (at the same temperature). As the total pressure of the gas falls, so does the partial pressure of water vapor. Knowing the initial humidity allows the humidity after expansion to be found. By varying the initial pressure or temperature or both, it is possible to obtain any chosen value of relative humidity or flow dew point.

Mixed Generator. Gas is humidified by any convenient method and is then mixed with drier gas. The value of humidity achieved depends on the mixing proportions and on the water content of the two gas streams.

Salts. Saturated (or unsaturated) salt solutions, and certain other chemicals, can be used to generate an environment of a particular relative humidity in an enclosed space. The value of relative humidity obtained depends on the particular chemical salt, the concentration of the solution, and the temperature of use, among other things.

Table 3. Standard Salt Solution for Humidity Calibration

Temperature (°C)	LiCl	MgCl ₂	Mg(NO ₃) ₂	NaCl	KCl
0		33.66 ± 0.33	60.35 ± 0.55	75.51 ± 0.34	88.61 ± 0.53
5		33.60 ± 0.28	58.86 ± 0.43	75.65 ± 0.27	87.67 ± 0.45
10		33.47 ± 0.24	57.36 ± 0.33	75.67 ± 0.22	86.77 ± 0.39
15		33.30 ± 0.21	55.87 ± 0.27	75.61 ± 0.18	85.92 ± 0.33
20	11.31 ± 0.31	33.07 ± 0.18	54.38 ± 0.23	75.47 ± 0.14	85.11 ± 0.29
25	11.30 ± 0.27	32.78 ± 0.16	52.89 ± 0.22	75.29 ± 0.12	84.34 ± 0.26
30	11.28 ± 0.24	32.44 ± 0.14	51.40 ± 0.24	75.09 ± 0.11	83.62 ± 0.25
35	11.25 ± 0.22	32.05 ± 0.13	49.91 ± 0.29	74.87 ± 0.12	82.95 ± 0.25
40	11.21 ± 0.21	31.60 ± 0.13	48.42 ± 0.37		82.32 ± 0.25
45	11.16 ± 0.21	31.10 ± 0.13	46.93 ± 0.47		81.74 ± 0.28
50	11.10 ± 0.22	30.54 ± 0.14	45.44 ± 0.60		81.20 ± 0.31
55	11.03 ± 0.23	29.93 ± 0.16			80.70 ± 0.35
60	10.95 ± 0.26	29.26 ± 0.18			80.25 ± 0.41
65	10.86 ± 0.29	28.54 ± 0.21			79.85 ± 0.48
70	10.75 ± 0.33	27.77 ± 0.25			79.49 ± 0.57
75	10.64 ± 0.38	26.94 ± 0.29			79.17 ± 0.66
80	10.51 ± 0.44	26.05 ± 0.34			78.90 ± 0.77

Humidity Chambers. Test chambers are available with both temperature and humidity control based on steam or spray injection, or on one of the principles outlined above.

Bottled Gases. Cylinders of compressed gas are available which supply gas of fixed humidity, at a selection of values of a few hundred ppm or below. The method relies on a special coating of the cylinder interior.

Used appropriately, any of the above methods can provide a stable humidity for use in comparing one hygrometer with another, already calibrated, one. This may mean supplying a flow of gas to pass through both instruments to be compared, or placing both instruments in a chamber where they are exposed to the same humidity. Humidity generators can be calibrated, but only at national standard level are they accurate and authoritative enough to act as absolute references.

Calibration Using Salt Solutions

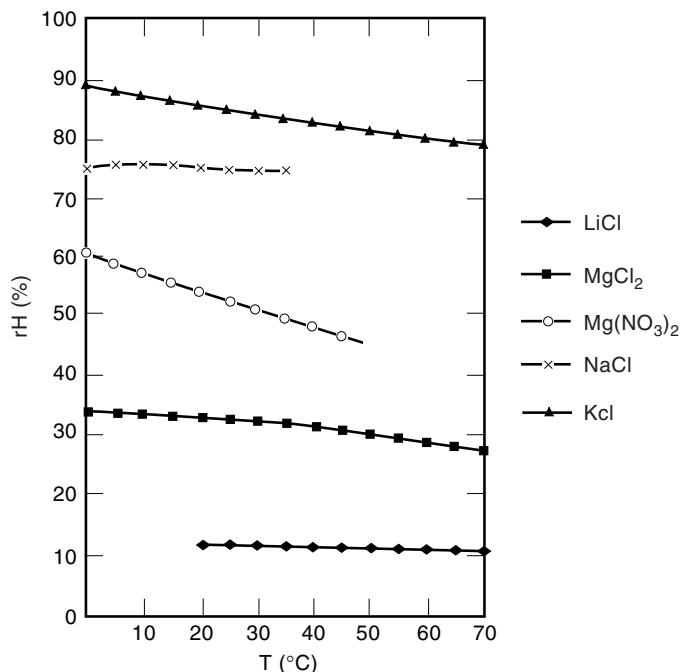
Salt solutions generate characteristic values of relative humidity which depend on the particular chemical salt, the concentration of the solution, and the temperature of use. Saturated salt solutions, with solid salt present, have the special property that a stable concentration—and hence a constant relative humidity—is maintained even if water migrates to or from the solution. Ready-made capsules are commercially available for a variety of relative humidity “fixed points.” Such capsules are designed to seal over the end of relative humidity probes, providing a small enclosed atmosphere of defined relative humidity. Certain of these, if in the form of a solution with a solid salt present, are reusable and can be calibrated to assign a traceable value to the reference humidity they provide.

Table 3 shows values of humidity generated over saturated solutions of various salts (13,31). Just as it is for relative humidity in general, temperature stability is critical for calibration with salt solutions, particularly since the properties of the salt solutions change with temperature (Table 3, Fig. 24). Stabilization after a change in temperature may take an hour, or many hours, depending on the size of the change. For freshly mixed

solutions, full stabilization may take a long time, sometimes weeks (31,32).

Some Practical Recommendations for Calibration of Hygrometers

Hygrometers should be calibrated in terms of the quantity to be measured. For example, the performance of a wet- and dry-bulb hygrometer cannot be characterized by calibrating the thermometers alone; a humidity calibration of the hygrometer as a whole is needed. Where a hygrometer consists of separate parts (e.g., probe and electronics) the pair should be calibrated together as one item and used together. Wherever possible, calibrations should be performed under the

**Figure 24.** Temperature stability of salt solutions.

intended conditions of use—that is, at similar values of humidity and temperature, and preferably under similar conditions of pressure, air flow, and so on. Hygrometer readings in some units of measurement (such as grams per cubic meter) are specific to gas type. Therefore a calibration in these units is valid only for a selected gas. This is so because the ratio of molecular masses of water and air, for instance, is about 0.62, whereas the ratio for methane is about 1.1. However, calibrations in terms of (say) parts per million by volume are not necessarily gas-specific, or only very slightly so (provided the instrument genuinely senses this quantity and does not simply give a numerically converted result from some other parameter) (33).

A general guideline for calibrating the various types of hygrometers might be as follows:

- Hygrometer types with relatively good stability (e.g., condensation dew-point hygrometers, wet- and dry-bulb hygrometers): 6 to 12 months after first calibration; thereafter every 1 to 12 years, depending on requirements.
- Moderately stable hygrometer types (e.g., electrical relative humidity instruments): perhaps 6 months after first calibration; thereafter every 6 to 12 months, depending on requirements.
- Less stable hygrometer types (e.g., aluminum oxide hygrometer) every 6 months, or sooner if desired.

Where a choice of intervals is suggested above, the longer intervals are for hygrometers confined to use in ordinary room or laboratory conditions, whereas the shorter intervals are for those used in more varied conditions. If used in extreme conditions or if accuracy is critical, then measurement checks of some kind at much shorter intervals are recommended. Where hygrometers are strongly prone to drift, this will remain a significant source of uncertainty almost irrespective of the frequency of calibrations.

LONG-TERM STABILITY

The measurement characteristics of an instrument will change with respect to time, due to gradual changes in its electrical or material components. Estimates of long-term stability, or drift, refer to the likely change in the instrument's measurement performance with respect to time. Regular checks of calibration should be made to quantify this problem. Although sometimes adjustments can be made to the hygrometer reading in the light of drift, performance could be compromised by the adjustment process, and the overall drift characteristics could be masked by frequent adjustments (34,35).

Speaking about the measurement characteristics of hygrometers, we also have to take into consideration the following terms:

Uncertainty. The uncertainty of a measurement is defined as the parameter characterizing the range in which the “true value” can be expected to lie. It defines a “margin of doubt” about a reading or estimated value, together with a level of confidence (normally 95%) that the “true value” will lie within this range. Often, the uncertainty for an instrument is specified for ideal operating conditions at a temperature of 20°C or 23°C.

Table 4. Effect of a Temperature Change of 1 °C at Various Levels of Temperature and Relative Humidity

Relative Humidity	Temperature				
	10 °C	20 °C	30 °C	50 °C	70 °C
10% rh	±0.7% rh	±0.6% rh	±0.6% rh	±0.5% rh	±0.5% rh
50% rh	±3.5% rh	±3.2% rh	±3.0% rh	±2.6% rh	±2.3% rh
90% rh	±6.3% rh	±5.7% rh	±5.4% rh	±4.6% rh	±4.1% rh

The method of use may make an important contribution to the overall uncertainty achieved.

Accuracy. Strictly speaking “accuracy” is a qualitative term only. For example, an instrument or measurement might be described generally as “accurate” or “not accurate.” However, the word “accuracy” continues to be used loosely in specifications to refer to the maximum difference that can be expected between the reading given by an instrument and the “true value” being measured.

Repeatability. In general terms, the repeatability of an instrument is the closeness of agreement of multiple readings repeated under the same conditions of measurement. To quantify repeatability, the spread or dispersion of results may be quoted, for example, in terms of a standard deviation of a number of readings.

Reproducibility. In general terms, reproducibility is an instrument's capacity to reproduce a previous measurement. This may be at a later date or after undergoing significant changes in conditions, for example, after a change of operator or location. To quantify reproducibility, the spread or dispersion of results may be quoted, for example, in terms of a standard deviation of a number of readings.

Temperature Influence on the Stability of Relative Humidity Measurement

Relative humidity is highly dependent on temperature. Roughly speaking, at room temperature, a change in dew point of 1°C corresponds to a change in relative humidity of 6% of the relative humidity value. At 50% rh, for example, an uncertainty in dew point of ±1°C corresponds to an uncertainty of ±3% rh. A change of 1°C in the measurement of the ambient temperature has almost exactly the same significance. The size of the effect under different conditions is illustrated in Table 4.

Overall, a useful rule of thumb is that ±1°C uncertainty in either dew point or temperature leads to an uncertainty of ±6% of the relative humidity value. The change in the relative humidity levels is not symmetric (36–38).

BIBLIOGRAPHY

1. D. M. Considine and G. Considine, *Process Instruments and Controls Handbook*, 3rd ed., New York: McGraw-Hill, 1985.
2. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 21–54.
3. E. R. Schramek, Psychrometer. In *Taschenbuch für Heizung und Klimatechnik*, Munich: Oldenbourg, 1997, pp. 305–307.
4. W. Lück, Psychrometer, In *Feuchtigkeit*, Munich: Oldenbourg, 1964, pp. 104–119.

5. D. R. Lide, Relative humidity from wet and dry bulb temperatures. In *CRC Handbook of Chemistry and Physics*, 77th ed., Boca Raton, FL: CRC Press, 1996–1997, pp. 15–22.
6. W. Lück, Die Wasseraufnahme der Stoffe. In *Feuchtigkeit*, Munich: Oldenbourg, 1964, pp. 48–56.
7. H. Kiessig and W. Koblit, *Der Anteil an Adsorptionswasser bei der Quellung von Cellulosefasern*, Papier 13, H. 15/16, 1959, p. 356.
8. F. Kollmann, Eine Gleichung der Sorptionsisotherme, *Naturwissenschaften*, **49**: 206–207, 1962.
9. T. Seiyama, Development and application of ceramic humidity sensors. In *Chemical Sensor Technology*, Vol. 1, Amsterdam: Elsevier, 1988, pp. 57–61.
10. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 41–53.
11. W. Lück, Sättigungsverfahren. In *Feuchtigkeit*, Munich: Oldenbourg, 1994, pp. 74–98.
12. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 30–31.
13. D. R. Lide, Standard salt solutions for humidity calibration. In *CRC Handbook of Chemistry and Physics*, 77th ed., Boca Raton, FL: CRC Press, 1996–1997, pp. 15–24.
14. D. R. Lide, Values of relative humidity at a selection of temperature and dew points. In *CRC Handbook of Chemistry and Physics*, 77th ed., Boca Raton, FL: CRC Press, 1996–1997, pp. 15–23.
15. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 14–25.
16. Valvo Sensoren, in *Feuchtesensor*, Herausgeber: Valvo Unternehmens-Bereich Bauelemente der Philips GMBH, Burcharstraze 19, 2000 Hamburg 1, 1980, pp. 48–52.
17. T. Seiyama, Humidity sensor using TiO_2 - SnO_2 Ceramics. In *Chemical Sensor Technology*, Vol. 2, Amsterdam: Elsevier, 1989, pp. 133–149.
18. H. Fischer and K. Heber, Impedanzveränderung eines Dünnschicht-Aluminium-Oxide. In *Industrielle Feuchtigkeitmesstechnik*, Ehningen, Germany: Expert Verlag, 1990, pp. 42–44.
19. Y. Dong and G. Feng, Effect of surface physical sorption on characteristic of coated quartz-crystal humidity sensor, *Sensors and Actuators*, **B24-25**: 62–64, 1994.
20. A. Wexler and W. H. King, The piezoelectric sorption hygrometer. In *Humidity and Moisture*, Vol. 1, New York: Elsevier, 1965, pp. 578–583.
21. E. Benes et al., Sensors based on piezoelectric resonators, *Sensors and Actuators*, **A48**: 1–22, 1995.
22. D. Hauden, S. Rousseau, and J. Gagnepain, Sensitivities of SAW-oscillators to temperature, forces and pressure: Application to sensor. In *Proc. 34th Annu. Frequency Control Symposium*, Philadelphia, 1980, pp. 735–743.
23. P. H. Huang, Sensor calibration of a SAW resonator for absolute humidity measurement in microelectronic packages, *Sensors and Actuators*, **B24-25**: 686–688, 1995.
24. V. Matko and D. Donlagic, Sensor for high-air-humidity measurement, *IEEE Trans. Instrum. Meas.*, **45**: 561–564, 1996.
25. W. Lück, Elektrolytisches Hygrometer. In *Feuchtigkeit*, Munich: Oldenbourg, 1964, pp. 129–135.
26. T. Seiyama, Design of polymer electrolytes-based humidity sensors. In *Chemical Sensor Technology*, Vol. 1, Amsterdam: Elsevier, 1988, pp. 117–132.
27. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 15–23.
28. J. K. Crawshaw, Electrolytic hygrometers for measurements in gases down to a few parts per million, *J. Sci. Instrum.* **36**: 121–126, 1959.
29. H. Fischer and K. Heber, Verfahren der Feuchtemessung. In *Industrielle Feuchtigkeitmesstechnik*, Ehningen, Germany: Expert Verlag, 1990, pp. 19–59.
30. A. Böhm and G. Schwotzer, Faseroptische Feuchtemessung, *Sensortechnik*, Heft 10, 13. May, 1997, pp. 48–51.
31. DIN 50 008, Kontantklimate über wässrigen Lösungen (Gesättigte Salzlösungen-Glycerin Lösungen), Deutsche Normen, 1981.
32. W. Lück, Die relative Luftfeuchte über wässrigen Lösungen. In *Feuchtigkeit*, Munich: Oldenbourg, 1964, pp. 56–59.
33. J. Pichlmaier, *Befeuchtung von Prüfgas für die Kalibrierung von Gas- und Feuchtesensoren*, Technisches Messen 58/12, Munich: Oldenbourg, 1991, pp. 471–477.
34. The Institute of Measurement and Control, *A Guide to the Measurement of Humidity*, London: Crown, 1996, pp. 45–52.
35. W. Lück, Psychrometer. In *Feuchtigkeit*, Munich: Oldenbourg, 1964, pp. 104–119.
36. Rotronic AG, *Feuchtemessung in Luft, Feuchte und Temperaturmessung*, Bassersdorf, Germany, 1996–1997, p. 146.
37. H. Fischer and K. Heber, Messgenauigkeiten für Gasfeuchte. In *Industrielle Feuchtigkeitmesstechnik*, Ehningen, Germany: Expert Verlag, 1990, pp. 26–59.
38. P. Profos and T. Pfeifer, Die Luftfeuchtemessverfahren. In *Handbuch der industriellen Messtechnik*, Munich: Oldenbourg, 1992, pp. 821–824.

VOJKO MATKO
University of Maribor

PYROMETERS

Infrared (IR) temperature measurement has been in use for a century. It has gained wide acceptance in industry and science wherever the target of interest cannot be contacted because it is too hot, moving or, for other reasons, not accessible to a contact probe such a resistance or thermocouple element. The worldwide annual market for IR pyrometers is now around 1/4 billion US dollars. There are about two dozen firms offering pyrometers, the most important of which are listed in Table 1.

The following sections deal with concepts and implementation of all practical pyrometers that are in use today or are about to enter the market.

FUNDAMENTALS OF PYROMETRY

At temperatures above absolute zero all bodies emit electromagnetic radiation. Planck's radiation law (1) describes the relationship between a body's surface temperature and the associated spectral radiance

$$L(w, T) = E(w, T)c_1w^{-5}[\exp(c_2/wT) - 1]^{-1} \quad (1)$$

where $L(w, T)$ is the spectral radiance as a function of wavelength w and absolute temperature T . $C_1 = 3.7413 \times 10^{-12} \text{ W} \cdot \text{cm}^2$, $C_2 = 1.4388 \text{ cm} \cdot \text{degree}$ are the first and second radiance constants, and $E(w, T)$ is the spectral emissivity of the surface, which also is a function of temperature. For a black body $E(w, T) = 1$. Figure 1 shows graphs of spectral radiance vs. wavelengths for different temperatures T , the so-called Planck curves. For all other bodies $E(w, T) < 1$. From Eq. (1) then

Table 1. List of Pyrometer Companies

Cino Co., Ltd., 32-8, Komano-cho, Itabashi-ku, Tokyo 173 Phone 011 3 3956 211; fax 011 3 3956 0915
Impac Electronic GmbH, Krifterler Strasse 32, D-60326 Frankfurt/M, Germany Phone 011-49-69-9 73 73-0; fax 011-49 69-9 73 73-181
IRCON, Square D Company, 7301 N. Caldwell Ave., Niles, IL 60648 Phone 708-967-5151; fax 708-647-0948
Keller GmbH, In der Garte 40, D-49479 Ibbenbüren, Germany Phone 011-49-5451-85-0; fax 011-49-5451-412
Land Instruments International, Ltd., Dronsfield, S18 1DJ, UK Phone +44 (1266) 417 691; fax +44 (1266) 290 274
Luxtron, 2775 Northwestern Pkwy, Santa Clara, CA 95051-0951 Phone 1-800-627-1666 or 1-408-727-1600; fax 1-408-727-1677
Mauer (Dr. Georg) GmbH, Industriegebiet 10, D-72664 Kohlberg, Germany Phone 011-49-7025-3031; fax 011-49-7025-6053
Mikron Infrared Co. Ltd., 16 Thornton Road, Oakland, NJ 07436 Phone 1-201-405-0900; fax 1-201-405-0090
Minolta Camera Co., Ltd., 30, 2-Chome, Azuchi-Machi, Higashi-Ku, Osaka 541, Japan
Pyrometer Instruments Co., Inc., 234 Industrial Pkwy., P.O. Box 70, Northvale, NJ 07647 Phone 201-768-2000; fax 201-768-2570
Quantum Logic Corp., Box 191, Westport CT 06881 Phone 203-226-3541; fax 203-227-8240
Raytek, Inc., 1201 Shaffer Rd., Box 1820, Santa Cruz, CA 95061-1820 Phone 408-458-1110; fax 408-458-1239
Tokio Seiko, 1-1 Kozura-Cho, Chigasaki-City, Kanagawa-Pref. 253, Japan Phone 011/ 81 467 546522; fax 011/ 81 467 546177
Williamson, 70 Domino Dr., Concord, MA 01742 Phone 1-617-369-9607

follows that $E(w, T)$ is the ratio of the radiance of a surface at a certain temperature T to the black body radiance at the same temperature T . A special case occurs when $E(w, T) = \text{constant} < 1$. Such objects are called gray bodies. The Planck curves exhibit several important features (see Fig. 1):

1. They do not intersect each other, which means that the radiance at any wavelength increases monotonically with temperature.
2. For each temperature there is a wavelength of peak emission that shifts to longer wavelengths with decreasing temperature. This shift is expressed by Wien's displacement law (2)

$$w_{\max} \cdot T = 2897.6 \mu\text{m} \cdot \text{degree}$$

3. If one sets $E(w, T) = 1$ and integrates Eq. (1) over all wavelength, one gets the Stefan-Boltzmann law for the total radiance of a black body:

$$L_{\text{tot}} = s \cdot T^4$$

where $s = 5.669 \times 10^{-12} \text{ W} \cdot \text{cm}^2 \cdot \text{degree}^{-4}$ is the Stefan-Boltzmann constant. If the object is *not* a black body, the

total radiance is given by

$$L_{\text{tot}} = E_{\text{eff}}^* s \cdot T^4$$

where E_{eff} is an effective emissivity.

Total Radiance Systems

The emissivity functions $E(w, T)$ depend on the chemical composition of the surface and its texture. Emissivity values have been measured and tabulated for many different materials, but even a thin coating by a different material can change the emissivity drastically. One important industrial example is hot steel being sheared in a processing plant, exposing a pure iron surface. The emissivity of pure steel is about 0.40 in the near infrared spectrum. Hot steel oxidizes rapidly, thereby increasing the emissivity, which can double the emissivity in minutes. Another example is from the semiconductor industry, namely that of a silicon wafer during thermal processing. Such a wafer would typically have on its surface films of materials with different optical refraction indices than that of the silicon substrate. The emissivity of such a wafer is substantially changed from that of the uncoated wafer due to optical interference effects, which occur on such coated surfaces, and which depend strongly on the wavelength of observation. Great temperature errors can result in both above cases if the change in emissivity is not taken into account.

The oldest automatic instruments measured the radiance over the whole infrared spectrum. This yields the most radiative power at the detector, which in early days was a thermocouple element of very low responsivity requiring a maximum of radiant power.

Narrow-Band Systems

With the development of more sensitive detection methods, particularly quantum detectors coupled to transistor amplifiers, it became feasible to measure small spectral portions of the radiant spectrum. There are several reasons why one wants to use small portions of the radiant spectrum:

1. The shorter the wavelength, the steeper the local exponential slope of radiance versus temperature. Operating at the shortest possible narrow-wavelength band minimizes temperature error due to emissivity uncertainty and calibration drift.
2. Some materials are opaque at certain wavelength bands. Spectral absorption curves for several materials are shown in Fig. 2.
3. Multiwavelength pyrometry.

Wavelength Choices

As illustrated by the Planck curves, the radiance peaks shift to longer wavelengths as the target temperature decreases. It is therefore necessary to match the operating wavelength band of the pyrometer to the temperature range of interest. Here one must consider the available radiation detectors (3). The major detectors and their associated temperature ranges are listed in Table 2. For free radiating targets, one should work at the shortest wavelength at which one still obtains reasonable signals. At this point the local exponential slope and the suppression of temperature errors are largest.

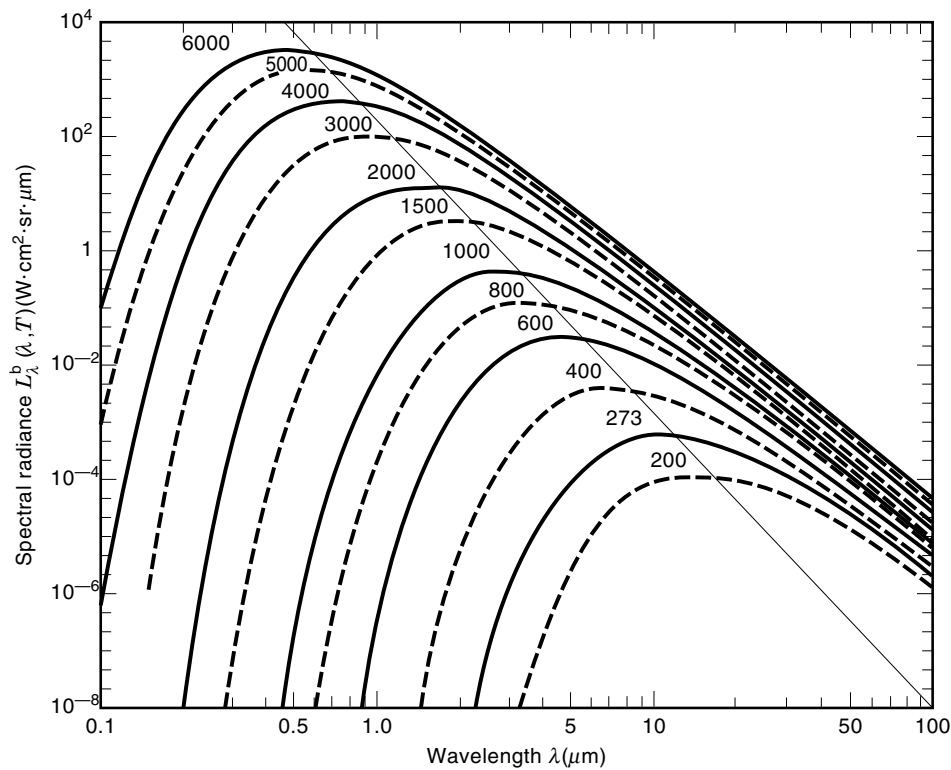


Figure 1. Planck curves (curves are for constant blackbody temperature, in K).

Disappearing-Filament System

Prior to the widespread use of infrared detectors, optical pyrometers were the instruments of choice. The operating principle is a visual comparison of the brightness of an electrically heated wire to that of the target seen through a red filter within the pyrometer’s viewing optics. By varying the electrical current in the wire, one can adjust its brightness until it blends into the target surface. The wire current is then a measure of the target’s radiance temperature. The drawback with these instruments is that they require operator judgment, which may differ from person to person. Also there is usually no provision for emissivity correction.

A newer version of this instrument projects a bright line of variable intensity into the field of view. The line illumination is provided by a small commercial lamp. Its intensity is monitored by a silicon photodetector. Another method is to hold the filament at a constant temperature, and use a narrow filter coupled with a neutral density filter to match the brightness of the filament with the target.

Dual-Wavelength Method

Dual-wavelength pyrometers exploit the nonlinearity of the Plank equation by determining the temperature from the radiance ratio at two different wavelengths (4). Here, uncertain emissivity ratio leads to error. The real advantage of dual-wavelength pyrometers over single-wavelength systems is the insensitivity of the dual-wavelength measurement to partial obstructions in the line of sight such as dust, steam, and vignetting, since both radiance values attenuate by the same factor. For the same reason there is no lower limit to the

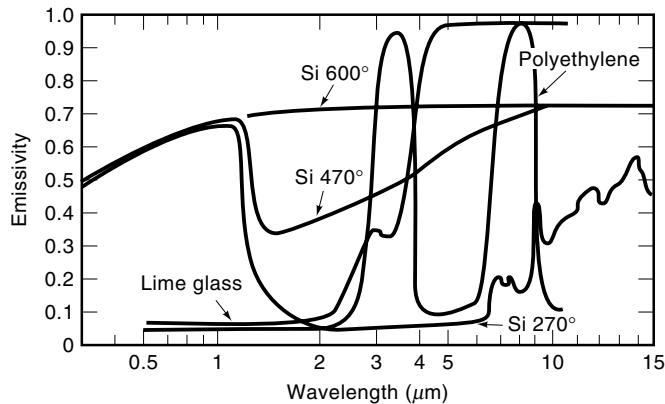


Figure 2. Spectral emissivity of materials: silicon at 270, 470, and 600°C; lime glass at 500°C; polyethylene at 50°C.

Table 2. Lower Temperature Limits

Lower Limit of Temperature	Infrared Detector ^a	Wavelength (μm)
600°C	Si	0.4–1.0
200°C	Ge and InGaAs	0.8–1.8
100°C	PbS	1.0–3.6
50°C	PbSe	1.5–5.8
–50°C	Pyroelectric detector	No wavelength dependency

^a Operating at room temperature.

target size, as long as the signals at the two wavelengths are measurable.

Active Method

In the following we distinguish between passive and active pyrometry. The former are the traditional instruments which only *receive* radiation, hence the term “passive.” New types of pyrometer have been introduced in the last decade, which augment the traditional optical receiver with a reflectometer, sending radiation to the target as well receiving radiation from it; thus the term “active.” The fundamental shortcoming of passive pyrometry, namely, its dependence on an estimate of emissivity, was recognized from its inception. For pure materials, such as most metals, the emissivity is known from special sample measurements. But even a thin oxide surface layer can change the emissivity drastically, leading to large temperature errors. In 1905, Rubin (5) proposed a method of measuring the target emissivity E by recognizing that for opaque bodies their hemispheric reflectivity R_{hemisph} is always the unity complement to its emissivity, $E = 1 - R_{\text{hemisph}}$. When a distant target reflects diffusely, R_{hemisph} (and therefore E) can be determined from a measurement of the reflected energy into a relatively small solid angle (6). In practice, one calibrates the reflectance measurement against a diffuse surface of known reflectivity, such as CaCO_3 at a known distance d_c and then scales all reflectance values of proper targets at distance d by the factor $(d/d_c)^2$. This is so, since for a particular receiver system the solid angle over which reflected radiation is collected scales with the square of the target distance. The reflectance measurement must, of course, be made at the same wavelength as the radiance measurement. This method is also referred to as “reflection-supported pyrometry.”

Furnace Applications

Objects are placed into furnaces to heat them up. By the nature of this process, such objects (targets) are surrounded by heat sources of higher radiance temperature than their own. Reflected ambient radiance can be as large or even larger than the target's thermal self-emission. An interesting situation occurs when the ambient irradiance is isotropic and its radiance temperature is equal to the target temperature. In that case the target appears as if it were a black body (Appendix 1). In general, however, the furnace is hotter than the target and one must separate the thermal target radiance from the reflected furnace radiation because only the former is associated with the target temperature. There is a family of active pyrometers that is designed for this application. These pyrometers have provisions to measure the average ambient radiance as well as the target reflectance. A microcomputer then subtracts the target radiance from the total radiance and then calculates the temperature from the former.

PRACTICAL DESIGNS

The following is a description of the pyrometer types that are commercially available at present. Total radiation and optical pyrometers are either being phased out or used for special applications only. Narrow-wavelength and dual-wavelength systems dominate the market. Active pyrometers, a newer type of

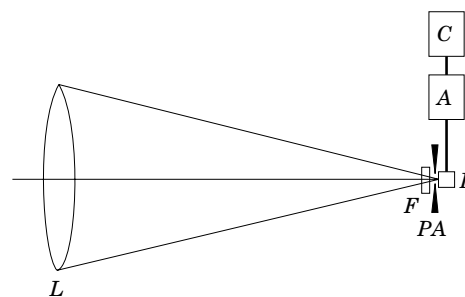


Figure 3. Passive single-wavelength system. D , detector; F , filter; L , objective lens; PA , pinhole aperture; A , amplifier; C , computer.

instrument, are gaining acceptance where accurate temperature measurements are required.

Narrow-Wavelength Pyrometers

The simplest pyrometer design, illustrated in Fig. 3, employs a lens (L), a pinhole aperture (PA), an optical filter (F), and a detector (D), with an output signal that is amplified by circuit (A) and then applied to microcomputer (C), which calculates the temperature from L and a manual input for the emissivity E via the Planck equation (1).

The temperature is then displayed in digital form (DS). One counterintuitive, but, nevertheless, valid feature of that design is that the radiance signal yielded by the detector is independent of the target distance. In other words, if the target moves out of focus, the detector signal remains the same. This very important feature is proven in Appendix 2. More advanced instruments are of the focusing type in which the lens (L) is adjustable to image the target surface onto the detector plane. An additional refinement is to reimagine the pinhole onto the detector with an auxiliary optical system (7). This latter design reduces stray radiation on the detector and thereby yields the highest degree of angular selectivity.

The dependence of temperature error on emissivity uncertainty can be analyzed by differentiating Eq. (1) with respect to E :

$$dT/T = (wT/c_2)dE/E \quad (2)$$

Equation (2) shows that for any given uncertainty in radiance or emissivity, the corresponding temperature error is smaller by the factor wT/c_2 ; that is, the shorter the wavelength, the smaller the temperature error, with all thing being equal. Of course, at shorter wavelengths the signal falls off exponentially. This puts a lower limit on the choice of wavelength for any given temperature. There are about a dozen companies that offer a complete assortment of narrowband pyrometers to cover the range of industrially important applications.

Dual-Wavelength Pyrometers

Figure 4 illustrates the design of one version of a dual-color system. A dichroic beam splitter M is placed in the optical path to separate the two wavelength components. One uses two filter-detector-amplifier systems $F1-D1-A1$ and $F2-D2-A2$ to measure the respective radiances. Both signals are applied to the microcomputer (C), which calculates the target temperature from the

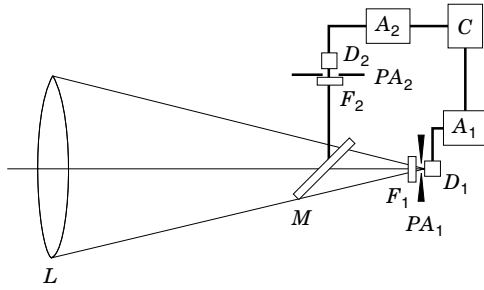


Figure 4. Passive dual-wavelength system. D_1 , detector for wavelength 1; D_2 , detector for wavelength 2; F_1 , filter for wavelength 1; F_2 , filter for wavelength 2; M , dichroic mirror; PA_1 , pinhole aperture for wavelength 1; PA_2 , amplifier for wavelength 2.

formula

$$T = [(c_2/T)(1/w_2 - 1/w_1)] / [\ln(L_1/L_2)(E_2/E_1)(w_2/w_1)^5] \quad (3)$$

with a manual input for emissivity ratio E_1/E_2 . The dual-wavelength systems on the market employ wavelength combinations in the near- and mid-infrared regions: 0.7/0.9, 0.9/1.55, and 3.2/3.7 μm (8).

Active Pyrometers

By adding a reflectometer function, an ordinary pyrometer is transformed into an active instrument illustrated in Fig. 5, where the beam of a pulsed light source (LA), such as a laser, is superimposed onto the pyrometer's optical axis by a beam combiner (M_2). To normalize the outgoing radiation pulse, a beam splitter (M_1) diverts a small portion of the outgoing radiation onto a second, reference detector (D_1), whose output, after amplification, V_2 , is applied to a microcomputer (C). The radiance received by detector D consists of a dc component V_0 associated with the thermal target radiance and a superimposed pulsed component of height V_1 , which is associated with the reflected pulsed radiation. Timing signals (TR), synchronously generated with the outgoing pulsed radiation, are used by the amplifiers (A) and (A_r) to separate the pulses from the dc component. Finally, a signal V_3 associated with the target distance d is also applied to the computer. When the lens (L) is carefully focused

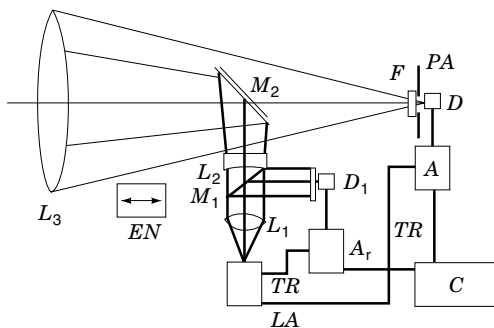


Figure 5. Active single-wavelength system. D , detector; D_r , reference detector for laser output; F , filter; A , signal amplifier; A_r , reference signal amplifier; L_1 , laser collimating lens; M_1 , beam splitter for the laser beam; L_2 , diverging lens; L_3 , objective lens; M_2 , mirror; PA , pinhole aperture; C , computer; EN , distance encoder.

onto the target, the position of this lens relative to its infinity focus location is a measure of the target distance from the lens. Thus, by indexing this relative lens position, monitored by the distance encoder EN , one generates signal V_3 to determine the value of d . This is then used in the following formula to calculate the in situ target emissivity:

$$E = 1 - c_{23}(V_1/V_2)/d^2 \quad (4)$$

The radiance component is proportional to V_0 ,

$$L = c_{20}V_0 \quad (5)$$

where c_{23} and c_{20} are instrument constants. After determining d from V_3 , the computer calculates the temperature T from V_0 , V_1 , and V_2 via Eqs. (4), (5), and (1).

FIBER-OPTIC SYSTEMS

All pyrometers of the preceding type can in principle be designed as fiber-optic models. The main advantage of fiber systems is their optical flexibility. While large-lens-type pyrometers require a straight, large-aperture line of sight, fiber systems make pyrometer measurements possible for very restricted and indirect optical access. Ordinary fiber cables have no imaging property; they simply collect over the acceptance angle of the fiber (about 20° for a typical clad quartz system). Also, normal fiber cables use polyvinyl chloride (PVC) jackets and must, therefore, stand off from a hot target. Hence the target spot would grow quickly with distance. One therefore places either a sapphire light pipe or a lens system in front of the fiber cable. A sapphire light pipe is extremely heat resistant and can be placed very close to a hot surface in order to make the target spot small. Another method is to stand off from a hot surface and use a lens to image the desired target spot onto the fiber end face. The other end of the fiber cable is placed close to a photodetection system, which sends its output to a signal-processing system. A schematic drawing of a fiber-optic pyrometer is shown in Fig. 6. Fiber-optic cables have wavelength restrictions. If we take 1 dB/m attenuation as a somewhat arbitrary spectral cutoff criterion, water-free quartz fibers transmit from about 0.5 μm to about 2.5 μm (9). Sapphire fibers transmit from 0.5 μm to 4 μm (10), and chalcogenide (As-Se-Te) fibers transmit from 3.0 μm to 9.5 μm (11). Together, these three fiber types cover the most important

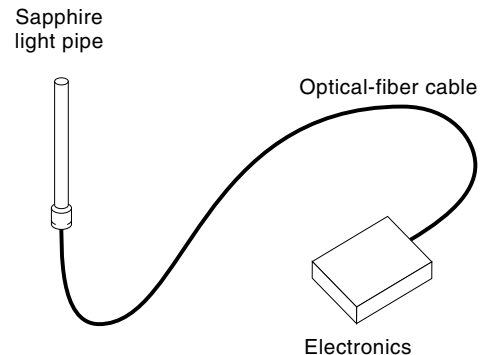


Figure 6. Typical fiber-optic pyrometer.

spectral ranges of pyrometry. In principle, therefore, all these types of pyrometers can be designed as fiber-optic models.

NEW DEVELOPMENTS

Active Dual-Wavelength Pyrometers

Active pyrometry operating at one wavelength has been making very accurate temperature measurements when the target surfaces are diffuse. But there are many more applications for which the surface scattering is somewhere between specular and diffuse, and for which single-wavelength active pyrometry will therefore not work. For these cases the dual-wavelength active technique was developed (12,13).

If one assumes that the polar scattering of a target surface is wavelength independent, one can prove that a measurement of the radiances at two wavelengths and the *ratio* of retroreflectances yields the target emissivity E_1 . Since only the ratio of reflectances is needed, the method works for any polar scattering pattern. The formula for E_1 developed in Ref. 13 has the form

$$E_1 = 1 - R + RB(E_1)^n \quad (6)$$

where

$$B = [L_2(w_2)^5/c_1]/[L_1(w_1)^5/c_1]^n \quad (7)$$

$$n = w_1/w_2 \quad (8)$$

Equation (6) can be solved by successive iteration. From E_1 and L_1 follows the temperature T via the Planck equation (1).

Figure 7 shows a system demonstrated in Ref. 13. A bifurcated optical fiber is used to superimpose the illuminating and optical receiver systems. The lamp radiation is mechanically chopped and the reflected energy is synchronously amplified by a lock-in circuit so as to distinguish it from the dc radiance photo signal. The optical receiver uses a di-chroic prism beam splitter to separate the two wavelength components. Additional spectral selection is provided by filters 1 and 2, located in front of detectors D_1 , D_2 , respectively. Reflection and thermal

emission signals for both wavelengths are digitized by an ADC and applied to a PC for calculation of emissivity and temperature via equations (6), (7), and (1).

Ripple Pyrometry

Ripple pyrometry, another interesting technique is used to measure and control the temperature of silicon wafers heated by quartz-halogen lamps in various rapid thermal processors (14). The ripple method dynamically determines emittances by sensing time-dependent components in radiation from the lamps and radiation reflected by the wafer surfaces; see Fig. 8. A sapphire light probe (S_1) is used to detect radiation from a region on the wafer surface. A second sapphire light probe (S_2) is directed towards the heating lamps and measures the radiation incident on the wafer surface, including secondary reflected radiation. The radiation from sapphire probes S_1 and S_2 is coupled via optical fibers to separate photodetectors. The quartz-halogen lamps are ac powered and therefore exhibit a 120 Hz ripple on top of a dc radiance component, which gives rise to ac photosignals dI_1 and dI_2 , as well as dc photosignals I_1 and I_2 . According to Ref. 14 the wafer radiance L_w and its reflectivity r_w are related to the ac and dc signals as follows:

$$L_w/(1 - r_w) = (s_1^{-1}I_1 - s_2^{-1}RI_2)/(1 - R) \quad (9)$$

where the effective emittance $1 - R$ and reflectance R are defined in terms of the ratio of the ac components in the two detector signals, according to

$$R = (s_1/s_2)(dI_1/dI_2) \quad (10)$$

Here s_1 and s_2 are instrument constants. The left-hand side term of Eq. (9) is the blackbody temperature associated with the true wafer temperature and thus yields that temperature. This method has been used to measure wafer temperatures during rapid thermal annealing, an important technique in manufacturing microelectronic devices. Wafer annealing at a temperature of 660°C was monitored with an accuracy of better than 3°C at 1 standard deviation (15).

There are two other types of pyrometer designs which aim to eliminate the emissivity error, multiwavelength pyrometer (16) and the polaradiometer (17), which works best for specular, metallic targets. Multiwavelength systems match the radiance

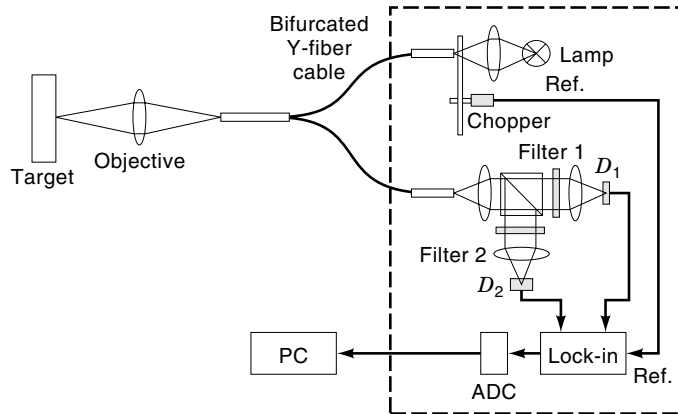


Figure 7. Two-color active pyrometer. ADC, analog-to-digital converter; D_1 , D_2 , photodetectors.

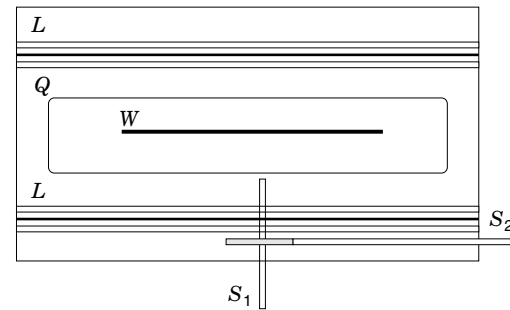


Figure 8. Schematic cross-sectional view of an RTP furnace showing wafer (W), lamps (L), quartz isolation tube (Q), and optical-fiber probes for the wafer (S_1) and the lamps (S_2). Arrows indicate the directions of the radiances in the model of Eq. (1).

at three or more wavelengths to the Planck formula with some assumption about the spectral emissivity characteristic of the target surface.

The polaradiometer illuminates a specular target spot from 45° off normal with radiation from a black body while observing, from the mirror image direction, the sum of reflected and thermally emitted radiance with a polarization-analyzing optical receiver. The radiation emitted into an angle far off normal by a hot metal is such that the dominant component of polarization is E_1 (in the plane of incidence) with a lesser component E_2 (perpendicular to the plane of incidence). Since Kirchhoff's law holds for each component, the originally unpolarized black body radiation, when reflected, has R_2 as the dominant component of reflection, the lesser component R_1 . One varies the temperature of the black body illuminator, until the combined intensity of emitted and reflected radiation is the same for the two principal polarizations. At that point the temperature of the target is equal to that of the black body.

Both the multiwavelength and the polarization-compensating pyrometers are limited to very special applications, and are not really commercial at this point.

APPENDIX A. BLACKBODY CONDITIONS IN FURNACES

When making temperature measurements of objects inside furnaces, the influence of reflected radiation must be taken into account in order to determine the correct target temperature. One way to do this is to define an effective emissivity E_{eff} . Consider a hollow enclosure of arbitrary shape with a small opening (Fig. 9).

Assume now that the inside surface is at uniform, elevated temperature. We call L_{bb} the black body radiance of the surface, r the surface reflectivity, and the physical emissivity is E is unity, where once again $E = 1 - r$. Radiation from any point A on the inside surface is the sum of the thermal emission at that point, EL_{bb} , and reflected energy $r(1 - r)L_{\text{bb}}$ emanation from all other points on the internal surface. That radiation itself is of course also the sum of thermal self-emission and reflected radiation from other points on the surface, and so forth at infinitum.

The observed radiance therefore is equal to the infinite sum:

$$L_{\text{obs}} = L_{\text{bb}}(1 - r) + L_{\text{bb}}(1 - r)r + L_{\text{bb}}(1 - r)r^2 + L_{\text{bb}}(1 - r)r^3 + \dots \quad (\text{A.1})$$

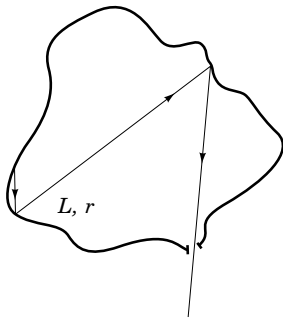


Figure 9. Hollow blackbody. L , surface radiance; r , surface reflectance. The enclosure at uniform temperature emits like a blackbody independent of the physical emissivity of the inside wall material.

This equation can be written in the form

$$L_{\text{obs}} = L_{\text{bb}}(1 - r)(1 + r + r^2 + r^3 + \dots) \quad (2)$$

Noting that the sum of the infinite series inside the curly bracket equals $1/(1 - r)$, one obtains

$$L_{\text{obs}} = L_{\text{bb}}(1 - r)[1/(1 - r)] = L_{\text{bb}} \quad (3)$$

which means the total radiance emanating from the small opening is that of a black body at the temperature of the internal wall, independent of the reflectivity (emissivity) of the wall. Note that this analysis ignores the effect of radiation escaping through the opening. This is permitted if the area of the opening is small compared to the total internal wall area of the hollow space.

APPENDIX B. OUT-OF-FOCUS TARGETS

Most on-line pyrometers have a fixed focus. In general one tries to select the optics such that the target is in focus. But in many practical cases, the target ends up somewhat out-of-focus because the installation procedure is not precise. Sometimes the target distance changes such as for objects on a conveyer belt. When the instrument is calibrated by focusing at the opening of a blackbody furnace and is used in the field against a target that is out of focus, the question arises: How does this affect the calibration? Not at all! In the following we will verify by analysis this experimental fact.

If the optics is focused at distance a_1 , but the target is located at distance a_3 , the power impinging on the detector can be calculated with the help of a ray optics construction illustrated in Fig. 10. A circular detector aperture of diameter d_1 , located at image distance b_1 from the lens, corresponds to a circular target area at object distance a_1 of diameter M_1 on the other side of the lens. The object and image distances are related as

$$1/a_1 + 1/b_1 = 1/f \quad (\text{B.1})$$

where f is the focal length of the lens. The diameter M_1 is given by

$$M_1 = d_1(a_1/b_1) \quad (2)$$

The optical power passing through the detector aperture is

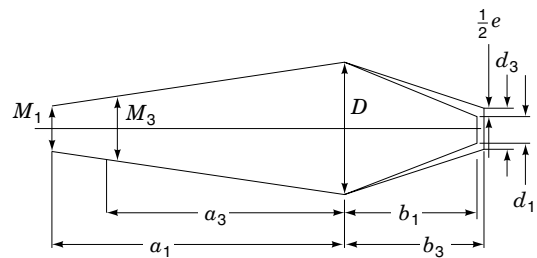


Figure 10. Spot diameters versus object distance. The edge ray construction is used to calculate the changes in target spot diameter and etendu.

proportional to the etendue:

$$(M_1)^2(D/a_1)^2 \quad (3)$$

where the term in the second set of parentheses represents the solid angle subtended by the lens as seen from the object plane. Imagine now that the target is moved to a new location at distance a_3 from the lens. By geometric analysis of the rays running from the edges of the lens to the edges of the target spot at a_1 , one can show that the spot diameter at distance a_3 is

$$M_3 = (a_3/a_1)(M_1 - D) + D \quad (4)$$

where D is the diameter of the lens.

Consider then a radiating disk of diameter M_3 at distance a_3 from the lens. The associated image distance b_3 on the other side of the lens is determined from

$$1/a_3 + 1/b_3 = 1/f \quad (5)$$

The image diameter

$$d_3 = d_1 + e \quad (6)$$

is larger than the detector aperture d_1 . Thus only a certain fraction of the imaged radiation passes through the aperture. A first-order approximation is to set the fraction of transmitted energy equal to

$$(d_1/d_3)^2 = [d_1/(d_1 + e)]^2 \quad (7)$$

where

$$e = (D/b_3)(b_3 - b_1) = Db_1(1/b_1 - 1/b_3) \quad (8)$$

Hence, the power passing through the aperture is proportional to

$$(M_3)^2(D/a_3)^2[d_1/(d_1 + e)]^2 \quad (9)$$

which is the etendue for an out-of-focus target.

In the following we will prove that the etendues (and hence the photosignals) for in-focus and out-of-focus targets are identical. For this purpose we may consider just the square roots of the respective etendues. Then, we normalize the aperture diameter $d_1 = 1$ in Eq. (B.9) without loss of generality. Next, we substitute e from Eq. (B.8), and M_3 from Eq. (B.4) into Eq. (B.9) and arrive at the expression:

$$D[1/a_3][(a_3/a_1)(M_1 - D) + D]/[1 + Db_1(1/b_1 - 1/b_3)] \quad (10)$$

The two rectangular brackets in the numerator of Eq. (B.10) can be consolidated and written as

$$M_1/a_1 + D(1/a_3 - 1/a_1) \quad (11)$$

From Eqs. (B.1) and (B.5) follow

$$1/a_3 - 1/a_1 = 1/b_1 - 1/b_3 \quad (12)$$

We substituted Eq. (B.12) in (B.11) and (B.10) to arrive at

$$D[M_1/a_1 + D(1/b_1 - 1/b_3)]/[1 + Db_1(1/b_1 - 1/b_3)] \quad (13)$$

In Eq. (B.13) we substitute $b_1 = a_1/M_1$ once, namely in the second square bracket before the parentheses. This results in the following expression:

$$D[M_1/a_1 + D(1/b_1 - 1/b_3)]/(a_1/M_1)[M_1/a_1 + D(1/b_1 - 1/b_3)] \quad (14)$$

Note that the terms in square brackets in the numerator and in the denominator of Eq. (B.14) cancel each other, leaving the expression $M_1(D/a_1)$. Hence, the etendue (and therefore the photosignal) for out-of-focus targets is equal to the etendue for in-focus targets. This approximation applies to cases for which the defocusing is moderate, that is where $b_3 - b_1 \ll f$.

For targets placed beyond the focal plane, the target diameter is given by

$$M_2 = (a_2/a_1)(M_1 + D) - D \quad (15)$$

A similar geometrical analysis as given in the preceding shows again that the signal for out-of-focus targets is the same as for in-focus targets. This very important feature of pyrometers is key to their use in factory application.

BIBLIOGRAPHY

1. M. Bass (ed.), *Handbook of Optics*, New York: McGraw-Hill, 1995, chap. 1, p. 13.
2. M. Bass (ed.), *Handbook of Optics*, New York: McGraw-Hill, 1995, chap. 1, p. 15.
3. Hamamatsu Corporation, Tech. Inf. SD-12, July 1993, p. 11.
4. J. Sunderland, US Patent No. 3,795,918, 1974.
5. H. Rubin, *Ann. Phys.*, **18**: 725 (1905).
6. A. Stein, *6th Int. Congr. Appl. Lasers Electro-optics*, 1987, p. 29.
7. *The Infrared Handbook*, Washington, DC: Office Naval Res., 1978, pp. 20–27.
8. Product literature from companies: Impact Electronic GmbH, Quantum Logic Corp., and Williamson given in Table 1.
9. M. Bass (ed.), *Handbook of Optics*, New York: McGraw-Hill, 1995, chap. 13, p. 15.
10. Advanced Crystal Products, Woughburn, MA.
11. *Optical Fiber Brochure*, Oxford, UK: Oxford Electronics Ltd.
12. A. Stein, US Patent No. 4,708,493 and European Patent No. EP0317653 B1.
13. G. Koska and J. Reger, *VDI Berichte 1379, Temperature '98*, p. 227.
14. A. T. Fiory et al., *Mater. Res. Soc. Symp. Proc.*, **303**: 139, 1993.
15. B. Nguyenphu, M. Oh, and A. T. Fiory, Temperature monitoring by ripple pyrometry in rapid thermal processing, *Mater. Res. Soc. Symp. Proc.*, **429**: 1996, pp. 291–296.
16. P. B. Coats, *Metrologia*, **17**: 103–109, 1981.
17. T. P. Murray, *Rev. Sci. Instr.* **38**: 791, 1967.

Reading List

J. T. Quinn, *Temperatures*, New York: Academic Press, 1983.

ALEX STEIN
Quantum Logic Corporation

ACOUSTIC VARIABLES MEASUREMENT

The science and technology of acoustics encompass sound, ultrasound, and infrasound in all media. There are many specialties in acoustics as well as interrelations to other branches of engineering, physics, and chemistry. Consequently, the list of quantities and their corresponding units is rather extensive (1), including both fundamental (e.g., mass) and derived (e.g., force) physical quantities as well as measures of level (e.g., sound pressure level). This article concentrates on the measurement of variables that are peculiar to acoustics.

Acoustical variables have evolved from two general disciplines: wave theory and circuit theory. Because of this, many acoustical variables have their counterparts or analogs in electromagnetic theory, electrical-network theory, mechanics, and heat—a convenience that must be used with care. Naturally, as in other branches of engineering and physical science, variables applicable to acoustics only have been introduced. As an exact science, acoustics is relatively young; hence, its quantitative development has been primarily in terms of the metric system, which accounts for the almost exclusive use of the SI units.

MEASURABLE ATTRIBUTES OF SOUND

Sound pressure is the most important quantity in acoustics, probably because it is the attribute of sound measurable by all hearing mechanisms. The term sound pressure (or acoustic pressure) refers to the increment over ambient pressure, or excess pressure, produced by a disturbance of the medium, not to the total ambient pressure (2). Similarly, acoustic displacement refers to the net motion of the molecules of the medium, in response to the disturbing force and not to the random motion of particles when the medium is “at rest.” Another measurable quantity is the density, which has the usual definition of mass per unit volume. When a sound wave is propagated in the medium, the ambient value of density changes in proportion to the changes in pressure. Particle velocity and acceleration have the same meaning as in any mechanical system, their magnitude depending on the strength of the disturbance. Particle velocity is not to be confused with the speed of propagation of the acoustic wave, which depends entirely on the properties of the medium and environmental parameters, such as temperature. Since by their nature all of the above quantities are alternating, frequency (or period) is one of their most basic parameters. In modern acoustical practice, frequency is measured by a frequency counter or signal analyzer after the acoustic signal has been converted into electrical. The unit of frequency is the hertz, Hz. The quantities already listed are of a fundamental nature because other quantities can be derived from them. The derived quantities include intensity, impedance, and a variety of coefficients and ratios, some of which are too specialized and, hence, outside the scope of this work, and others are discussed in separate articles.

Sound Pressure and Sound Pressure Level

The American National Standard (3) defines instantaneous sound pressure as the difference between total pressure and the atmospheric static pressure that exists at a given point in space, at a particular instant of time, in a stated frequency

band. The SI unit is the pascal, Pa, which results from a 1 newton force uniformly applied over an area of 1 m². In acoustical practice, the term sound pressure or effective sound pressure is the root mean square of the instantaneous sound pressures determined over a specified time interval, at a point in space. Because of the wide range of pressures attending acoustic phenomena (from micro- to kilo-pascals), it is frequently more convenient to compare a sound pressure to some reference value, rather than state its magnitude in pascals. This ratio, converted to the logarithmic scale, is called sound pressure level in dB. It is defined as follows:

$$SPL = 10 \log \left(\frac{p}{p_{\text{ref}}} \right)^2 = 20 \log \frac{p}{p_{\text{ref}}} \quad (1)$$

The reference pressure, p_{ref} , must always be specified. The standard value for gases is 20 μPa ; for other media, $p_{\text{ref}} = 1 \mu\text{Pa}$, unless otherwise specified.

Particle Displacement, Velocity, and Acceleration

In the acoustics of fluids, the number of applications requiring a direct measurement of variables describing particle motion is very small. For instance, in the case of a loudspeaker it is the motion of the vibrating diaphragm rather than the displacement of the air particles that one measures. When particle displacement is required, it can usually be calculated from pressure measurements and the properties of the acoustic medium, especially since no simple instruments are available to perform this task. The opposite is true for solid media, where vibratory displacement, velocity, and acceleration are measured routinely by a variety of sensors and instrumentation. In terms of their relative usage in vibration measurements, accelerometers have considerable advantages over displacement and velocity sensors (4).

There are two measures of velocity of an acoustic medium: the particle velocity and the volume velocity. The particle velocity, defined at a point in the medium, is the velocity of a portion of the medium small enough that the variations of acoustic parameters within the volume of the portion can be considered infinitesimal. It is measured in m/s. The volume velocity is the volume of the medium displaced per unit time through a specified area. Thus, it is equal to the integral of the particle velocity over the area, and its unit is the cubic meter per second (5). Particle acceleration has the usual meaning of the rate of change of velocity, and is measured in m·s⁻². In the United States, acceleration is also expressed in units of g_n , the standard acceleration of free fall, equal to 9.80665 m·s⁻², and acceleration levels are often stated with reference to 1 μg . As in the case of acoustic pressure, when working with these variables, one must distinguish between instantaneous and effective root-mean-square (rms) values, which have the same meaning as for the electrical variables (6).

Measures of the Strength of an Acoustic Signal

The sound strength depends on the amount of energy carried by the wave. Two measures of this property are commonly used: the acoustic energy density flux and acoustic intensity. The energy transported by acoustic waves through a fluid medium comprises two components: the kinetic energy associated with particle motion, and the potential energy stored during compression and rarefaction of successive fluid elements. The

energy density is measured in joules per cubic meter. Because the instantaneous particle velocity and acoustic pressure are functions of both position and time, the instantaneous energy density, ϵ , is not constant throughout the fluid. Its time average, $\bar{\epsilon}$, at any point in the fluid can be expressed by a simple formula only if the relationship between particle velocity, u , and pressure, p , is known. For example, in a plane simple harmonic wave,

$$p = \rho_0 c u \quad (2)$$

hence

$$\bar{\epsilon} = \frac{1}{T} \int_0^T \epsilon dt = \frac{1}{2} \rho_0 U^2 = \frac{1}{2} \frac{P^2}{\rho_0 c^2} \quad (3)$$

where ρ_0 is the medium density, c is the speed of wave propagation, and P and U are the amplitudes of the acoustic pressure and particle velocity respectively (7). In sound fields where such relationships do not apply, pressure and velocity have to be measured by sampling over the area or volume of interest and appropriately summed. Investigation of jet noise is a good example (8).

Acoustic intensity, I , of a sound wave is defined as the mean rate of flow of energy through a unit area normal to the direction of propagation (5). Consequently, it represents the power of a sound wave incident on a unit area of a detector. However, unlike their optical counterparts, the simple acoustical detectors do not directly respond to intensity, but rather to pressure, displacement, velocity, or acceleration. But, since energy, and hence intensity, varies as the square of these quantities, one needs only to find the factor of proportionality. For example, basic acoustical theory shows that for a plane free progressive wave the intensity, I , and acoustic pressure, p , are related by the following equation:

$$I = \frac{1}{\rho_0 c} p^2 \quad (4)$$

The unit of intensity is the watt per square meter. For the calculation of intensity levels, $10^{-12} \text{ W} \cdot \text{m}^{-2}$ is often taken as the reference intensity. Measurements of acoustic intensity and acoustic energy-density flux are the subject of a separate article.

Frequency Response Functions

Acoustic variables, individually or in combination, depend on various parameters, frequency being the most common. Since most variables by their nature are phasors, they are representable by complex quantities having a real and an imaginary part or, alternately, magnitude (also called amplitude) and phase. Magnitude information is sought more often than phase; however, the dependence of both magnitude and phase on frequency are both very important characteristics of an acoustic or electro-acoustic system. Typically the magnitude is expressed in normalized form based on some convenient reference, which must be stated. Moreover, the ratio can be expressed in decibels instead of a fraction when the quantities can be related to energy or power, as was done for sound pressure level in Eq. (1). For phase, a common reference value is the phase at a specified frequency. In Fig. 1 the theoretical amplitude and phase response of a Bruel & Kjaer microphone is compared to

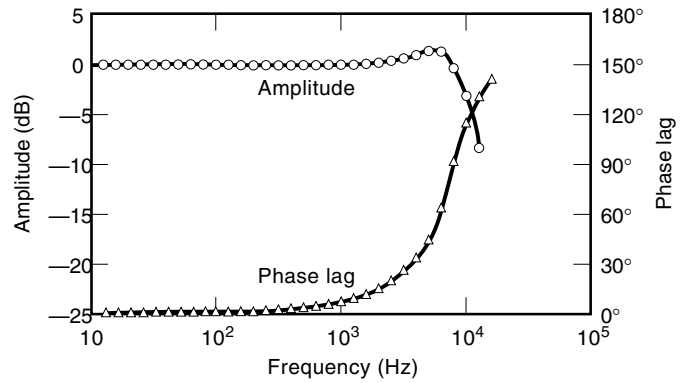


Figure 1. Example of a frequency-response function: relative amplitude and phase response of B&K type 4146 microphone. Solid lines: theoretical; symbols: experimental. See Ref. 9.

the measured response (9). It should also be noted that not all frequency-response functions represent ratios of the same denomination: they can, for example, be pressure-to-velocity ratios considered in the next section.

DESCRIPTIVE PARAMETERS AND FUNCTIONS

The behavior of acoustic systems is described not only by the basic variables discussed in the preceding section, but also by a collection of quantities derived from them. For a complete treatment of these parameters and functions the reader is referred to the rather extensive specialized literature, Ref. 10 being a good starting point, while in this article we consider only those deemed most important and commonly used.

Impedance

The concept of impedance was introduced into acoustics by A.G. Webster in 1919 by analogy with electrical circuits (11). Webster saw the value of extending the significance of the concept of impedance to the case of mechanical oscillations and acoustic wave propagation. The concept has been found to be of great value in both theoretical acoustics and applications. Impedance is a very useful quantity connecting the excitation and response of a system (see Fig. 2). Electrical impedance is defined as the ratio of the phasor representing an alternating voltage to the phasor representing a corresponding current (6). As such, it is a complex quantity having magnitude and phase angle (or, alternatively, a real and an imaginary part), is time invariant, but generally frequency dependent. Impedance, therefore, is one of the frequency-response functions listed in Table 1 which are used to describe the behavior of a linear system.

Before proceeding, it is proper to warn the reader of the dichotomy in choosing the sign of the reactance. It arises from the

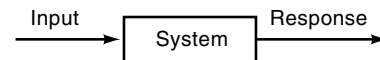


Figure 2. Block diagram of a single input/single output system for defining frequency-response functions of Table 1. The complex function of frequency, given by the output-to-input ratio, characterizes the system.

Table 1. Frequency-Response Functions in Common Use

Output Q Input	Pressure	Force	Displacement	Velocity	Acceleration
Pressure	SPL				
Force		Force ratio	Compliance	Mobility	
Displacement		Stiffness	Displ. ratio		
Velocity		Impedance		Vel. ratio	
Acceleration		Mass			Accel. ratio

different representations of a traveling wave used in physics and engineering. A simple illustration readily makes this clear. Consider, for example, a plane wave of acoustic pressure, p , traveling in the direction of the x -axis. The physicist usually represents it by

$$p = P_+ e^{-i(\omega t - kx)} + P_- e^{-i(\omega t + kx)} \quad (5)$$

where P_+ and P_- are the amplitudes of the forward and backward traveling waves respectively, k is the wave number, and ω is the angular frequency of the alternating pressure. The electrical engineer prefers the following notation:

$$p = P_+ e^{j(\omega t - kx)} + P_- e^{j(\omega t + kx)} \quad (6)$$

The reason for these choices is that the physicist likes to think of the forward wave as traveling in the positive- x direction, while the engineer focuses on the phasor representation $P_+ e^{j\omega t}$. Elementary analysis (12) readily shows that the former choice ($-i$) leads to a negative sign for mass reactance and a positive sign for compliant reactance. In order to make the results of electrical circuit theory directly applicable to acoustics, the latter ($+j$) choice is followed in this article.

Types of Impedance Used in Acoustics. In accord with national and international standards, impedance of acoustic systems is specified in three different ways; hence, the type of impedance selected by the user must be carefully stated. The first of these, Z_a , is called acoustic impedance and is defined as the quotient of the sound pressure, p , on a given surface, S , divided by the volume velocity, U , through the surface:

$$Z_a = \frac{p}{Su} = \frac{p}{U} \quad (7)$$

where u is the particle velocity and S is the area. Its unit is the acoustic ohm, which has the dimensions of $\text{Pa} \cdot \text{s} \cdot \text{m}^{-3}$.

The second type of impedance is the specific acoustic impedance, Z_s , defined as the quotient of the sound pressure and the particle velocity:

$$Z_s = \frac{p}{u} \quad (8)$$

The unit is the rayl, equal to $\text{Pa} \cdot \text{s} \cdot \text{m}^{-1}$.

The third type of impedance, called the mechanical impedance, Z_m , is defined as the quotient of the force exerted on a given area and the resulting particle velocity:

$$Z_m = \frac{pS}{u} \quad (9)$$

The unit of Z_m is the mechanical ohm, equal to $\text{N} \cdot \text{s} / \text{m}$.

Each type of impedance has certain advantages in the solution of specific problems. For instance: the acoustic impedance does not vary with changes in cross section of a conduit down which a sound wave is traveling; specific acoustic impedance in a plane progressive wave is nearly the same at all points; mechanical impedance may be used directly in equivalent circuits. Moreover, like its electrical counterpart, acoustic impedance can be realized in either lumped or distributed form. When one or more dimensions of an acoustic system are not small as compared to the wavelength of sound, it is no longer possible to describe the system by lumped parameters.

Acoustic Compliance. Acoustic compliance is that parameter of an acoustical circuit which accounts for volume displacement without acceleration (5). The physical law governing the compression of a volume of acoustic medium is

$$p(t) = \frac{1}{C_a} \int U(t) dt \quad (10)$$

where $p(t)$ is the instantaneous pressure acting to compress the volume of the medium, $U(t)$ is the instantaneous volume velocity of the medium flowing into the volume, and C_a is the acoustic compliance. In the steady state of an acoustic pressure varying sinusoidally at an angular frequency ω ,

$$P = \frac{U}{j\omega C_a} \quad (11)$$

where P and U are complex rms quantities. It is evident from Eq. (11) that acoustic compliance has the units of $\text{m}^5 \cdot \text{N}^{-1}$. Note also that the acoustic impedance of a purely compliant element is

$$Z_a = -j \frac{1}{\omega C_a} \quad (12)$$

As a circuit element, acoustic compliance is a two-terminal device, but one of its terminals must always be at "ground potential," that is, one terminal is the outside of the enclosure containing the compressed medium. The reciprocal of compliance, called stiffness, is equally prevalent, the choice being a matter of convenience. Moreover, the adjective *dynamic* is sometimes used with the terms compliance and stiffness, especially in situations where the values of the static and dynamic parameters are substantially different.

Acoustic Inertia. The acoustic mass or inertance, m_a , is a quantity associated with a mass of an acoustic medium accelerated by a net force which acts to displace the medium without appreciably compressing it:

$$p(t) = m_a \frac{dU(t)}{dt} \quad (13)$$

As a consequence of this definition, the dimensions of acoustic mass are $\text{kg}\cdot\text{m}^{-4}$. In the sinusoidal steady state, the value of the parameter m_a multiplied by ω yields the magnitude of the imaginary part of the acoustic impedance resulting from the inertia or effective mass of the medium:

$$p = j\omega m_a U \quad (14)$$

As an acoustic-circuit element, acoustic mass can be realized as a tube open at both ends. The tube must have rigid walls and be short enough so that the medium in it moves as a whole without appreciable compression. Furthermore, the assumption of zero pressure at the ends requires that the tube diameter be small relative to the wavelength of sound in order to reduce its radiation impedance. The effect of the radiation impedance is to increase slightly the apparent length of the tube. One end of an acoustic mass is usually driven by a source, while the other end can be terminated in a boundless medium, a larger cavity, or a damping element.

Acoustic Resistance. Resistance may be contributed to an acoustic system by a number of different causes, but it is always associated with dissipative losses. The physical law defining acoustic resistance is

$$p(t) = R_a U(t) \quad (15)$$

Being the real part of acoustic impedance, Z_a , acoustic resistance has the same unit dimensions, $\text{Pa}\cdot\text{s}\cdot\text{m}^{-3}$, also called the acoustic ohm. In lumped form it can be realized by having the sound go through short capillary tubes or a fine-mesh screen.

The Helmholtz Resonator. A very instructive example of an acoustic system containing the three basic elements (compliance, inertance, and resistance) is provided by the Helmholtz resonator, shown in Fig. 3 alongside its electrical analog, the series RLC circuit. The lumped-parameter values can readily be calculated in terms of its dimensions and properties of the acoustic medium, as is done in Ref. 13 and other textbooks on acoustics. Thus,

$$C_a = \frac{V}{\rho_0 c^2}, \quad m_a = \frac{\rho_0 L}{S}, \quad R_a = \frac{R_r}{S^2} + R_{eq} \quad (16)$$

where R_r is the radiation resistance of the neck opening, and R_{eq} represents the effect of all damping present in the system, and all other quantities have the previously defined meaning.

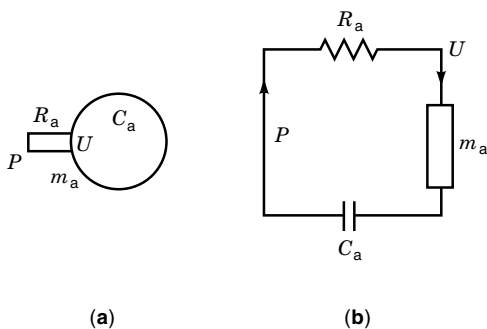


Figure 3. A Helmholtz resonator consisting of acoustic compliance, mass, damping, and radiation resistance (a), and a schematic representation of its electrical equivalent (b).

Acoustic Absorption

When sound traverses matter, be it solid, liquid, or gaseous, the acoustic intensity decreases with distance. This fact is a consequence of the properties of the wave itself and the medium through which it travels. In the absence of boundaries this propagation loss is attributable mainly to spreading (e.g., the so-called inverse r^2 loss), scattering, and absorption by the medium. When boundaries are present, any diminution of intensity upon reflection from them is governed by the nature of the reflecting surfaces which in addition to presenting an impedance mismatch and absorbing acoustic energy can diffuse the acoustic field by virtue of their roughness. Absorption is a consequence, primarily, of the viscosity of the medium, and is usually accounted for by a complex value of the propagation velocity. Accordingly, the wave number must be complex also. In a boundless medium, this is accounted for by an extra factor in the spreading-loss expression. Likewise, when sound is reflected from a medium which has absorption, the transmitted and reflected waves are affected: inhomogeneous waves are produced along the boundary, and total internal reflection cannot take place (14).

Absorption Coefficient. The simplest theory of absorption yields the following expression for the acoustic pressure of a wave propagating in a viscous medium:

$$p = p_0 e^{-\alpha x} e^{j(\omega t - kx)} \quad (17)$$

where α is the absorption coefficient, x is the distance traveled by the wave, ω is the angular frequency, and k is the wave number. Accordingly, one of the standard methods of measuring sound absorption coefficient is the classical impedance-tube method (15). Yet even in this simple model, the absorption coefficient is a function of not only viscosity, but also of frequency and the speed of sound. To obtain good agreement between predicted and measured values of α , other processes may have to be built into the theoretical model. For example, in air, humidity becomes important, while in sea-water, chemical relaxation measurably affects the results at low frequencies.

An alternate definition of the sound absorption coefficient is as the fraction of the sound energy incident on a surface which is absorbed or dissipated. A convenient measure of the effectiveness of any absorptive treatment is, quite naturally, the reverberation time, that is, the time interval during which the sound is audible after its source has been turned off. Quantitatively, the reverberation time is defined as the time required for the level of the sound to drop by 60 dB.

The preceding discussion considered the effect of absorption on the propagation of sound in a definite direction. There are, however, many situations where sound is not expected to travel from one specified point to another, but rather to fill a volume of space with a uniform intensity. Such a sound field is called diffuse or reverberant, and is usually produced by multiple reflections. This process lends itself to straightforward analysis, which was originally performed by Sabine and can be found in one of several variants in most books on acoustics, for example in chapter 6 of Ref. 7. The analysis is based on a few realistic assumptions and involves the room dimensions as well as properties of the absorbers. When the average absorption coefficient, $\bar{\alpha}$, for the room is small, the Sabine formula in metric

units reads:

$$T = 0.161 \frac{V}{S\bar{\alpha}} \quad (18)$$

where T is the reverberation time, V is the room volume, and S is the area of the bounding surfaces. This formula can also be used to measure the average absorption coefficient of the room, and in turn the absorption coefficient of an individual absorber. Since

$$\bar{\alpha} = \frac{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n}{S_1 + S_2 + \cdots + S_n} \quad (19)$$

where S_i and α_i are corresponding areas and absorption coefficients, any α_i can be calculated if all the others are known and $\bar{\alpha}$ is measured. This is the basis of a standard method of measuring the absorption coefficient of materials.

DATA ACQUISITION, VALIDATION, AND ANALYSIS

Some of the most important measurement activities should take place before data are acquired. Planning for data acquisition and analysis should address several issues: the selection of transducers, evaluation of instruments to be used in the measurement system, measurement location and duration, and protocol for estimating uncertainties in the measurement. This is particularly important if a new method of measurement is contemplated or the procedure of a standard method is not adequately documented. Prior to data acquisition, the measurement system must undergo a complete calibration of all components as well as an end-to-end calibration of the system. The lack of a proper end-to-end electrical calibration is probably the most important source of errors in dynamic data measurements. Modern instruments, though fast and versatile, still require their user to be fully conversant with the different types of averaging, sampling rate, filtering, and statistical analysis that are available.

It is of paramount importance to carry out a reliable validation procedure prior to the analysis of data. Validation in this context means ensuring that the recorded signal correctly represents the measurand of the variable required for subsequent analysis. Signal clipping, excessive background noise, inadequate bandwidth of instrumentation, undesirable coupling, spurious signal pickup all make validation an essential component of the measurement procedure. Poor coherence between the input calibration signal and the output at the relevant channels of the data acquisition system can readily invalidate seemingly good results. The coherence function should have a value in excess of 0.99 at all frequencies of interest (16). If the practice is to record the measured signals on either analog or digital storage devices for later analysis, the signals can be validated prior to analysis. When data are acquired and analyzed on-line, complete validation prior to analysis may be impracticable; however, many of the desired data validation operations can be automated by knowledge-based computer programs, resulting in substantial reduction in time and labor (17).

Characterization of Measuring Instruments

In its elemental form, an instrument for measuring an acoustic variable is a transducer which converts acoustical energy into

another form, usually electrical. Since both the input and output measurands are phasor quantities, it is very convenient to describe performance of the transducer by its transfer function, that is the ratio of the output to the input. This function is called the sensitivity of the transducer and, like any complex quantity, may be expressed in terms of magnitude and phase, as in Fig. 1. Each of these components exhibits a definite dependence on frequency and, to a lesser extent, on environmental parameters, such as temperature, humidity, static pressure, magnetic field, and so on. For a specific application consider first the microphone, probably the most important instrument in acoustical metrology.

Measurement of Microphone Response. When a microphone is placed in a sound field, the field will be disturbed. The disturbance results in the sound pressure on the diaphragm being different from that of the undisturbed field. This pressure deviation is accounted for by the so-called free-field correction, which depends on the dimensions of the microphone, the shape (including any protective grid), and the angle of incidence of the sound. When sound arrives from many directions, the free-field response must be averaged over all directions to yield the correction necessary to compensate for the disturbing influence of the microphone. This is known as the random-incidence correction, required in situations such as machine shops and reverberation chambers. See Fig. 17.9 of Ref. 18.

There are two methods for expressing microphone frequency response: the free-field sensitivity and the pressure sensitivity. The free-field sensitivity is defined as the ratio of the microphone's electrical output to pressure applied on the diaphragm when the applied pressure is that which would exist at the microphone location in the absence of the microphone. The pressure sensitivity of a microphone is defined as the ratio of the electrical output to pressure at the diaphragm when the pressure is uniform and in phase over the entire sensing surface of the microphone. Thus, the actual response of a microphone in a sound field can be represented by the sum of its pressure response and the applicable free-field correction. The free-field and pressure responses of a microphone are identical at low frequencies, but begin to diverge at frequencies where the sound wavelength becomes comparable to the dimensions of the microphone. This information can be presented in absolute terms or relative to the value at some reference frequency.

The free-field response data is the most applicable description of microphone sensitivity as a function of frequency when the microphone is operated in an open space, several wavelengths removed from a sound source and any reflecting boundaries. The free-field correction curves are relevant only when the sound field is diffracted or reflected by the microphone body. When the microphone forms part of a natural boundary, the pressure response for sound normally incident upon the diaphragm should be used. For example, when measuring the sound pressure acting on an aircraft wing, the microphone should ideally be mounted so that its diaphragm is flush with the surface of the wing. As long as the diaphragm's input impedance is large compared with the radiation impedance loading the diaphragm, the microphone will not disturb the field, and the pressure response will adequately represent the microphone's performance. There exist standard methods of calibrating microphones, based on the above definitions, as well as the so-called pistonphone calibrator which is easier to use, but less accurate. The pistonphone calibration is most

suitable for use at frequencies where the wavelengths are very long compared with the dimensions of both the microphone and the volume of air entrapped within the pistonphone. Recently the pistonphone technique has been adapted to phase-response measurements (Chapter 9 of Ref. 18).

Regardless of the method employed the microphone response is typically displayed in two plots: amplitude vs. frequency, and phase vs. frequency, as illustrated in Fig. 1. Until the 1980s users were interested mostly in the magnitude of the microphone sensitivity, but in recent years the demand for phase response information has greatly increased, primarily because of its importance to sound intensity measurements. What is needed in practice is the relative phase response of two microphone systems, that is, the phase difference between them. Methods involving phase angles of microphone response are the subject of active research (19).

Sound-Level Meter. Although the availability of a calibrated microphone together with an amplifier and indicating meter is all that is required to construct an instrument for measuring sound pressure, such instruments have not come into common use. Instead, a related instrument with increased capabilities has evolved. It is the sound level meter, which has become the basic instrument for all sound- and noise-level measurements, particularly those made in the field. Figure 4 shows a photograph of a representative sound level meter. These instruments come in different shapes and sizes as well as signal-processing capabilities, but all comprise a microphone, an amplifier, frequency weighting networks, and an output-indicating meter, and are constructed according to national and international standards (20). The sound level meter measures the weighted sound pressure level, defined as follows:

$$SL = 10 \log \frac{(p)_{av}^2}{(p)_{ref}^2} \quad (20)$$

where p is the acoustic pressure being measured, and p_{ref} is a reference pressure, which must always be specified. Normally, for airborne sound $p_{ref} = 20 \mu\text{Pa}$, and for underwater sound, $p_{ref} = 1 \mu\text{Pa}$. Although the unit of the displayed quantity is the decibel, Eq. (20) will yield the same value as Eq. (1) only when no spectral weighting, resulting in a flat frequency response, is used. Usually, however, one of three or four (A, B, C, D) weighting characteristics is employed, which modifies the response at low frequencies, and consequently the indicated average reading. The precise dependence of these weighting networks on frequency has been standardized (20). (Originally they were intended to make the sound level meter respond approximately like the human ear to the loudness of the sounds being measured.) When reporting data in an isolated context, one should state not only the reference pressure, but also the weighting, for example: “the A-weighted $SL = 43 \text{ dB}$.”

Displacement Transducers. In acoustics it is the vibratory displacement of a surface, rather than the particle displacement of the acoustic medium, that is commonly measured. To accomplish this function a number of successful methods have been developed, using either contacting or noncontacting probes. The method of choice appears to be some type of optical interferometry, especially since the advent of affordable good-quality lasers. An interferometer, based on one of several



Figure 4. An example of the basic instrument for acoustical measurements, a sound level meter (courtesy of Brüel & Kjær Instruments, Inc.).

classical configurations, usually Michelson or Mach-Zehnder, can be custom built from off-the-shelf components, is easy to set-up, and offers the user the assurance of accuracy traceable to the wavelength of the laser light. Many systems are available commercially, and some are designed to respond to either displacement or velocity. In principle, the behavior of an interferometer does not depend on frequency, but practical instruments do have limitations on dynamic range as a function of frequency.

Velocity Transducers. Faraday's law of electromagnetic induction is the principle of the moving-coil transducer, an intrinsically velocity transducer that finds application in both microphones and vibration pickups. The voltage induced in the coil is, therefore, proportional to the relative velocity of the coil motion. An instrument that has gained great popularity in recent years is usually referred to as the laser vibrometer, marketed by manufacturers worldwide for both laboratory and field use. It operates on the Doppler principle whereby light reflected from the vibrating surface experiences a frequency shift which is proportional to the instantaneous velocity of the vibration. In the fiber-optic version the lens probe is decoupled

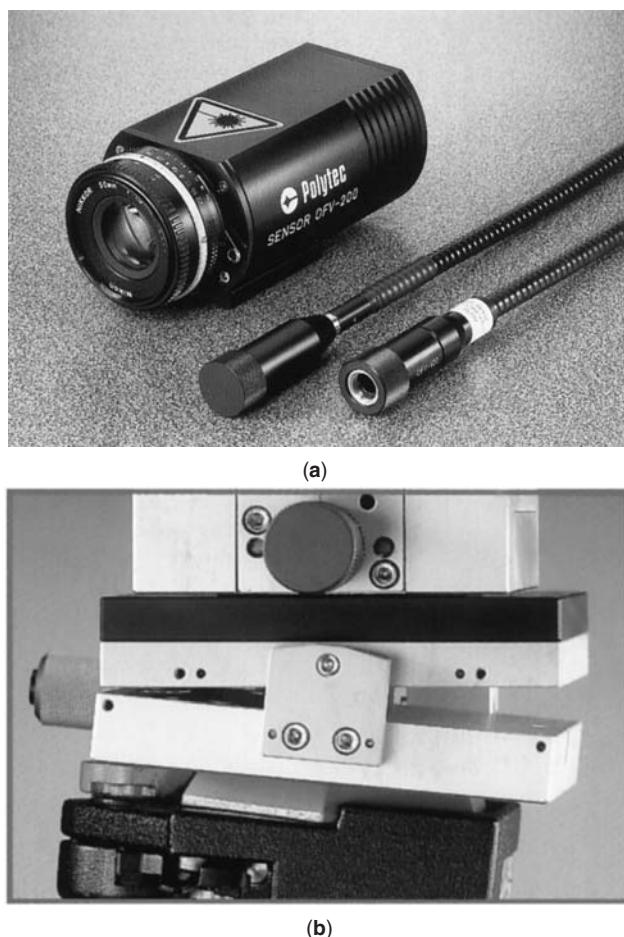


Figure 5. Components of a laser Doppler vibrometer, a versatile vibration measuring instrument: (a) sensor and optical fiber probes; (b) positioning stage for x , y , z , tip, and tilt adjustment (courtesy of Polytec PI, Inc.).

from the interferometer system by an optical fiber probe allowing the experimenter to reach locations not accessible otherwise. Figure 5 shows one of several commercially available laser vibrometers.

Accelerometers. Transducers that directly sense particle acceleration and that are employed to measure vibrations of solid surfaces are by far more common than those which directly sense their velocity or displacement. There exist several general types of accelerometers with many specialized implementations of each type. The piezoelectric accelerometer usually is the optimum choice because of its very wide frequency and dynamic ranges. For specific information the reader is referred to publications on vibration measurements (21,22) and product literature.

Mechanical-Impedance Heads. In the measurement of mechanical impedance, a force transducer and a velocity transducer must be used simultaneously. Some manufacturers supply devices containing a force transducer and an accelerometer, in which case the output of the latter must first pass through an integrator. Like all joint measurements, mechanical impedance measurements present special problems. The force and

velocity transducers must be much closer together than the shortest wavelength of interest in order for the two transducers to be considered as a single transduction system. Moreover, because of the complex nature of impedance, both amplitude and phase responses of each transducer must be accurately known. Mechanical resonances of structures other than the one whose impedance is measured can also affect the results.

Application of Correlation Techniques

Correlation methods originally were used to obtain statistical measures of randomly varying signals, for example to characterize the sound field in an enclosure. However, it soon became apparent that because of their inherent advantages, methods based on measuring auto- and cross-correlation of acoustic signals could rival many measurements traditionally using deterministic signals. The chief among these advantages is the ability to discriminate against noise, the latter being defined as the unwanted part of a signal. This is particularly important in situations of low signal-to-noise ratio. Another is the opportunity to design pseudo-random test signals which can generate the desired results more rapidly than measurements using deterministic signals. To illustrate we present two specific applications, particularly relevant to this article: measurements of the speed of sound and of transmission loss.

The feasibility of measuring the speed of sound follows from the form of the cross-correlation function of the same signal recorded at two points along the path of propagation. If the separation between the microphones is d , the maximum of the correlation function $R_{12}(\tau)$ occurs when the time delay $\tau = d/c$. Knowing the separation and noting the value of time delay, the speed of propagation, c , is readily obtained (23). In this respect the correlation method resembles the pulse method. They differ primarily in the nature of the test signal, which should be random and broadband for the correlation method, but for the pulse method it usually takes the form of a pulse-modulated sine wave. This method is particularly convenient for measurements in solids or liquid-filled tubes, where more than one mode may simultaneously be excited (24). As an example, consider the measurement of the speed of propagation of breathing waves in a tube with soft walls (25).

As a second example, consider the problem of measuring sound isolation and damping (23,24). The experimental set-up, shown schematically in Fig. 6, consists of a loudspeaker driven by a noise generator and power amplifier, a microphone with

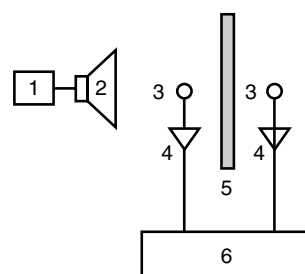


Figure 6. Equipment layout diagram for the measurement of sound transmission loss of an absorptive panel by a correlation method: (1) random noise generator and amplifier; (2) loudspeaker; (3) microphones; (4) signal conditioners; (5) test panel; (6) correlator and data recorder.

signal conditioner, and a correlator. The microphone picks up the direct sound plus any reflected and diffracted sounds. All those signals have different arrival times and, consequently, the maxima of their correlation with the source signal are appropriately delayed. The transmission loss, TL, equals the difference between the first maxima of the correlation functions, $R_0(\tau)$ and $R_p(\tau)$, corresponding respectively to the direct paths in the absence and presence of the test panel. (The delay in the signal that has passed through the panel is caused by the sound speed in the panel material being slower than in air.)

$$TL = 20 \log \frac{R_0(\tau_1)}{R_p(\tau_1)} + 20 \log \frac{K_0}{K_p} \text{ dB} \quad (21)$$

where K_0 and K_p are the respective microphone-channel gains.

For accurate results it is important that the level of the transmitted sound greatly exceed the ambient noise level. The constraint on the sound source is that it have sufficient bandwidth.

Slight variations of the technique just described can be used to measure the reflection coefficient of a wall and the absorption coefficient of a reverberant room (24). It should also be noted that the correlation measurements here described, as well as all others, have their counterparts in the frequency domain and can be measured in terms of spectral functions.

ELECTROMAGNETIC-ACOUSTICAL ANALOGIES

There exist well-defined analogies between acoustical and electrical systems. Researchers familiar with electromagnetism and circuit theory can readily transfer much useful information to acoustics by a judicious application of a few rules and by keeping in mind the fundamental distinctions between sound and electromagnetic waves, particularly with regard to the variables in each system. For example, the most often measured quantity of an acoustic field, the sound pressure, is a scalar satisfying the Helmholtz equation, while the electromagnetic field quantities are vectors satisfying Maxwell's equations. These differences notwithstanding, the results of most boundary-value problems are mathematically analogous (26).

As an illustration consider the two-dimensional reflection problem. The acoustic pressure, p , can be shown (27) to be analogous to the tangential component of the electric field, E_y , and the normal component of particle velocity, u_z , to be analogous to the magnetic field, H_x , as evidenced by the following relationships:

$$u_z = \frac{1}{j\omega\rho} \cdot \frac{\partial p}{\partial z} \quad \text{and} \quad H_x = \frac{c}{j\omega\mu} \cdot \frac{\partial E_y}{\partial z} \quad (22)$$

Consequently, the reflection coefficients in the two cases differ only by their constant factors, which themselves have similar physical meaning, if one recalls that ρc and $\sqrt{\epsilon/\mu}$ characterize the wave impedance of the respective media.

The same analogy can generate other important results. For example, the well-known Snell's and Brewster's laws have identical forms in both systems. However, in three-dimensional problems, it is usually not possible to find acoustical counterparts to a general solution of Maxwell's equations, except for the so-called geometrical optics case, when Kirchhoff's approximation holds (27). It is because under that condition it is

Table 2. Electroacoustical Analogies

Acoustical	Electrical
Sound pressure	Potential difference (voltage)
Volume velocity	Electric current
Volume displacement	Electric charge
Acoustic mass (inertance)	Inductance
Acoustic compliance	Capacitance
Acoustic resistance	Resistance
Acoustic impedance	Electric impedance

permissible to treat the electromagnetic-field quantities as scalars.

The analogies extend beyond the variables describing the behavior of systems in terms of field quantities. Individual elements of electromagnetic and acoustic circuits are characterized by analogous parameters that preserve the same frequency and space dependence, for example, when mechanical impedance is represented by its electrical analog. This can be especially useful in transferring results of electric-circuit theory to the design of acoustical and vibration measurements. The formulation of circuit-theory analogies follows definite rules, discussed by Olson (28), Kurtze (29), and many others. Table 2 shows some commonly employed analogies between acoustical and electrical variables and parameters.

As an illustration consider the task of calculating the transmission loss provided by a rigid wall whose area mass density is m . This basic problem of sound transmission between two media can also be solved by the application of Thevenin's theorem, resulting in the circuit of Fig. 7(b), where acoustic pressure is the analog of voltage, and particle velocity is the analog of current (30). Accordingly,

$$u = \frac{2p_1}{2\rho c + j\omega m} \quad \text{and} \quad p_2 = p_1 \frac{2\rho c}{2\rho c + j\omega m} \quad (23)$$

Hence, the transmission loss, in decibels, becomes

$$TL = 20 \log \left| \frac{p_1}{p_2} \right| = 20 \log \sqrt{1 + \left(\frac{\omega m}{2\rho c} \right)^2} = 10 \log \left[1 + \left(\frac{\omega m}{2\rho c} \right)^2 \right] \quad (24)$$

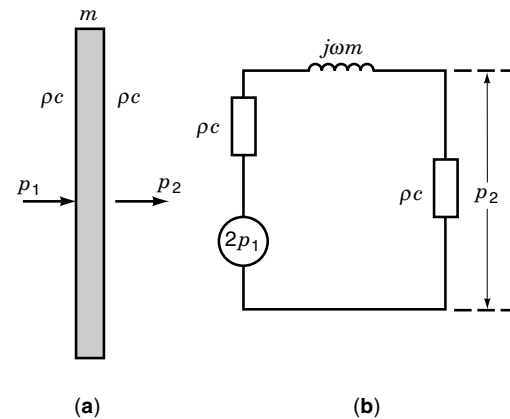


Figure 7. Derivation of an expression for the sound transmission loss by the application of Thevenin's theorem to an equivalent electrical circuit: (a) an acoustic panel of area mass density m , placed in a medium of specific acoustic impedance ρc ; (b) an analogous electrical circuit.

While the answer originally sought is correct, the reader is cautioned not to regard the two parts of Fig. 7 as equivalent in every respect. To understand this warning, consider, for example, the particle velocities on both sides of the partition. Clearly, they are in the same ratio as the pressures since the acoustic impedances on both sides are equal to ρc , but according to the schematic diagram the same particle velocity “flows” on either side of the wall.

It likewise follows that the matrix equations representing acoustical and electric systems are analogous. This is very fortuitous because the elegant and easily understood theory of multiports can serve as the foundation for important analytical and experimental investigations. For example, two- and three-port models of microphones and accelerometers facilitate the development and implementation of methods for calibrating them. When two microphones are coupled by an acoustic medium, the system can be represented by a four-port model, which provides the basis for a reciprocity calibration of standard microphones, elucidated in chapter 4 of Ref. 18.

BIBLIOGRAPHY

1. ANSI/IEEE Std. 260.4-1996, *American National Standard Letter Symbols and Abbreviations for Quantities Used in Acoustics*, New York: Acoust. Soc. Am., 1996.
2. F. Trendelenburg, *Einführung in die Akustik*, 3rd ed., chap. 2, Stuttgart: Springer-Verlag, 1961.
3. ANSI S1.13-1995 (ASA 118-1995), *American National Standard, Measurement of Sound Pressure Levels in Air*, New York: Acoust. Soc. Am., 1995.
4. D. N. Keast, *Measurements in Mechanical Dynamics*, chap. 6, New York: McGraw-Hill, 1967.
5. L. L. Beranek, *Acoustics*, chap. 1, New York: McGraw-Hill, 1954.
6. IEEE Standard 100-1996, *IEEE Standard Dictionary of Electrical and Electronics Terms*, 6th ed., New York: IEEE Press, 1996.
7. A. D. Pierce, *Acoustics*, 3rd ed., New York: Acoustical Society of America, 1994.
8. L. Maestrello, Relationship between acoustic energy density flux near the jet axis and far-field acoustic intensity, in H. T. Nagamatsu (ed.), *Aeroacoustics: Jet and Combustion Noise; Duct Acoustics*, New York: AIAA with MIT Press, 1975.
9. A. J. Zuckerwar, Theoretical response of condenser microphones, *J. Acoust. Soc. Amer.* **64**: 1278–1285, 1978.
10. M. J. Crocker (ed.), *Encyclopedia of Acoustics*, vol. 1, New York: Wiley, 1997.
11. A. G. Webster, Acoustical impedance, and the theory of horns and of the phonograph, *Proc. Natl. Acad. Sci.*, **5**: 275–282, 1919.
12. L. L. Beranek, *Acoustical Measurements*, revised ed., New York: American Institute of Physics, 1988.
13. L. E. Kinsler et al., *Fundamentals of Acoustics*, 3rd ed., chap. 10, New York: Wiley, 1982.
14. D. G. Tucker and B. K. Gazey, *Applied Underwater Acoustics*, chap. 3, Oxford: Pergamon, 1966.
15. M. G. Jones and P. E. Stiede, Comparison of methods for determining specific acoustic impedance, *J. Acoust. Soc. Amer.*, **101** (5): 2694–2704, 1997.
16. R. B. Randall, *Frequency Analysis*, 2nd ed., chap. 6, Naerum: Bruel & Kjaer, 1977.
17. Military Handbook *Guidelines for Dynamic Data Acquisition and Analysis*, prepared by Harry Himelblau, Jet Propulsion Laboratory, Pasadena, CA, 1990.
18. G. S. K. Wong and T. F. W. Embleton (eds.), *AIP Handbook of Condenser Microphones*, New York: AIP Press, 1995.
19. M. J. Crocker (ed.), *Encyclopedia of Acoustics*, vol. 4, New York: Wiley, 1997.
20. ANSI S1-4, *American Standard Specification for General-Purpose Sound-Level Meters*, New York, 1961.
21. C. M. Harris (ed.), *Handbook of Shock and Vibration*, 4th ed., New York: McGraw-Hill, 1995.
22. M. Serridge and T. R. Licht, *Piezoelectric Accelerometer and Vibration Preamplifier Handbook*, Naerum: Bruel & Kjaer, 1987.
23. A. K. Novikov, *Correlation Measurements in Ship Acoustics*, chap. 3, Leningrad: Sudostroyeniye, 1971.
24. I. I. Klyukin and A. E. Kolesnikov, *Acoustical Measurements*, 3rd ed., chap. 8, Leningrad: Sudostroyeniye, 1982.
25. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, chap. 11, New York: McGraw-Hill, 1953.
26. H. Severin, *Zur Analogie Akustischer und Elektromagnetischer Randwertprobleme*, in *Akustische Beihefte*, Bd. 1, Stuttgart 1959.
27. E. L. Shenderov, *Wave Problems of Hydroacoustics*, Leningrad: Sudostroyeniye, 1972.
28. H. F. Olson, *Dynamical Analogies*, 2nd ed., Princeton: Van Nostrand, 1958.
29. G. Kurtze, *Physik und Technik der Lärmbekämpfung*, Karlsruhe: 1964.
30. I. I. Klyukin, *Fight against Ship Noise and Sonic Vibration*, 2nd ed., chap. 3, Leningrad: Sudostroyeniye, 1971.

M. ROMAN SERBYN
Morgan State University

TIME, FREQUENCY, AND PHASE

FREQUENCY AND TIME INTERVAL METERS

Frequency and time interval meters are broadly applied to measure frequency, period, and other time-related waveform parameters, including pulse duration, rise and fall times, duty cycle, frequency or period ratio, and also the time relationships between two different signals. The most popular meters, digital meters, measure frequency and time by comparing an unknown quantity to a reference. The reference sources of these meters are usually either built-in or externally applied to control the gate which passes the signal to the counters. The reference sources can work independently of the source of absolute time reference, or they can be synchronized with the absolute time reference to measure the actual time (1). The meters with references provide very accurate readings of frequencies or time intervals, including periods. Less accurate meters with calibrated tuned circuits or frequency-to-voltage converters are applied when less precise frequency or time measurements are required.

The detailed descriptions of the frequency and time standards, frequency stability, frequency transfer, clocks, oscillators, frequency synthesis, time and frequency measurements are included in the following articles of this encyclopedia: FREQUENCY STANDARDS, CHARACTERIZATION; FREQUENCY STABILITY; CLOCKS IN TELECOMMUNICATIONS; MEASUREMENT OF FREQUENCY, PHASE NOISE AND AMPLITUDE NOISE.

The frequency and time interval measurements are interrelated, and for higher frequencies, it is more convenient and more accurate to measure frequency and calculate time parameters. For lower frequencies, the period measurements are more reliable and the frequency is then calculated.

FREQUENCY METERS

Frequency represents the count of the number of events of a repetitive waveform occurring per unit of time. The accepted unit of frequency is hertz (Hz), corresponding to the number of cycles per second. Other, larger units are kilohertz (kHz or 10^3 hertz), megahertz (MHz or 10^6 hertz), gigahertz (GHz or 10^9 hertz), and terahertz (THz or 10^{12} hertz). Smaller units are millihertz (mHz or 10^{-3} hertz) and microhertz (μ Hz or 10^{-6} hertz).

The simplest frequency meters are passive or active absorption meters, which are applied for periodic signals whose frequencies are within the radio frequencies and also within microwave frequency ranges. As shown in Fig. 1, the meter is composed of an LC resonance circuit, a diode detector, and a meter indicating the maximum reading when the circuit is tuned to the frequency of the radiating source. The frequency ranges are selected by changing the coils. The variable capacitor used to tune the LC circuit is calibrated in frequency or wavelength units. The microwave absorption meters have their resonance circuits constructed as resonating cavities with tunable dimensions.

The other absorption meters, called dip oscillators (2), involve processes of energy absorption from external sources in

order to mix them with the internally generated sine wave signals (Fig. 2). When the oscillator frequency is the same as the source frequency, the meter indicates the minimum indication since the differential frequency component is a direct current (dc) signal, which does not pass through the circuit capacitor.

Frequency and time interval measurements can be performed applying an oscilloscope. The distance considered as a period of an observed wave is computed either visually or by a built-in digital counter. The visual reading leads to large errors reaching several percent. The built-in digital meters do not differ from classical frequency counters, which are described later in this text.

The block diagram of a single-channel, low-frequency (below several hundred megahertz) counter is depicted in Fig. 3, and the typical waveforms are shown in Fig. 4. The unknown frequency signal after shaping is applied to the main gate operating like an AND logic gate, which is opened by the main gate control signal for a defined amount of time. The gate timing is selected according to the preferred setting of the time base controlled by the time base generator or clock. The unknown frequency, f_x , is then compared with a built-in standard. Through the open gate the input signal pulses are applied to the digital counter that accumulates the count and displays it according to the measurement cycles programmed in the measurement cycle control unit. The total count on n pulses refers to the main gate control pulse duration, T_{base} , so the unknown

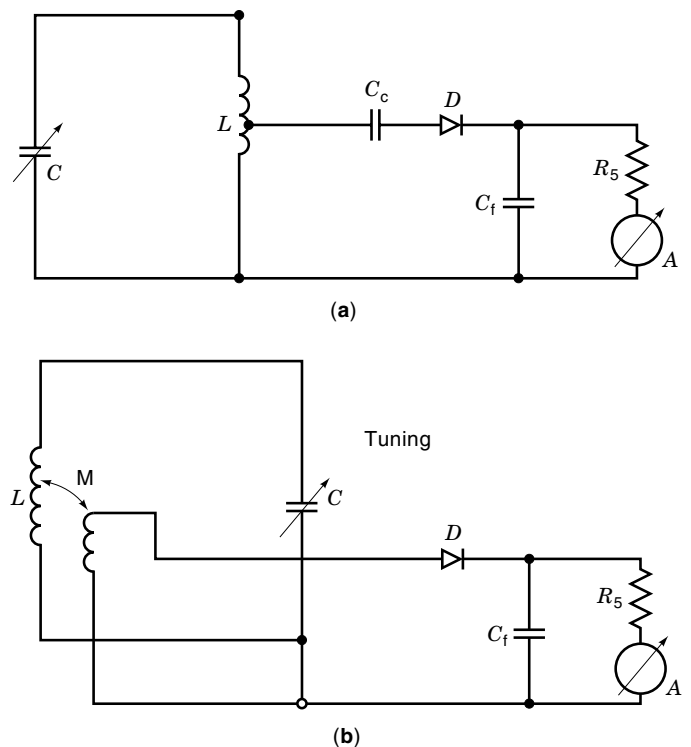


Figure 1. Passive absorption wavemeter. (a) Capacitive-coupled. (b) Transformer-coupled.

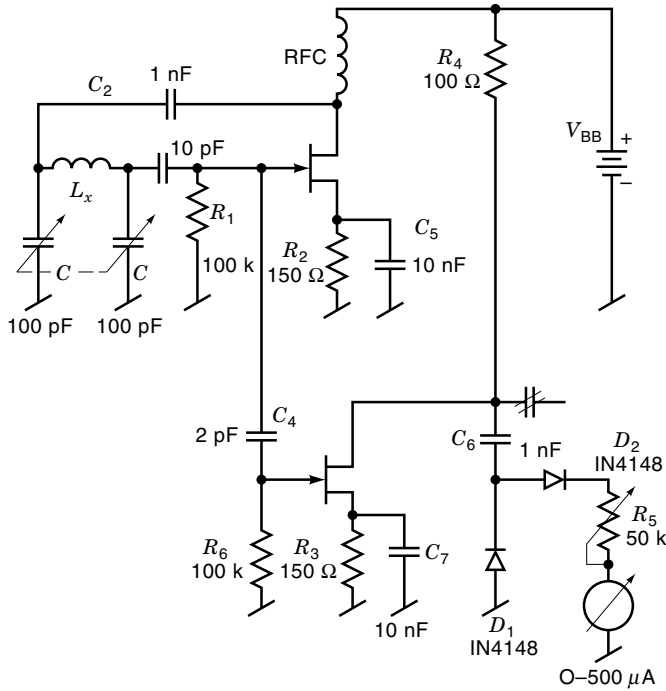


Figure 2. Active absorption wavemeter (dip meter).

frequency is expressed as

$$f_x = \frac{n \text{ pulses}}{T_{\text{base}}} \quad (1)$$

Inaccuracy of Eq. (1) is caused by the discrete nature of pulse counting. The error involved is called ± 1 count quantization error (3). Additional errors are added due to short- and long-term instability of the time base oscillator (clock) and due to the input conditioning unit trigger uncertainty. This error can be reduced by accumulation of many measurements and averaging. The error is then reduced $\sqrt{N}$ times, where N is the number of measurement cycles.

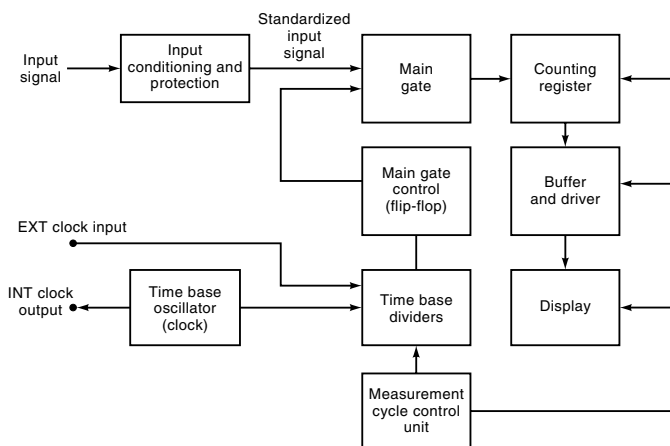


Figure 3. Frequency mode of a counter.

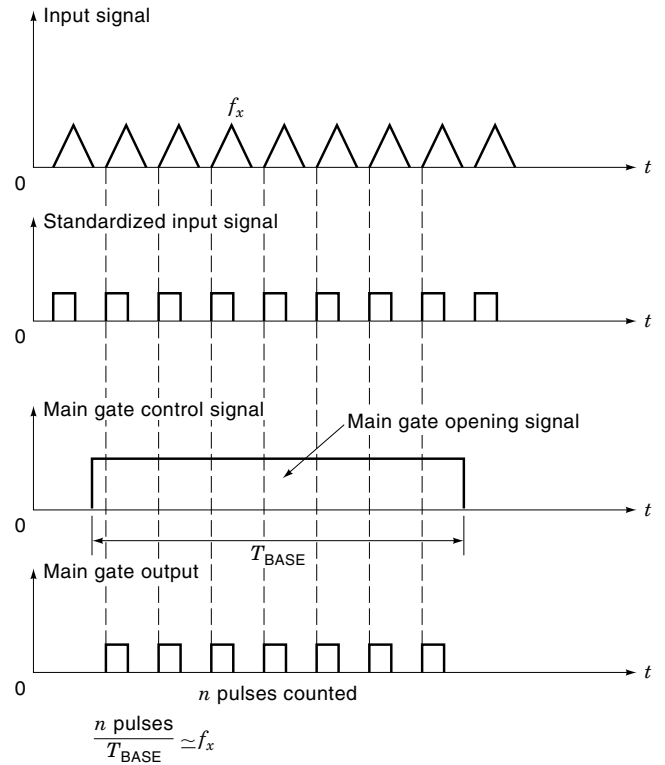


Figure 4. Waves in a frequency counter.

TIME INTERVAL MEASUREMENT

To achieve time interval measurement, we change the frequency counter configuration by swapping the function of the time base with the function of the input signal. The input signal opens the main gate for the pulses applied from the time base oscillator through the time base dividers, as shown in Figs. 5 and 6. The count accumulated in the counter register multiplied by the time base period yields the unknown time interval duration:

$$T_x = nT_{\text{clk}} \quad (2)$$

The basic errors are similar to the ones discussed before. More discussion of errors is included later in the article.

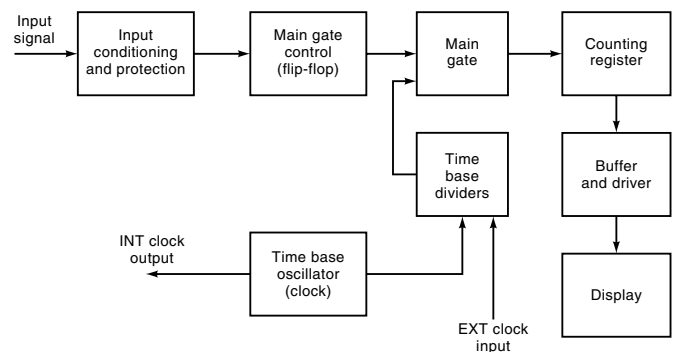


Figure 5. Period measurement mode of a counter.

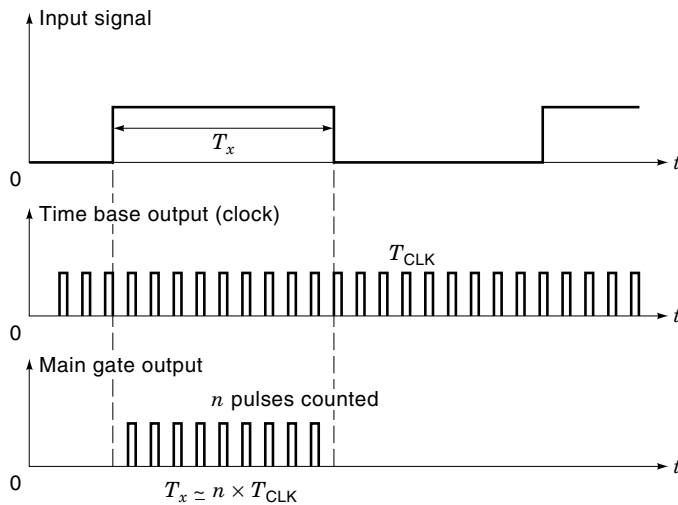


Figure 6. Waves in period meter.

A two-channel, start-stop, time interval meter can measure the time between two independent events or it can measure the period or the pulse width of one waveform. When two events are separately measured, the two signals are applied independently to two inputs, A and B. When a single signal is investigated, the inputs of the meter are connected together or the same signal is applied to both inputs. One input conditioning system is set to one signal slope, and the other system is set to the other slope. The process is carried out the way the previous meter operates (Figs. 7 and 8).

RATIO MEASUREMENT MODE

Two input counter shown in Fig. 9 can be used to find the ratio of two frequencies. One signal is applied to one channel and the other channel is used to gate the first signal. In all cases when the frequencies of the two channels are different, the smaller frequency signal gates the higher frequency signal. The accuracy of measurement is very critical when both frequencies are close one to each other. The effect of ± 1 count becomes then more significant and has to be reduced by averaging.

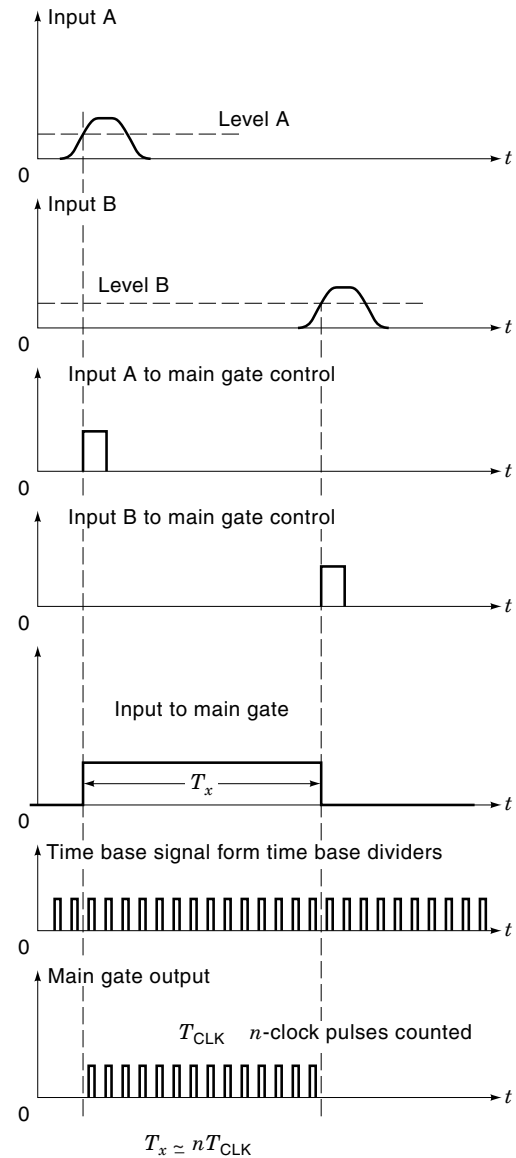


Figure 8. Waves in the time interval meter.

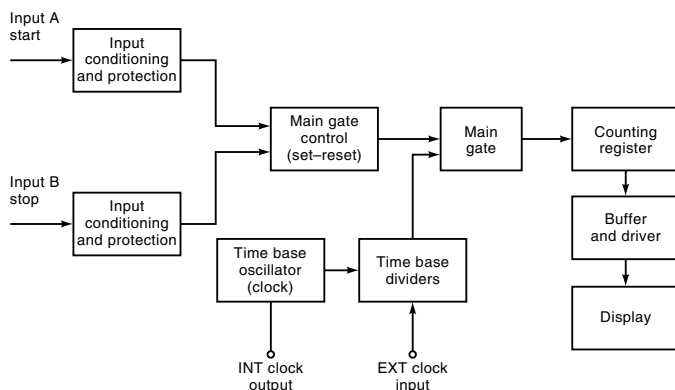


Figure 7. Time interval measurement mode.

RECIPROCAL COUNTER

The reciprocal counter measures the period and calculates the frequency from the period. At lower frequencies, the period measurements show smaller effect of the ± 1 count error than the frequency measurements. When the signal frequency exceeds the clock frequency, the frequency measurement is more accurate. Some reciprocal counters switch their mode from period to frequency measurement (3).

In the following sections the two major digital counter components are described. The first of them is the input conditioning unit, and the second is the time base oscillator.

INPUT CONDITIONING CIRCUIT

A block diagram of an input conditioning circuit is shown in Fig. 10 and its more detailed version of it in Fig. 11. The

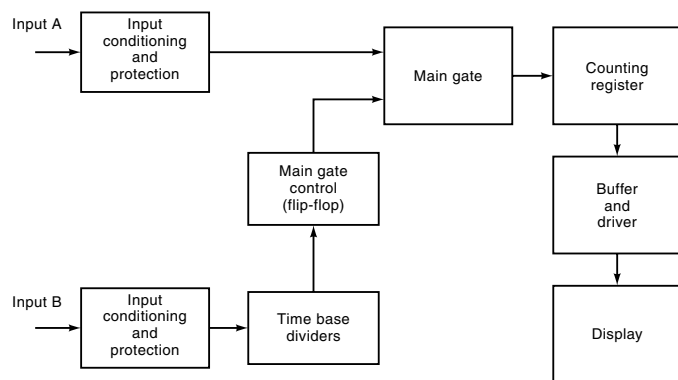


Figure 9. Ratio measurement mode of a counter.

major function of the circuit is to convert a periodic input signal into a standardized rectangular pulse wave whose pulses can be counted during gating time when the frequency is measured. When the period or the time interval is measured, the pulses from the gating signals are counted during the interval defined by the input signal. The input conditioning circuit, especially the trigger amplifier and the comparator with hysteresis, provides proper shaping of the input signals of various amplitudes, shapes, and speeds in order to drive digital circuits of the meter. At the same time, the effects of the input signal impurities, noise, and interference are also minimized by the trigger and comparator circuits. An amplitude limiter and sometimes an additional fuse protect the internal circuits when the input signal levels exceed the dynamic capability of the circuit.

DIGITAL CIRCUITS OF THE COUNTERS

Typical digital parts of a modern frequency or time interval meter include specialized integrated circuits. Each of them has a high frequency oscillator, main gate, time base counter, multidecade data counter, latches, seven-segment decoders, digit multiplexers, and multidigit multiplexed LED-display drivers. The circuits require very few external components, like crystal, several resistors, capacitors, display, switches, and power supply to operate. Circuits like these are capable of measuring frequency, period, and time interval, counting pulses, and determining frequency ratio. Maximum frequency range for

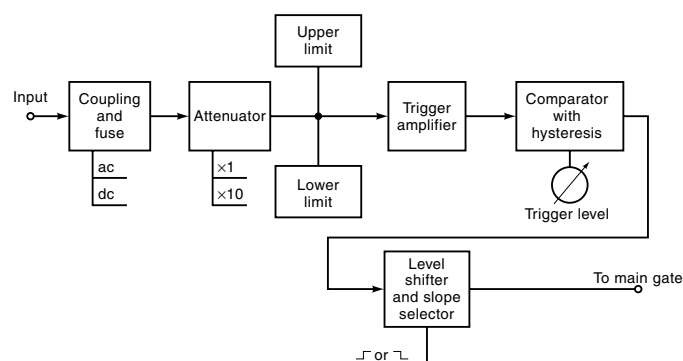


Figure 10. Block diagram of an input circuit.

direct count does not exceed several tenths of MHz. Additional prescaler circuits can extend the range of operation to several hundred MHz or even above 1 GHz. Some of these circuits, like UPB582, MC12009, and CD4017, are available on the market; some are custom designed.

The details of design and analysis of various circuits, including flip-flops, counters, analog and digital comparators, analog and digital phase-locked-loop circuits, and others are given in other articles of this encyclopedia. The related articles are: ASYNCHRONOUS CIRCUITS; ASYNCHRONOUS SEQUENTIAL CIRCUITS; COMPARATOR CIRCUITS; COUNTING CIRCUITS; PHASE-LOCKED LOOPS; SEQUENTIAL CIRCUITS; and MULTIVIBRATORS.

TIME BASE OSCILLATORS

The standard oscillators are discussed in FREQUENCY STABILITY. Here, only a short introduction is included to review the limitations. The time base oscillators are usually high-stability, high-accuracy crystal oscillators. In more expensive and complex solution, the crystal oscillator is mounted in an oven to maintain constant temperature during regular operation so the temperature effects on the oscillator frequency can be minimized. This guarantees relative inaccuracies close to 1×10^{-8} while nonstabilized oscillators, even if they are temperature-compensated [temperature-controlled crystal oscillators (TCXO)], show errors at least two orders of magnitude higher. The rubidium frequency reference oscillators have much better accuracy expressed as an error close to 10^{-10} (4). Much more expensive, cesium beam frequency standards have their inaccuracies on the order of 10^{-11} . The most complex, maser oscillators, mostly used as frequency standards, reach the level of error close to 10^{-15} . The accuracy figures cited here are related to one-year change (5). Due to the cost and weight, the majority of counters employ various crystal oscillators. Many of them have temperature-stabilized ovens for better stability [oven-controlled crystal oscillators (OCXO)]. The most popular frequencies of the time base oscillators are 1 MHz and 10 MHz, which allows for establishing very convenient meter frequency ranges. High precision crystals can be produced for fundamental mode of operation up to about 300 MHz (9). Two examples of block diagrams of the crystal oscillators are shown in Fig. 12. In the first circuit [Fig. 12(a)] the quartz operates as an inductor with a high-quality factor that maintains very good frequency stability. The second circuit [Fig. 12(b)] oscillates when the quartz impedance reaches its minimum, which occurs at the frequency of series resonance. The transistor solutions are shown in Fig. 13. Lower frequency signals necessary to control the counter main gate are obtained from the frequency dividers. A typical time base divider unit is shown in Fig. 14 for both modes of operation, frequency measurement and time interval (or period) measurement.

ERRORS IN DIGITAL FREQUENCY AND TIME INTERVAL METERS

An electronic counter measuring frequency or time interval involves (1) a reference oscillator or a clock and (2) a discrete counting of pulses within a certain amount of time defined by the gating process. The types of measurement errors include the following:

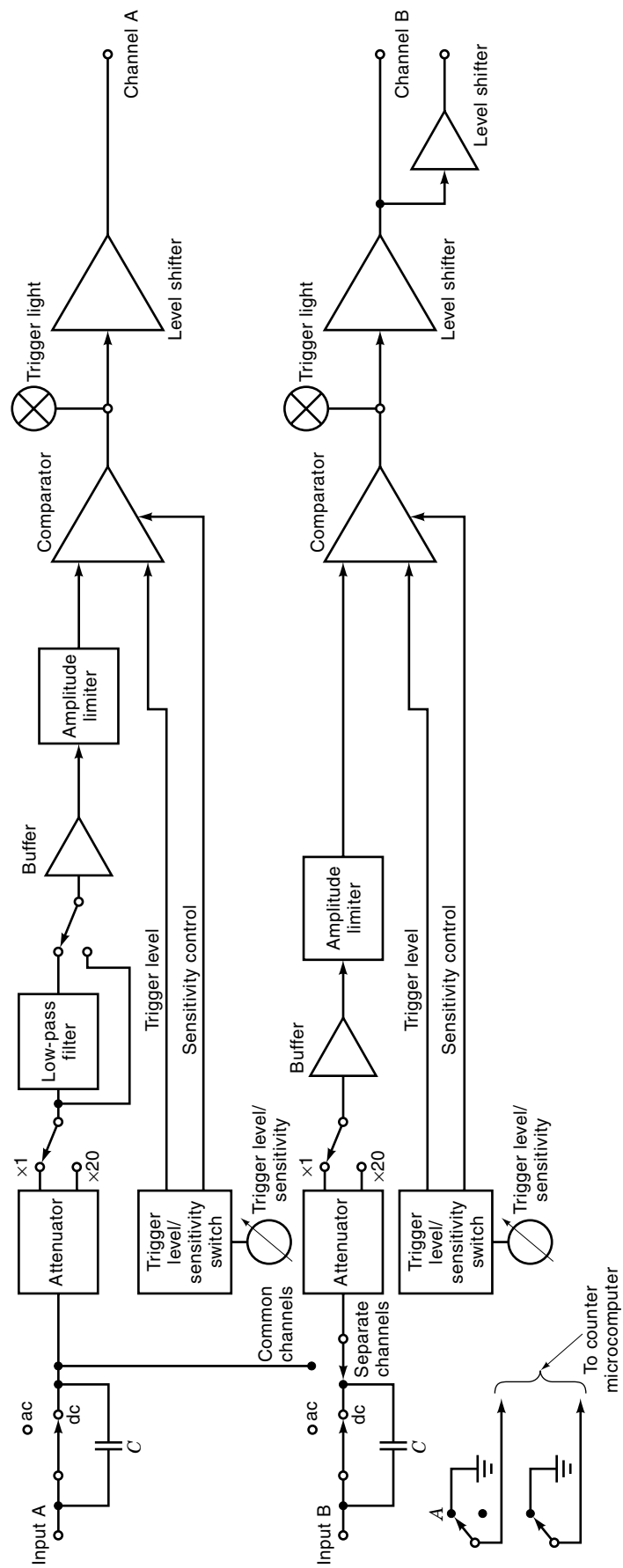


Figure 11. Input circuit of the HP 5315 counter.

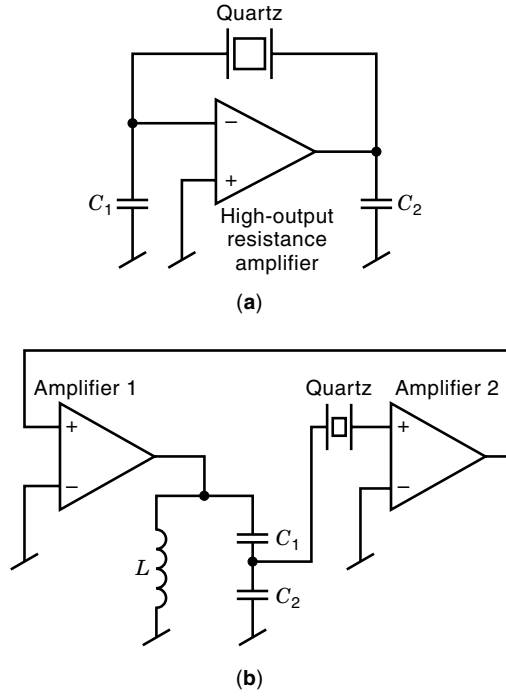


Figure 12. Block diagrams of crystal oscillators. (a) Collpits oscillator. (b) Crystal oscillator with quartz coupling two stages.

- Time base errors result from inaccuracy and from short-term and long-term instability of the time base oscillator.
- Trigger error, of random character, is caused by the noise. The error is a result of random instability of the gate pulse width caused by the noise of different thresholds.
- The systematic error, affecting only time interval measurements, originates in any differences between the start and stop trigger levels caused by hysteresis of the internal comparator, differences in delays, and the signal slew rate (3,6).
- The ± 1 count error is due to lack of synchronization of the gating pulses and the counted pulses.

The numerical values of different error components strongly depend upon the quality of the reference oscillator. The inaccuracies of the various clock circuits are between 10^{-8} and 10^{-6} .

The trigger error, important in the period and the time measurements, can be expressed in terms of the noise voltage levels and the triggering signal speed as follows (3):

$$e_t = \frac{\sqrt{2(e_i^2 + e_m^2)}}{dV_{in}/dT} \quad (3)$$

where

e_i is the root mean square (rms) voltage noise of the instrument input circuits,
 e_m is the rms voltage noise of the signal source limited by the counter bandwidth,
 dV_{in}/dT is the slew rate of the input signal at the trigger point.

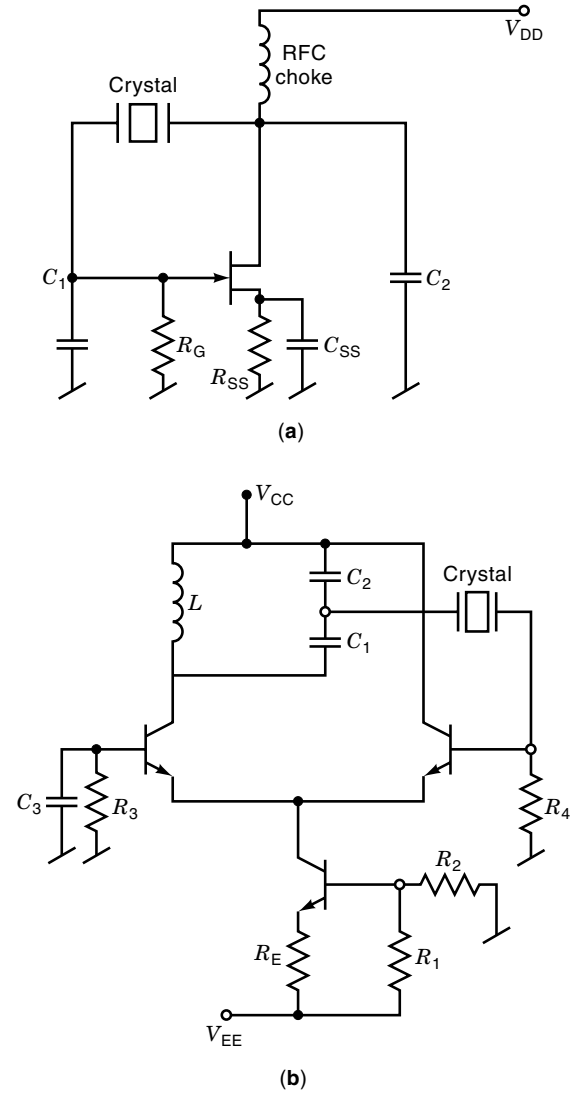


Figure 13. Circuit implementation of crystal oscillators. (a) Collpits oscillator. (b) Crystal oscillator circuit with quartz coupling two stages.

The level of the noise component can reach several hundred microvolts. The slew rates of the pulses vary from several volts per nanosecond to several volts per microsecond or less, which leads to the error values between 1 ps and 1 ns.

The systematic error, which strongly depends on the differences in delays the trigger level settings, affects the time interval measurements when short pulses are measured. This error is usually minimized through careful design and calibration.

The ± 1 count error, which is random in nature, becomes critical when a small number of pulses is counted within the frame of the gating pulse. Then a missing or additional pulse counted creates a relative error of great significance. These error effects can be summarized as follows:

$$\frac{\Delta f}{f} = \pm \frac{1}{f_{in} T_{GATE}} \quad (4)$$

for frequency measurements and

$$\frac{\Delta T}{T} = \pm T_{clk} / T_{in} \quad (5)$$

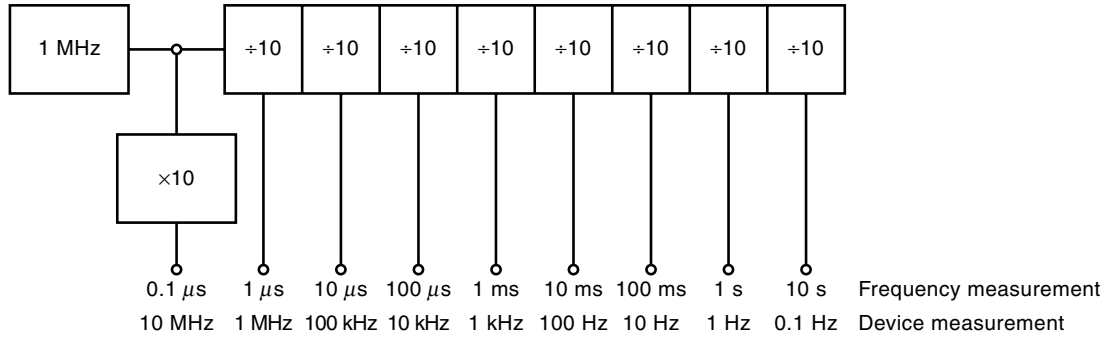


Figure 14. Time base divider unit.

for period measurements, where f_{in} and T_{in} are frequency and period of the signal under test, T_{clk} is the time base period, and T_{GATE} is the length of the counter gating pulse.

If $f_{clk} = 10$ MHz and $f_{in} = 10$ kHz, then the frequency error equals $\pm 10^{-7}$ and $\Delta T/T = \pm 10^{-3}$ for 1 s gating pulse.

HIGH-FREQUENCY METERS

Direct measurement of frequency is now available to the frequencies close to 1 GHz. Higher-frequency measurements above 1 GHz are achieved by means of frequency down-conversion using prescaling with frequency dividers (Fig. 15), transfer oscillators (Fig. 16), and frequency translation by mixing (Figs. 17 and 18). The frequency prescaling involves a process of frequency division in the fast digital counters. Then a classical low-frequency counter measures the input signal frequency. Typical division ratios are 2 to 16. The maximum frequency divided in a stable manner can reach several gigahertz (HP 5386A). Frequencies of signals reaching 25 GHz can be measured with help of transfer oscillators. The down-conversion, which is based on the heterodyne conversion, is applied to provide measurements within the frequency ranges comparable with the previous method, although the harmonic conversion extends the frequency to millimeter wave region.

In the following sections, the transfer oscillator and frequency translation methods are discussed.

TRANSFER OSCILLATOR

The transfer oscillator provides the down-conversion to extend the frequency measurement range by means of phase locking a

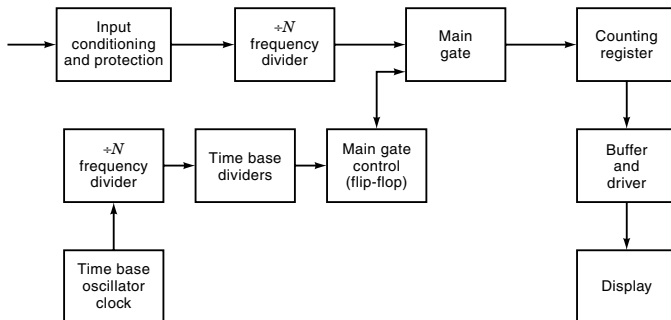


Figure 15. High-frequency counter with down conversion.

low-frequency oscillator to the high-frequency repetitive input wave. The high-frequency signal of the frequency f_x is sampled every n th period at a rate adjusted by the low-frequency oscillator of the frequency f_1 , as in coherent sampling scopes. The low frequency and the number n are computed to find the high frequency (3).

Figure 16 shows the block diagram and basic waves illustrating operation of a transfer oscillator. The upper part includes the phase-locked-loop system composed of a voltage-controlled oscillator, VCO_u , producing narrow sampling pulses; sampling gate; video amplifier, acting as a low-pass filter; phase detector, comparing phases of a reference oscillator signal with a signal filtered by the video amplifier. The lower part has the same sampling gate and video amplifier as the upper channel. Additionally, the channel includes a frequency mixer with a filter, and a low-frequency counter measuring frequency and frequency ratio.

When the loop is in lock, the input signal of frequency f_x is sampled by the narrow pulses at frequency f_1 established by the voltage-controlled oscillator VCO_u . The frequency f_1 is much lower than f_x , and is adjusted automatically in the loop. The output signal of the sampler is composed of the train of samples whose amplitudes vary like the input signal amplitude but at a much lower rate (f_{u1}). This signal, filtered by the video amplifier, has the frequency close to the reference frequency oscillator f_{ur} . The signal frequencies in the lock-in conditions follow the equation

$$f_x = nf_1 - f_{ur} = nf_1 - f_{ur} \quad (6)$$

where f_1 is the sampling pulse frequency, f_x is the unknown frequency, f_{ur} is the reference frequency, which in lock equals f_{uf} , n is an integer number. Typically, $f_{ur} \ll nf_1$, so nf_1 is close to f_x . The upper channel sends the signal of frequency f_1 to the low-frequency counter shown at the bottom of Fig. 16. The counter also determines the frequency transfer coefficient, n , and finally the product of n and f_1 . The lower channel involves the same sampling processes as the upper channel but the sampling frequency is offset to $f_1 + f_0$, where f_0 is a constant frequency much lower than f_1 . By offsetting the frequency f_1 , the unknown integer number n is transferred to the frequency f_0 . The rate of the signal variations after sampling and filtering by the lower channel video amplifier is determined by the frequency f_{lf} . The frequency relations for the lower channel are as follows:

$$f_{lf} = n(f_1 + f_0) - f_x \quad (7)$$

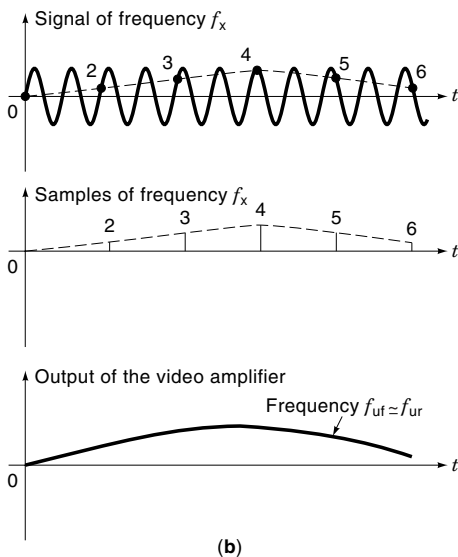
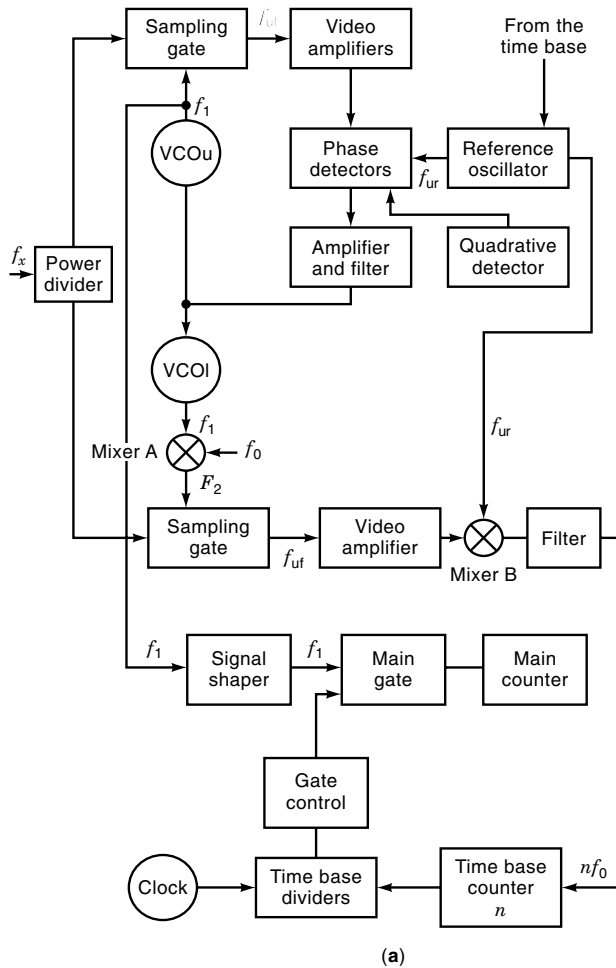


Figure 16. Transfer oscillator. (a) Block diagram. (b) Time waves in the upper part of the circuit.

since

$$f_2 = f_1 + f_0 \quad (8)$$

hence

$$f_{1f} = nf_0 + f_{uf} = nf_0 + f_{ur} \quad (9)$$

The signal of frequency f_{1f} is mixed with the signal of frequency f_{ur} and later filtered to pass only the signal of frequency nf_0 . The low frequency counter measures the frequency ratio of nf_0 and f_0 to determine the number n , and this number determines the length of the time base to measure the frequency nf_1 . In this way all components are available to find the frequency f_x , which is expressed as

$$f_x = nf_1 - f_{ur} \quad (10)$$

In practice, $f_{ur} \ll nf_1$ and $f_x = nf_1$.

FREQUENCY TRANSLATION

Frequency translation or heterodyne down-conversion system is shown in Fig. 17. The conversion involves mixing the input signal of unknown frequency with a harmonic of the reference source. The harmonic number, k , is adjusted automatically by the processor, which registers the signal from the signal detector and tunes the YIG/PIN switch filter until harmonic frequency of f_1 from the comb generator (generator of many harmonic signals) produces a noticeable detector signal. The comb generator is controlled through the frequency by the clock reference signal of high stability. The multiplier output signal frequency is typically 100 MHz to 500 MHz. The flowchart of Fig. 18 illustrates the system operation.

A combination of both previously described methods is applied in harmonic heterodyne counters to extend the measurement range to millimeter and submillimeter waves.

Similar techniques are used in optical measurements, although all operations are performed on optical signals.

LOW-FREQUENCY METERS

A classical digital frequency counter can measure signal frequencies by determining periods. The frequencies are then computed from the periods. Another method to measure low frequencies is applied in the counters, which have frequency multipliers. Figure 19 shows the frequency multiplier with the phase-locked-loop system. The loop is in lock when the subharmonic frequency of the voltage-controlled oscillator is adjusted to satisfy the phase condition imposed by the phase detector. The phase difference between the detector input signals must be confined between 0° and 180° , with the center of 90° ; or for some detectors the range of 180° is centered at 0° in order to produce a stable dc output to control the voltage-controlled oscillator. When the frequencies are not correctly aligned, the loop is not in lock, and the output voltage of the low-pass filter oscillates. If both frequencies are closer together, the oscillating frequency falls within the low-pass filter bandwidth, and the loop "captures" the input signal to lock it. The voltage-controlled

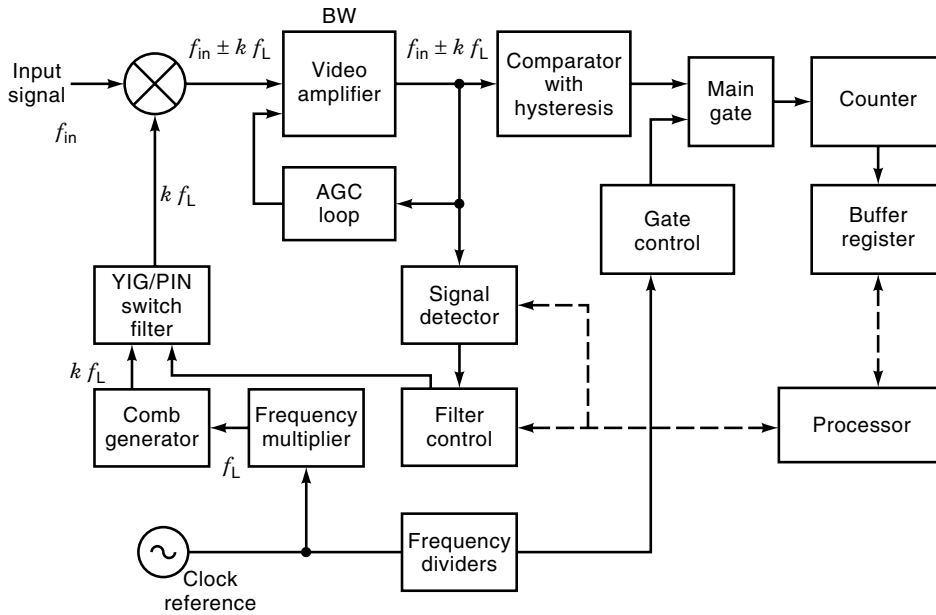


Figure 17. High-frequency counter with frequency translation (heterodyning).

oscillator is tuned automatically by the negative feedback loop of the system.

Another interesting method is applied in low-frequency measurements of rotating objects. This method is called a stroboscopic method. The low-frequency stroboscopic method involves high-intensity flashes of light, whose frequency can be controlled by a source of very stable but adjustable frequency signals. The light is directed on the mark on the rotating object whose frequency is to be measured. For the proper frequency of the flashlight, full synchronization of the light pulses

with a rotating object produces a single motionless image of the marker. The synchronization can be also achieved at the rotational speeds, which are multiples or submultiples of the actual speed of the rotating object. Figure 20 shows the basic measuring configuration with a stroboscopic device and the most basic patterns observed for various speeds. More complex synchronization patterns are described in Ref. 7.

The high-frequency stroboscopic method is applied in sampling oscilloscopes described in OSCILLOSCOPES. The sampling scopes were also called the stroboscopic oscilloscopes during very early stages of the development of this technique.

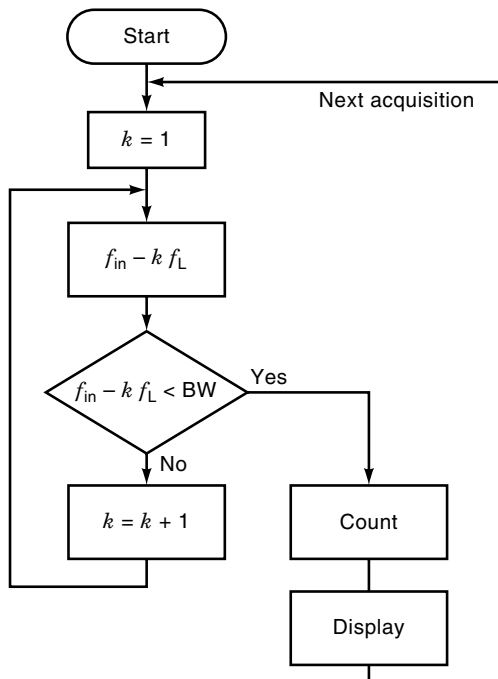


Figure 18. Flowchart illustrating operation of a high-frequency counter with frequency translation.

EXAMPLES OF CALIBRATION OF FREQUENCY AND TIME INTERVAL METERS

The processes of calibration are usually done by comparison of time or frequency of the internal time bases of the meters with available, more accurate standards whose parameters are traceable to national or international references. The standard

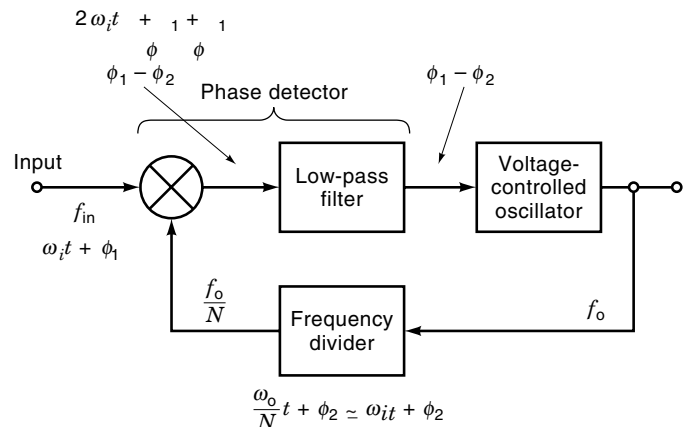


Figure 19. Frequency multiplier applied in low-frequency counters.

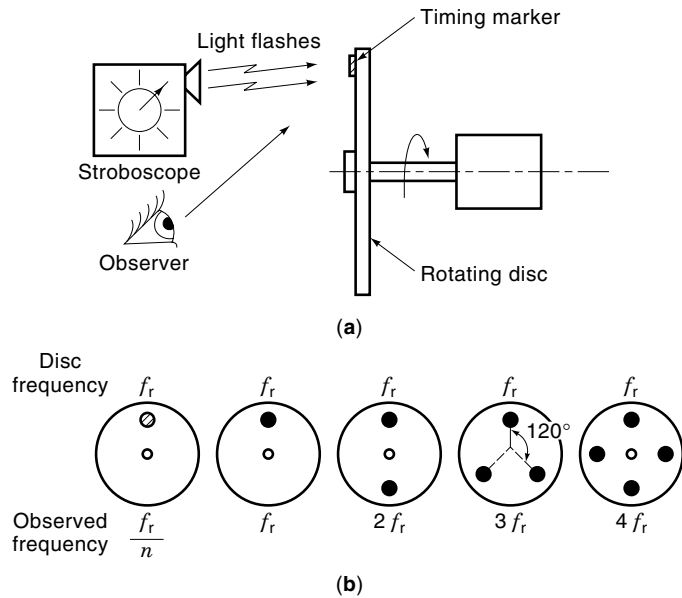


Figure 20. A stroboscope. (a) Measurement setup. (b) Stroboscopic patterns.

frequencies and time standards are accessible directly through radio, TV, and satellite signals, including GPS (1). The inaccuracy of atomic standards used to deliver reference signals reaches the level of 10^{-13} . Commercially available built-in standards (4) have their accuracy close to 10^{-10} . The external sources serve to either introduce automatic frequency corrections or establish the error bands. The frequency comparison can be done by means of the vector voltmeters, the linear phase comparators, oscilloscopes, and frequency counters measuring the differential frequency. Figures 21 through 23 show a few examples of the measurement setups in which the frequency of the counter time base are compared with the applied reference. The oscilloscope Lissajous patterns are described in OSCILLOSCOPES. Figure 21 illustrates another oscilloscope method that registers the frequency drift of an oscillator under test in relation to the oscillator of high stability. The rate of movement of the observed wave indicates the drift. If the wave displayed on the

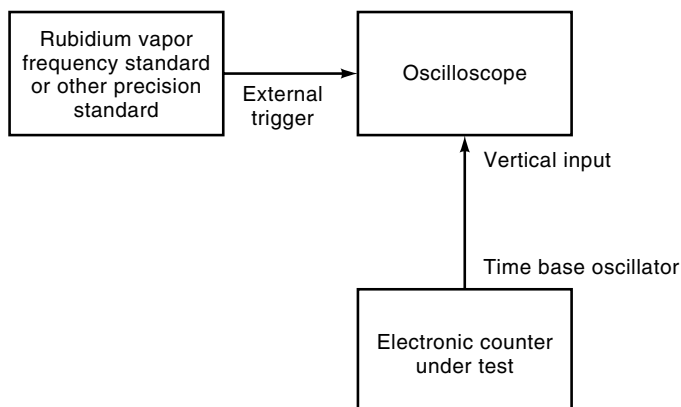


Figure 21. Oscilloscope tests of the frequency drift.

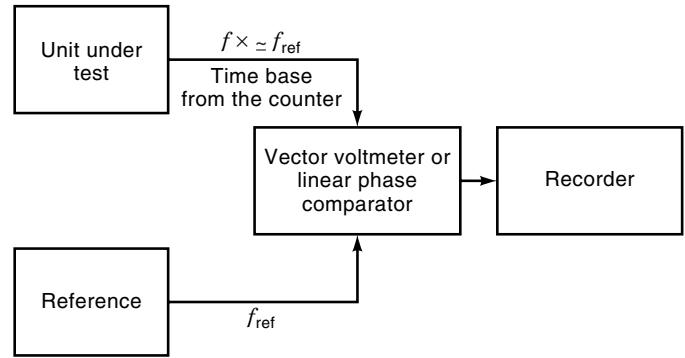


Figure 22. Frequency comparison testing circuit with a vector voltmeter or a phase comparator.

screen moves to the left, the tested frequency is higher than the frequency of the standard. If the wave moves right, the counter frequency is lower. Frequency error can be calculated from the speed of motion of the wave. If a 1 MHz wave is displayed using a 100 kHz reference signal for triggering and the observed drift is one full cycle per 10 s, then the frequency error 0.1 cycle per second related to 1 MHz gives 10^{-7} relative error due to the drift. The vector voltmeter method is illustrated in Fig. 22. The phases of the reference and tested signals are compared instead of frequencies. Frequency difference is calculated from the phase change $\Delta\Phi$ over a predetermined amount of time Δt according to the definition of the frequency:

$$\frac{\Delta f}{f} = \frac{\Delta\Phi}{2\pi\Delta t_f} \quad (11)$$

where. The factor 2 can be replaced with 360° when the phase is expressed in degrees.

In Fig. 23, the frequency counter is used to compare the frequency of the unit under test and the standard source frequency. The standard source clocks the frequency counter that measures the frequency counter under test.

Precision time and frequency transfer procedure at a remote site by means of the GPS system is described in (8). A site clock can be compared with universal time coordinated (UTC at NIST) anywhere in common view of a GPS satellite with NIST branch located in Boulder, Colorado.

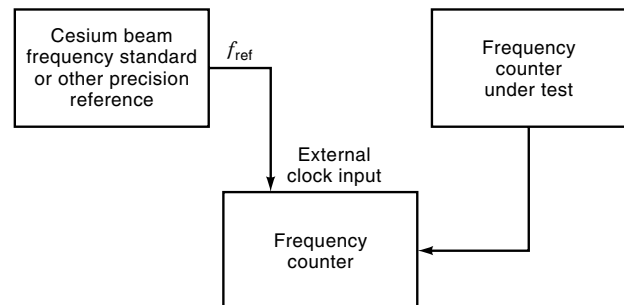


Figure 23. Frequency comparison with a counter.

BIBLIOGRAPHY

1. GPS and Precision Timing Applications, Application Note 1272, Palo Alto, CA: Hewlett Packard.
2. J. J. Carr, *Elements of Electronic Instrumentation and Measurement*, Upper Saddle River, NJ: Prentice-Hall, 1996.
3. Fundamentals of the Electronic Counters, Application Note 200, Palo Alto, CA: Hewlett Packard.
4. SR 625 2 GHz Rubidium Stabilized Frequency Counter [Online], Stanford Research Systems. Available www: <http://www.sr.com/sr625.html>
5. Timekeeping and Frequency Calibration, Application Note 52-2, Palo Alto, CA: Hewlett Packard.
6. Fundamentals of Time Interval Measurements, Application Note 200-3, Palo Alto, CA: Hewlett Packard.
7. R. S. Figolia and D. E. Besley, *Theory and Design for Mechanical Measurements*, New York: Wiley, 1995.
8. M. A. Weiss and D. W. Allen, An NBS calibration procedure for providing time and frequency at a remote site by weighting and smoothing of GPS common view data, *IEEE Trans. I&M*, **IM-36** (2), June 1987.
9. PCB and Chassis Mounted Oscillators, HCD Research Co., <http://www.hcdresearch.co.uk/oscillat/welcome.htm>

Reading List

- H. M. Berlin and F. C. Getz, Jr., *Principles of Electronic Instrumentation and Measurement*, Columbus, OH: Merrill Publishing Company, 1988.
- C. F. Coombs, Jr., editor-in-chief, *Basic Electronic Instrument Handbook*, New York: McGraw-Hill, 1972.
- W. D. Cooper and A. D. Helfrick, *Electronic Instrumentation and Measurement Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1985.
- Fundamentals of Microwave Frequency Counters, Application Note 200-1, Palo Alto, CA: Hewlett Packard.
- Fundamentals of Quartz Oscillators, Application Note 200-2, Palo Alto, CA: Hewlett Packard.
- Fundamentals of Time and Frequency Standards, Application Note 52-1, Palo Alto, CA: Hewlett Packard.
- HP 5371A Frequency and Time Interval Analyzer, Application Note 358-1, Palo Alto, CA: Hewlett Packard.
- HP 5386A Frequency Counter, Technical Data, Palo Alto, CA: Hewlett Packard.
- HP 5370B Universal Time Interval Counter, Application Note 191-7, Palo Alto, CA: Hewlett Packard.
- S. Wolf, *Guide to Electronic Measurements and Laboratory Practice*, Englewood Cliffs, NJ: Prentice-Hall, 1983.
- S. Wolf and R. F. M. Smith, *Student Reference Manual for Electronic Instrumentation Laboratories*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

ANDREW RUSEK
Oakland University

FREQUENCY STANDARDS, CHARACTERIZATION

In this article we describe the characterization of frequency standards following the general definitions accepted by the IEEE, the International Telecommunication Union-Radiocommunications (ITU-R), and International Radio

Consultative Committee (CCIR) (1–3). In using the term “frequency standard” we imply that changes in the frequency $\Delta\nu$ of the device are small compared to its nominal frequency ν_0 and that therefore the frequency would be about the same if we were to remeasure it. This permits us to treat the variations in fractional frequency $\Delta\nu/\nu_0$ as small compared to 1, and this greatly simplifies the mathematics of the characterization. We describe what is meant by accuracy and frequency stability of a frequency standard. The variations in frequency (which define the frequency stability) can be classified into two basic types, random and systematic. To characterize the random variations in frequency the systematic effects must be removed from the data. Special statistical techniques other than the standard variance must then be used to quantify the random variations in frequency, because some of the noise processes are not stationary. By this we mean that the mean of the frequency noise changes slowly with time. As part of the statistical treatment we describe how to determine the confidence intervals for the estimates of the various types of random frequency noise.

ACCURACY

Frequency standards generate a periodic voltage signal, whose ideal frequency is defined as a specific number of oscillations per second. The second is the agreed unit of time, based on the energy difference between two energy levels of the unperturbed cesium atom (4), and it is this definition that allows the specification of accuracy of a frequency standard. Generally there will be offsets or biases in the actual frequency of a standard when compared to the ideal or defined value (according to the definition of the second) due to systematic and random effects (5–7). The *accuracy* of a frequency standard, more recently described as an uncertainty, is a measure of the confidence relative to a physical model of the standard and the evaluation process. To evaluate the accuracy of a frequency standard, a list of known sources of frequency offsets is made, and the offset or bias due to each source and its uncertainty are carefully measured or computed. The uncertainty is a proper summation of the estimates of various systematic offsets and random noise (8). For example, a frequency standard with an output frequency of 10 MHz and an uncertainty of 10^{-8} has an frequency error of $\leq \pm 0.1$ Hz.

STABILITY

Figure 1 shows the output voltage signal of an ideal *frequency standard* as a function of time. The maximum value V_0 is the nominal amplitude of the signal. The time required for the signal to repeat itself is the period T of the signal. The nominal frequency ν_0 of the signal is the reciprocal of the period, $1/T$. This voltage signal can be represented mathematically by a sine function,

$$v(t) = V_0 \sin \theta = V_0 \sin(2\pi \nu_0 t) \quad (1)$$

where the argument $\theta = 2\pi \nu_0 t$ of the sine function is the nominal phase of the signal. The time derivative of the phase θ is $2\pi \nu_0$ and is called the nominal angular frequency ω_0 . In the frequency domain, this ideal signal is represented by a delta

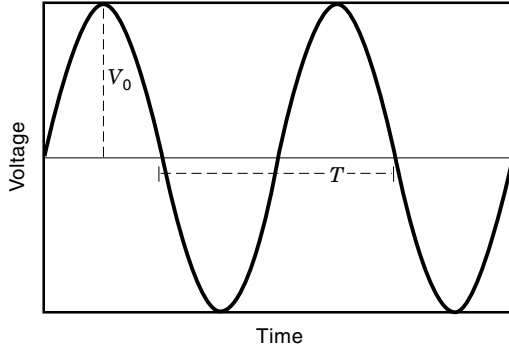


Figure 1. Ideal output voltage of a frequency standard.

function located at the frequency of oscillation. Since this signal is ideal, there are no known sources that cause frequency shifts, and thus its frequency is *totally* accurate and its uncertainty is 0. Furthermore, its frequency, phase, and amplitude are constant; therefore the signal is also stable in frequency, phase, and amplitude.

In real situations, the output signal of an oscillator (source) or frequency standard has noise. Such a *noisy signal* is illustrated in Fig. 2. In this example we have depicted a case where the noise power is much less than the signal power. Frequency instability is the result of fluctuations in the period of the signal. Amplitude instability is the result of fluctuations in the peak values of the voltage. Phase instability is the result of fluctuations in the zero crossings. Since the period (and thus the frequency) of the signal is related to its phase, frequency instability and phase instability are directly related.

Figure 3 shows the *power spectrum of a noisy signal* (power as a function of frequency) as measured by a spectrum analyzer. Although the maximum power occurs at the frequency of oscillation, other peaks are observed at frequencies of $2\nu_0$, $3\nu_0$, $\dots$, $n\nu_0$, where n is a positive integer. These frequencies are called *harmonics* of the fundamental frequency ν_0 ; $2\nu_0$ is the second harmonic, $3\nu_0$ is the third harmonic, and so on. The power at these harmonic frequencies will depend on the design of the source. The spectrum around the fundamental frequency displays power sidebands at frequencies above the carrier (upper sideband) and at frequencies below the carrier (lower sideband). These power sidebands are the result of phase fluctuations and amplitude fluctuations in the signal.

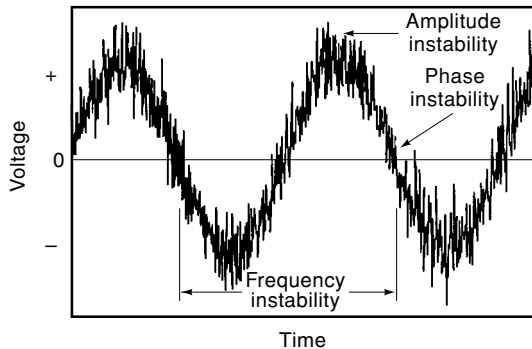


Figure 2. Output voltage of a noisy signal.

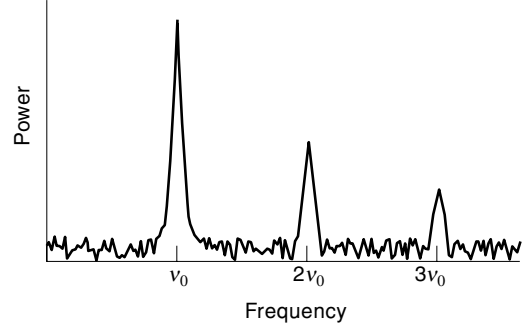


Figure 3. Power spectrum of a noisy signal.

While the power spectrum gives an idea of the total noise of a signal, it does not give information about the relative magnitude of its phase instabilities and amplitude instabilities. Furthermore, at frequencies close to ν_0 it is difficult to separate the noise power from the power of the fundamental frequency. Therefore, special measurement techniques are needed to measure phase instabilities and amplitude instabilities in sources.

A *noisy signal* can be mathematically represented by

$$v(t) = [V_0 + \epsilon(t)] \sin[2\pi\nu_0 t + \phi(t)] \quad (2)$$

where $\epsilon(t)$ represents amplitude fluctuations (amplitude deviation from the nominal amplitude V_0) and $\phi(t)$ represents phase fluctuations (phase deviation from the nominal phase $2\pi\nu_0 t$). (1). The instantaneous frequency of this signal is defined as

$$\nu(t) = \frac{1}{2\pi} \frac{d}{dt} (\text{phase}) = \nu_0 + \frac{1}{2\pi} \frac{d}{dt} \phi(t) \quad (3)$$

Frequency fluctuations refer to the deviation $\nu(t) - \nu_0$ of the instantaneous frequency from the nominal frequency. Fractional frequency fluctuations, denoted as $y(t)$, refer to frequency fluctuations normalized to ν_0 ; that is,

$$y(t) = \frac{\nu(t) - \nu_0}{\nu_0} = \frac{1}{2\pi\nu_0} \frac{d}{dt} \phi(t) \quad (4)$$

Equation (4) indicates that there is a direct relation between phase fluctuations and fractional frequency fluctuations. Therefore, if a signal exhibits a certain amount of phase fluctuation, it also exhibits frequency fluctuation given by Eq. (4). The *time deviation* $x(t)$ of a signal is the integral of $y(t)$ from 0 to t . Thus one can write

$$y(t) = \frac{d}{dt} x(t) \quad (5)$$

The frequency stability of a frequency standard is affected by random noise processes and by systematic, deterministic changes. Systematic effects often dominate the frequency stability in nonlaboratory environments. The sensitivity of the standard to temperature, humidity, atmospheric pressure, magnetic field, and radiation may play a role (9–11). Generally, frequency (and thus phase) stability is divided into three regions: short-term, medium-term, and

long-term stability. (Though the term frequency stability is used throughout most of the literature, the term actually refers to instabilities in the frequency.)

Short-term frequency stability refers to the random, nonsystematic fluctuations that are related to the signal-to-noise ratio of the device. In quartz crystal resonators this refers to the region dominated by white phase noise, where the time between observed frequencies (sample time) is less than a second. In atomic frequency standards the short-term stability also includes white frequency noise and extends to sampling times of several minutes.

Medium-term frequency stability refers to the region where flicker noise dominates. The sampling time characteristic of this region is a function of the type of frequency standard. Short-term and medium-term frequency stability can be characterized either in the frequency domain or in the time domain, after known systematic effects have been removed. Frequency-domain characterization and measurements are generally used when the sample time of interest is less than a second. For sampling times longer than a second, time-domain measurements are used to characterize frequency stability.

Long-term frequency stability includes random-walk frequency-noise processes in addition to systematic, deterministic changes in frequency observed when the sampling time is long. The long-term, systematic frequency change is called *frequency drift* (2,12). Drift includes frequency changes due to changes in the components of the source in addition to those due to external parameters such as temperature, humidity, pressure, magnetic field, and radiation (9–11). *Frequency aging*, on the other hand, refers to the long-term systematic frequency change due to changes in the components of the source, independent of parameters external to the source (2,12).

Frequency-Domain Characterization

Phase fluctuations in the frequency domain, or *phase modulation* (PM) noise, are characterized by the power spectral density (PSD) of the phase fluctuations, given by (13)

$$S_\phi(f) = \mathcal{F}[\mathcal{R}_\phi(\tau)] = \int_{-\infty}^{\infty} \mathcal{R}_\phi(\tau) e^{-j\omega\tau} d\tau \quad (6)$$

where $\mathcal{F}[\]$ is the Fourier transformation and $\mathcal{R}_\phi(\tau)$ is the autocorrelation function of the phase fluctuations, given by

$$\mathcal{R}_\phi(\tau) = \langle \phi(\tau)\phi(t-\tau) \rangle \quad (7)$$

A more practical definition of $\text{PSD}[\phi(t)]$ is

$$S_\phi(f) = \text{PSD}[\phi(t)] = [\phi(f)]^2 \frac{1}{\text{BW}} \quad (8)$$

where $[\phi(f)]^2$ is the mean squared phase deviation at an offset frequency f from the frequency ν_0 (called the carrier in this context), and BW is the bandwidth of the measurement system (1,13–15). The offset frequency f is also called the Fourier frequency. The units for $S_\phi(f)$ are rad^2/Hz . Equation (8) is defined for $0 < f < \infty$; nevertheless it includes fluctuations from the upper and lower sidebands and thus is a double-sideband measure.

The PM noise measure recommended by the IEEE (1,14,15) is $\mathcal{L}(f)$, defined as

$$\mathcal{L}(f) \equiv \frac{S_\phi(f)}{2} \quad (9)$$

At Fourier frequencies far from the carrier frequency, where the integrated PM noise from ∞ to f (the Fourier frequency) is less than 0.1 rad^2 , $\mathcal{L}(f)$ is equal to the single-sideband phase noise. The units for $\mathcal{L}(f)$ are decibels below the carrier in a 1 Hz bandwidth (dBc/Hz).

Frequency fluctuations in the frequency domain, or *frequency modulation* (FM) noise, are characterized by the power spectral density of the fractional frequency fluctuations, given by

$$S_y(f) = \text{PSD}[y(t)] = [y(f)]^2 \frac{1}{\text{BW}} \quad (10)$$

where $y(f)^2$ represents the mean squared fractional frequency deviation at an offset (Fourier) frequency f from the carrier (1,13–15). $S_y(f)$ is defined for Fourier frequencies $0 < f < \infty$, and its units are inverse hertz.

The conversion between $S_y(f)$ and $S_\phi(f)$ can be obtained from Eq. (4). Applying the Fourier transformation to both sides of Eq. (4), squaring, and dividing by the measurement bandwidth results in

$$S_y(f) = \left(\frac{1}{2\pi\nu_0} \right)^2 (2\pi f)^2 S_\phi(f) = \left(\frac{f}{\nu_0} \right)^2 S_\phi(f) \quad (11)$$

Amplitude fluctuations in the frequency domain, or *amplitude modulation* (AM) noise, are characterized by the power spectral density of the fractional amplitude fluctuations, given by

$$S_a(f) = \text{PSD} \left[\frac{\epsilon(t)}{V_0} \right] = \left(\frac{\epsilon(f)}{V_0} \right)^2 \frac{1}{\text{BW}} \quad (12)$$

where $\epsilon(f)^2$ represents the mean squared amplitude deviation at an offset frequency f from the carrier (1). $S_a(f)$ is defined for Fourier frequencies $0 < f < \infty$, and its units are inverse hertz.

In free-running sources, the FM noise is usually modeled by the sum of five different power laws or noise types as

$$S_y(f) = \sum_{\gamma=-2}^2 h_\gamma f^\gamma = h_{-2} f^{-2} + h_{-1} f^{-1} + h_0 f^0 + h_1 f^1 + h_2 f^2 \quad (13)$$

where $h_{-2} f^{-2}$ represents random-walk frequency noise, $h_{-1} f^{-1}$ represents flicker frequency noise, $h_0 f^0$ represents white frequency noise, $h_1 f^1$ represents flicker phase noise, and $h_2 f^2$ represents white phase noise. Similarly, the PM noise can be modeled by

$$S_\phi(f) = \sum_{\beta=-4}^0 k_\beta f^\beta = k_{-4} f^{-4} + k_{-3} f^{-3} + k_{-2} f^{-2} + k_{-1} f^{-1} + k_0 f^0 \quad (14)$$

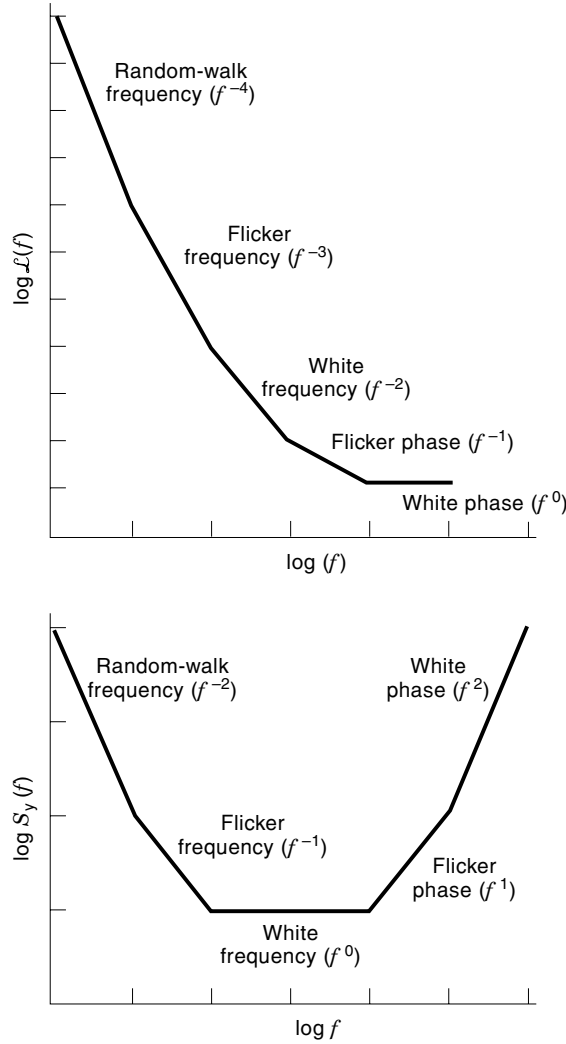


Figure 4. PM and FM noise characteristics of a source.

where $k_{-4}f^{-4}$ represents the random-walk frequency noise, $k_{-3}f^{-3}$ represents flicker frequency noise, $k_{-2}f^{-2}$ represents white frequency noise, $k_{-1}f^{-1}$ represents flicker phase noise, and k_0f^0 represents white phase noise. Notice that $S_\phi(f)$ and $S_y(f)$ have different slopes for a specific type of noise, as implied by Eq. (11). Equation (11) can be used to obtain the conversion between the $S_\phi(f)$ and $S_y(f)$ coefficients, yielding

$$k_\beta = v_o^2 h_\gamma \quad \text{for} \quad \beta = \gamma - 2 \quad (15)$$

Figure 4 shows the common noise types characteristic of the PM noise and the FM noise of a source (1,14–17). Usually a source exhibits two or three of the noise types shown in the plots (17).

The AM noise of a source can typically be modeled by the sum of three different power laws or noise types:

$$S_a(f) = \sum_{\gamma=-2}^0 h_\gamma f^\gamma = h_{-2}f^{-2} + h_{-1}f^{-1} + h_0f^0 \quad (16)$$

where $h_{-2}f^{-2}$ represents random-walk amplitude noise, $h_{-1}f^{-1}$ represents flicker noise, and h_0f^0 represents white amplitude

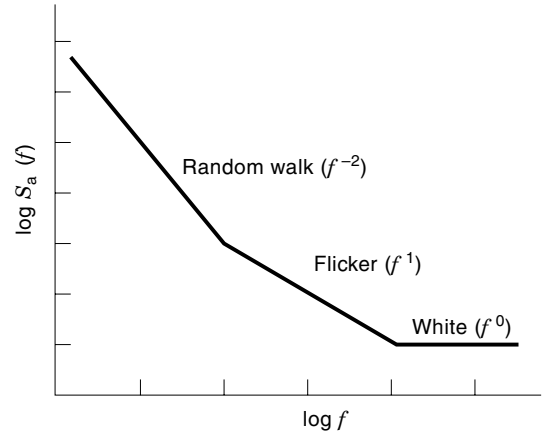


Figure 5. AM noise characteristics of a source.

noise (18). Figure 5 shows the common noise types characteristic of the AM noise of a source.

Upper and lower PM sidebands are always equal and 100% correlated. Likewise the upper and lower AM sidebands are always equal and 100% correlated. This is true even when the RF spectrum is not symmetric about the carrier (18a). The phase between AM and PM noise varies randomly with time for broadband additive noise (18a).

Time-Domain Characterization

In the time domain, the *fractional frequency stability* of a signal is usually characterized by the Allan variance, a type of two-sample frequency variance given by (1,13–15)

$$\sigma_y^2(\tau) = \frac{1}{2} (\bar{y}_{i+1} - \bar{y}_i) \quad (17)$$

where $\bar{y}_i$ is the average fractional frequency of interval i given by

$$\bar{y}_i = \frac{1}{\tau} \int_{t_i}^{t_i+\tau} y(t) dt = \frac{1}{\tau} [x(t_i + \tau) - x(t_i)] = \frac{1}{\tau} (x_{i+1} - x_i) \quad (18)$$

In practical situations only a finite number of fractional frequency samples are available, and the Allan variance is approximated by

$$\sigma_y^2(\tau) \approx \frac{1}{2(M-1)} \sum_{i=1}^{M-1} (\bar{y}_{i+1} - \bar{y}_i)^2 \quad (19)$$

where $M = N - 1$ is the number of frequency samples (N is the number of time samples) (1,14,15). The Allan variance can also be expressed in terms of time samples using $\bar{y}_i = (x_{i+1} - x_i)/\tau$:

$$\sigma_y^2(\tau) \approx \frac{1}{2(N-2)\tau^2} \sum_{i=1}^{N-2} (x_{i+2} - 2x_{i+1} + x_i)^2 \quad (20)$$

The square root of the Allan variance is called the Allan deviation, $\sigma_y(\tau)$.

When time samples are taken every τ_0 seconds, the Allan variance can be computed for several sampling times $\tau = n\tau_0$ where $n > 0$. For $n > 1$, overlapped samples can be used to compute σ_y^2 as shown in Fig. 6, providing better confidence

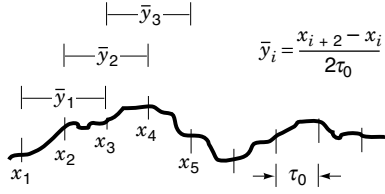


Figure 6. Computation of $\{\bar{y}_i\}$ for the overlapped Allan variance and $n = 2$.

intervals (1,19). An expression for the fully overlapped Allan variance can be derived using Fig. 6 and Eq. (19). For $\tau = n\tau_0$ Eq. (19) becomes

$$\sigma_y^2(\tau) \approx \frac{1}{2(M-2n)} \sum_{i=1}^{M-2n} (\bar{y}_{i+n} - \bar{y}_i)^2 \quad (21)$$

which can also be expressed in terms of time-domain data by substituting $\bar{y}_i$ for $(x_{i+n} - x_i)/n\tau_0$:

$$\sigma_y^2(\tau) \approx \frac{1}{2(N-2n)\tau^2} \sum_{i=1}^{N-2n} (x_{i+2n} - 2x_{i+n} + x_i)^2 \quad (22)$$

Figure 7(a) shows a log-log plot of the Allan variance as a function of the sampling time τ for a source that exhibits all five common noise types. The slopes of the white PM noise and the flicker PM noise are the same; therefore these two noise types cannot be separated using this plot. The Allan variance can often be modeled by the sum of four different power laws:

$$\sigma_y^2(\tau) = \sum_{\mu=-2}^1 p_\mu \tau^\mu = p_{-2}\tau^{-2} + p_{-1}\tau^{-1} + p_0\tau^0 + p_1\tau^1 \quad (23)$$

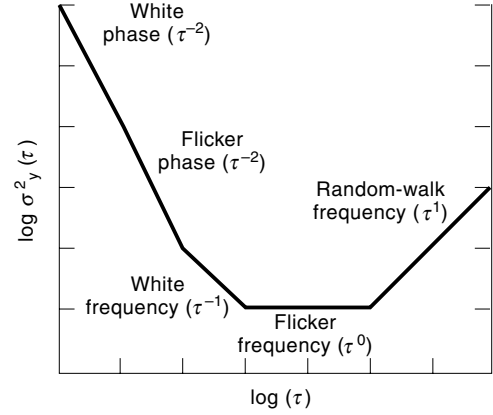
where $p_{-2}\tau^{-2}$ represents white phase and flicker phase noise, $p_{-1}\tau^{-1}$ represents white frequency noise, $p_0\tau^0$ represents flicker frequency noise, and $p_1\tau^1$ represents random walk frequency noise.

When the dominant noise type in the short term is flicker PM or white PM, the modified Allan variance can be used to improve the estimate of the underlying frequency stability of the sources (1,14,20). Here a new series $\{\bar{x}_i\}$ is created by averaging n adjacent phase (time) measurements of duration τ_0 . The average fractional frequencies are computed from the $\{\bar{x}_i\}$, as illustrated in Fig. 8. For N time samples and $\tau = n\tau_0$, the resulting modified Allan variance is

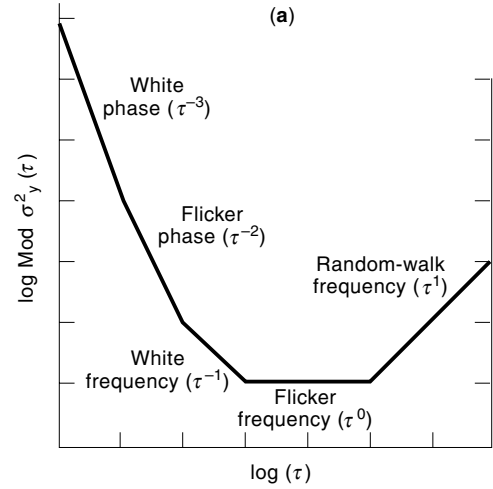
$$\text{Mod } \sigma_y^2(\tau) \approx \frac{1}{2(N-3n+1)} \sum_{i=1}^{N-3n+1} (\bar{y}'_{i+n} - \bar{y}'_i)^2 \quad (24)$$

Equation (24) can also be expressed in terms of the initial time-domain data $\{x_k\}$:

$$\begin{aligned} \text{Mod } \sigma_y^2(\tau) &\approx \frac{1}{2\tau^2 n^2 (N-3n+1)} \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2 \\ &\quad (25) \end{aligned}$$



(a)



(b)

Figure 7. $\sigma_y^2(\tau)$ and $\text{Mod } \sigma_y^2(\tau)$ for the five noise types.

where

$$\bar{y}'_i = \frac{\bar{x}_{i+n} - \bar{x}_i}{\tau} \quad (26)$$

and

$$\bar{x}_i = \frac{\sum_{k=0}^{n-1} x_{i+k}}{n} \quad (27)$$

$$\bar{x}_i = x_{i+2} + x_{i+1} + x_i$$

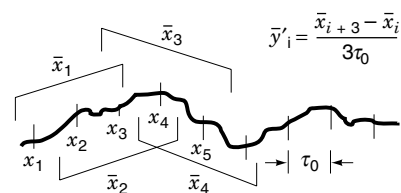


Figure 8. Computation of $\{\bar{x}_i\}$ and $\{\bar{y}'_i\}$ for the Modified Allan variance and $n = 3$.

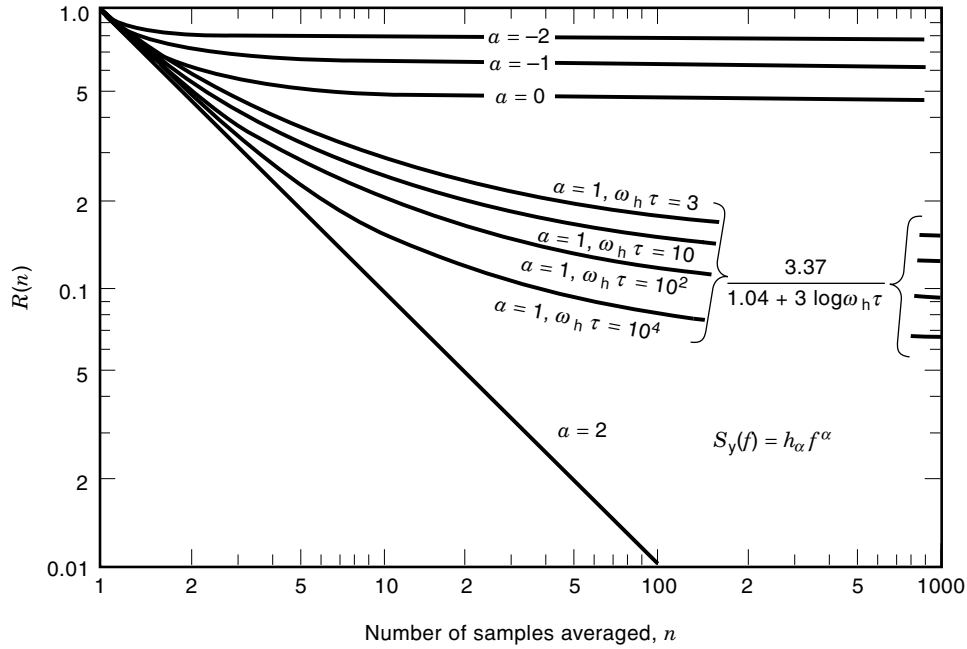


Figure 9. Ratio of the modified Allan variance to the Allan variance, $R(n) = [\text{Mod } \sigma_y^2(\tau)]/\sigma_y^2(\tau)$, as a function of n (15). $\tau = n\tau_0$.

Here $\bar{x}_i$ is the phase (time) averaged over n adjacent measurements of duration τ_0 . Thus $\text{Mod } \sigma_y(\tau)$ is proportional to the second difference of the phase averaged over a time $n\tau_0$. Viewed from the frequency domain, $\text{Mod } \sigma_y(\tau)$ is proportional to the first difference of the frequency averaged over n adjacent samples. The square root of the modified Allan variance is called the modified Allan deviation $\text{Mod } \sigma_y(\tau)$.

Figure 9 shows the ratio $R(n) = [\text{Mod } \sigma_y^2(\tau)]/\sigma_y^2(\tau)$ as a function of n for all five types of noise processes (15). For random-walk FM, flicker FM, and white FM the ratio is constant for $n \geq 5$. Therefore, $\text{Mod } \sigma_y^2(\tau)$ and $\sigma_y^2(\tau)$ have the same slope for these noise types. For white PM noise ($\alpha = 2$), the slope is $-1/n$; therefore the slope of $\text{Mod } \sigma_y^2(\tau)$ is equal to the slope of $\sigma_y^2(\tau)$ divided by τ . Finally, for flicker PM noise the ratio asymptotically reaches $3.37/(1.04 + 3 \ln \omega_h \tau)$. In this case and for $n > 10$, the slope of $\text{Mod } \sigma_y^2(\tau)$ is approximately τ^{-1} . Figure 9 also shows that $\text{Mod } \sigma_y^2(\tau)$ is considerably smaller than $\sigma_y^2(\tau)$ for white PM and flicker PM noise. Not only does $\text{Mod } \sigma_y^2(\tau)$ provide a different slope for white PM noise and flicker PM noise, allowing the separation of the two noise processes (see Fig. 7b); it can also speed the stability measurements. If a system is limited by white and flicker PM noise at short average times, using $\text{Mod } \sigma_y^2(\tau)$ reduces the measurement time required to observe white FM, flicker FM, and random-walk FM at longer averaging times, in comparison with that required when using $\sigma_y^2(\tau)$ (15).

At long averaging times when the ratio $N\tau_0/(2n\tau_0)$ is close to 1, the Allan variance has a bias related to its insensitivity to odd noise processes in the phase (time) fluctuations (odd with respect to the midpoint). In these situations an extension of the Allan variance that removes this bias can be used to characterize the frequency stability of a source. This variance, $\sigma_{y,\text{TOTAL}}^2(\tau)$, is obtained by extending the $\{x_i\}$ in both directions and then computing the Allan variance from the new $\{x'_i\}$

sequence (21–23). Figure 10 illustrates this extension of $\{x_i\}$: on the left side the extension is the inverted mirror image of $\{x_i\}$ with respect to x_1 ; on the right side it is the inverted mirror image of $\{x_i\}$ with respect to x_N . How far this extension depends on the maximum value n_m of n . For N time data points, n_m is the integer part of $(N-1)/2$. The far-left data point is $x'_{2-n_m} = 2x_1 - x_{n_m}$; the far-right data point is $x'_{N+n_m} - 1 = 2x_N - x_{N-n_m} + 1$. Thus $\sigma_{y,\text{TOTAL}}^2(\tau)$ is given by

$$\hat{\sigma}_{y,\text{TOTAL}}^2(\tau) = \frac{1}{2(N-2)} \sum_{i=2}^{N-1} (\bar{y}_i - \bar{y}_{i-n})^2 \quad (28)$$

Equation (28) can be expressed in terms of $\{x'_i\}$ using $\bar{y}'_i = (x'_{i+n} - x'_i)/\tau$ as

$$\hat{\sigma}_{y,\text{TOTAL}}^2(\tau) = \frac{1}{2\tau^2(N-2)} \sum_{i=2}^{N-1} (x'_{i-n} - 2x'_i + x'_{i+n})^2 \quad (29)$$

For a detailed description of $\sigma_{y,\text{TOTAL}}^2(\tau)$ see Refs. 21–23.

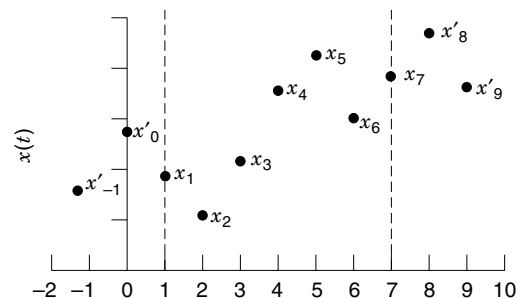


Figure 10. Extension of $\{x_i\}$ for $\sigma_y^2(\tau)$.

CONVERSION BETWEEN TIME-DOMAIN MEASURES AND FREQUENCY-DOMAIN MEASURES

Frequency-domain data $S_\phi(f)$ can be converted to time-domain data $\sigma_y^2(\tau)$ using the relation (1,24)

$$\sigma_y^2(\tau) = \frac{2}{(\pi \nu_0 \tau)^2} \int_0^{f_h} S_\phi(f) \sin^4(\pi f \tau) df \quad (30)$$

Equation (30) is derived by expressing both $S_\phi(f)$ and $\sigma_y^2(\tau)$ in terms of the autocorrelation function $\mathcal{R}_\phi(\tau)$ of the random process $\phi(t)$ and then combining the two expressions to cancel $\mathcal{R}_\phi(\tau)$ (13). Similarly,

$$\sigma_y^2(\tau) = \frac{2}{(\pi \tau)^2} \int_0^{f_h} S_y(f) \frac{\sin^4(\pi f \tau)}{f^2} df \quad (31)$$

Expressions for $\text{Mod } \sigma_y^2(\tau)$ (15,24), obtained using a similar procedure, are

$$\text{Mod } \sigma_y^2(\tau) = \frac{2}{n^4(\pi \nu_0 \tau_0)^2} \int_0^{f_h} \frac{S_\phi(f) \sin^6(\pi \tau f)}{\sin^2(\pi \tau_0 f)} df \quad (32)$$

$$\text{Mod } \sigma_y^2(\tau) = \frac{2}{n^4(\pi \tau_0)^2} \int_0^{f_h} \frac{S_y(f) \sin^6(\pi \tau f)}{f^2 \sin^2(\pi \tau_0 f)} df \quad (33)$$

The inclusion of f_h as the upper limit of the integral assumes that the term inside the integral is multiplied or “filtered” by an infinitely sharp low-pass filter with cutoff frequency f_h . Table 1 shows the results of Eqs. (31) and (33) for the five types of noise for $2\pi f_h \tau \gg 1$ (1,14–16,24). The results will depend on the type of filter assumed. While an infinitely sharp filter was assumed in Eqs. (30)–(33), actual measurement systems have different filter response.

Expressions have also been derived for a single-pole filter (15,25,26). These expressions, along with those in Table 1, constitute the boundaries for $\sigma_y^2(\tau)$ and $\text{Mod } \sigma_y^2(\tau)$, given a specific PSD of phase fluctuations (15). For this reason it is important to specify the filter frequency response, including the high cutoff frequency, when specifying the Allan and modified Allan variances of a source.

Generally, conversion from $\sigma_y^2(\tau)$ or $\text{Mod } \sigma_y^2(\tau)$ to the frequency domain is not possible, unless specific information about the noise characteristics is known. Greenhall demonstrated that several different spectral densities of random processes can have the same Allan variance (27). However, in the case where the spectral density follows the noise model in Eqs. (13)–

(14), a one-to-one correspondence between $S_\phi(f)$ and $\sigma_y^2(\tau)$ and $\text{Mod } \sigma_y^2(\tau)$ is found, except that for white PM and flicker PM noise $\sigma_y^2(\tau)$ exhibits the same slope, corresponding to τ^{-2} . Often, uniqueness fails more generally. Some sources have internal phase-locked loops, and their noise spectra deviates from the model in Eqs. (13)–(14) (28); others exhibit 60 Hz and other peaks that will affect $\sigma_y^2(\tau)$ (16,29). Generally, multivariate analysis should be used to obtain frequency-domain coefficients for each type of noise from time-domain data (30).

CONFIDENCE INTERVALS FOR $\sigma_y^2(\tau)$ AND $\text{Mod } \sigma_y^2(\tau)$

The Allan variance is defined as the first difference of average fractional frequencies, averaged over an infinite time. Since only a finite number M of frequency samples can be taken, we can only estimate the Allan variance and deviation, and the confidence of this estimate depends on the number of samples.

A simple method to obtain confidence intervals is to use the chi-squared distribution function. The Allan variance has a chi-squared distribution function given by

$$\chi^2 = \text{df} \frac{\hat{\sigma}_y^2(\tau)}{\sigma_y^2(\tau)} \quad (34)$$

where df is the number of degrees of freedom (16). The Allan variance is the sum of the squares of the first differences of adjacent fractional frequency values. If all the first-difference values were independent, then the number of degrees of freedom would be equal to the number of first difference values. This is not the case, and thus other procedures have been used to compute the number of degrees of freedom for $\sigma_y^2(\tau)$ (16). Table 2 shows analytical (empirical) equations that approximate the number of degrees of freedom for the fully overlapped Allan variance (16). The equation depends on the noise type. For nonoverlapped estimates, n in Table 2 is equal to 1, and N refers to the *equivalent* number of time samples for $\tau = n\tau_0$ given by $\text{Int}((N-1)/n) + 1$, where $\text{Int}(\)$ refers to the integer part.

Usually a $(p \times 100)\%$ confidence interval is computed, where p is the probability that the true Allan variance or Allan deviation is within the computed confidence interval. The $(p \times 100)\%$ confidence interval for the overlapped Allan

Table 1. Conversion Factors for $\sigma_y^2(f)$ and $\text{Mod } \sigma_y^2(f)$ (1, 14–16, 24)

Noise Type	$S_y(f)$	$\sigma_y^2(f)$	$\text{Mod } \sigma_y^2(f)$
Random walk frequency	$h_{-2}f^{-2}$	$(2\pi^2\tau/3)h_{-2}$	$5.42h_{-2}\tau$
Flicker frequency	$h_{-1}f^{-1}$	$(2 \ln 2)h_{-1}$	$0.936h_{-1}$
White frequency	h_0f^0	$h_0/2\tau$	$h_0/4\tau$
Flicker phase	h_1f^1	$\frac{1.038 + 3 \ln(\omega_0\tau)}{4\pi^2} h_1 \frac{1}{\tau^2}$	$\frac{3.37}{4\pi^2} h_1 \frac{1}{\tau^2}$
White phase	h_2f^2	$\frac{3f_h}{4\pi^2} h_2 \frac{1}{\tau^2}$	$\frac{3f_h}{4\pi^2} h_2 \frac{1}{n\tau^2}$

Table 2. Empirical Equations for the Number of Degrees of Freedom When Computing Confidence Intervals for the Overlapped Allan Variance (16)

Noise Type	α	No. of Degrees of Freedom
White phase	-2	$\frac{(N+1)(N-2n)}{2(N-n)}$
Flicker phase	-1	$\exp \left[\ln \left(\frac{N-1}{2n} \right) \ln \left(\frac{(2n+1)(N-1)}{4} \right) \right]^{1/2}$
White frequency	0	$\left(\frac{3(N-1)}{2n} - \frac{2(N-2)}{N} \right) \frac{4n^2}{4n^2+5}$
Flicker frequency	+1	$\frac{2(N-2)^2}{2.3N-4.9}$ for $n=1$ $\frac{5N^2}{4n(N+3n)}$ for $n \geq 2$
Random-walk frequency	+2	$\frac{N-2}{n} \frac{(N-1)^2 - 3n(N-1) + 4n^2}{(N-3)^2}$

Table 3. Confidence Intervals for the Nonoverlapped and Fully Overlapped $\sigma_y(\tau)$ (16) and for the Fully Overlapped Mod $\sigma_y(\tau)$ (31–33)^a

n	Noise Type	Confidence Interval (%)					
		Nonoverlapped $\sigma_y(\tau)$		Fullyoverlapped $\sigma_y(\tau)$		Fullyoverlapped Mod $\sigma_y(\tau)$	
		–	+	–	+	–	+
2	White PM	4.1	4.8	2.9	3.2	3.1	3.4
8		7.7	10.1	2.9	3.2	5.2	6.1
32		13.6	23.1	3.0	3.4	9.7	14
2	Flicker PM	3.7	4.3	2.9	3.1	3.0	3.3
8		7.1	9.0	3.6	4.0	5.7	6.8
32		12.7	20.7	5.2	6.1	11	16
2	White FM	3.6	4.0	2.8	3.0	3.0	3.2
8		6.8	8.6	4.8	5.6	5.8	7.0
32		12.5	20.1	8.8	12	11	16
2	Flicker FM	3.2	3.5	2.6	3.0	2.9	3.2
8		6.1	7.4	5.1	6.0	5.8	7.1
32		11.1	16.8	9.9	14	11	16
2	Random-walk FM	3.0	3.3	3.0	3.3	3.2	3.5
8		5.7	6.8	5.7	7.0	6.4	8.0
32		10.4	15.2	11	16	12	19

^a Confidence intervals for the nonoverlapped $\sigma_y(\tau)$ were obtained using df in Table 2 (16). The degrees of freedom used for the fully overlapped $\sigma_y(\tau)$ were computed using numerical methods and are approximately equal to those obtained using Table 2 (16, 19). Confidence intervals for the fully overlapped Mod $\sigma_y(\tau)$ were obtained from Ref. 33. $N = 1025$.

variance is given by

$$\chi^2 \left(\frac{1-p}{2} \right) < df \frac{\hat{\sigma}_y^2(\tau)}{\sigma_y^2(\tau)} < \chi^2 \left(p + \frac{1-p}{2} \right) \quad (35)$$

$$\frac{df}{\chi^2 \left(p + \frac{1-p}{2} \right)} \hat{\sigma}_y^2(\tau) < \sigma_y^2(\tau) < \frac{df}{\chi^2 \left(\frac{1-p}{2} \right)} \hat{\sigma}_y^2(\tau) \quad (36)$$

where the chi-squared value $\chi^2(C)$ for $(p \times 100)\%$ confidence can be obtained from chi-squared distribution tables or from several computer programs. The $(p \times 100)\%$ confidence interval for the Allan deviation is

$$\sqrt{\frac{df}{\chi^2 \left(p + \frac{1-p}{2} \right)}} \hat{\sigma}_y(\tau) < \sigma_y(\tau) < \sqrt{\frac{df}{\chi^2 \left(\frac{1-p}{2} \right)}} \hat{\sigma}_y(\tau) \quad (37)$$

The chi-squared distribution can also be used to find the confidence intervals for Mod $\sigma_y^2(\tau)$.

Walter (31) and Greenhall (32) have derived expressions for the number of degrees of freedom of Mod $\sigma_y^2(\tau)$ using different procedures. These expressions are complicated and will not be presented here. The two methods yield similar results (33). Table 3 shows the confidence intervals for $\sigma_y(\tau)$ with no overlap and full overlap (16,19), and for Mod $\sigma_y(\tau)$ (31–33) for the five noise types. In general, the confidence intervals for the fully overlapped $\sigma_y(\tau)$ are smaller than those for the nonoverlapped $\sigma_y(\tau)$. For random-walk FM noise, the confidence intervals for the nonoverlapped and the fully overlapped $\sigma_y(\tau)$ are approximately the same, although Table 3 shows a small degradation when using fully overlapped estimates. This degradation is due to the approximations used in the analytical expressions. Table 3 also shows that the confidence intervals for the fully overlapped Allan deviation are smaller than the ones for the fully overlapped modified Allan deviation. Nevertheless, the modified deviation is generally smaller than the Allan deviation, and thus the absolute confidence intervals for the two are similar.

BIBLIOGRAPHY

1. E. S. Ferre-Pikal et al., Draft revision of IEEE Std 1139-1988: Standard definitions of physical quantities for fundamental frequency and time metrology—random instabilities, *Proc. 1997 IEEE Int. Freq. Control Symp.*, 1997, pp. 338–357.
2. R. L. Sydner and D. W. Allan (eds.), *Handbook Selection and Use of Precise Frequency and Time Systems*, International Telecommunication Union, Geneva, Switzerland, 1997.
3. International Radio Consultative Committee (CCIR), Report 580, *Characterization of Frequency and Phase Noise*, 1986, pp. 142–150.
4. H. Hellwig, *Frequency Standards and Clocks: A Tutorial Introduction*, Technical Note 616, Washington, DC: U.S. National Bureau of Standards, 1977.
5. J. R. Vig and F. L. Walls, Fundamental limits on the frequency stabilities of crystal oscillators, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **42**: 576–589, 1995.
6. *IEEE Guide for Measurements of Environmental Sensitivities of Standard Frequency Generators*, IEEE Std. 1193, Piscataway, NJ: IEEE Press, 1994.
7. W. D. Lee et al., The accuracy evaluation of NIST-7, *IEEE Trans. Instrum. Meas.*, **44**: 120–124, 1995. An example of evaluating the uncertainties of an atomic frequency standard.
8. B. N. Taylor and C. E. Kuyatt, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, Technical Note 1297, Washington, DC: National Institute of Standards and Technology, 1994.
9. F. L. Walls, The influence of pressure and humidity on the medium and long-term frequency stability of quartz oscillators, *Proc. 42nd Annu. Symp. Freq. Control*, Baltimore, MD: 1988, IEEE Catalog No. 88CH2588-2, pp. 279–283.
10. J. J. Gagnepain, Sensitivity of quartz oscillators to the environment: Characterization methods and pitfalls, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **37**: 354, 1990.
11. F. L. Walls and J. J. Gagnepain, Environmental sensitivities of quartz oscillators, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **39**: 241–249, 1992.
12. J. R. Vig and T.R. Meeker, The aging of bulk acoustic wave resonators, filters and oscillators, *Proc. 45th Annu. Symp. Freq. Control*, Los Angeles, CA: 1991, IEEE, Catalog No. 91CH2965-2, pp. 77–101.

13. J. A. Barnes et al., Characterization of frequency stability, *IEEE Trans. Instrum. Meas.*, **IM-20**: 105–120, 1971.
14. D. W. Allan et al., Standard terminology for fundamental frequency and time metrology, *Proc. 42nd Annu. Symp. on Freq. Control*, Baltimore, MD: 1988, IEEE Catalog No. 88CH2588-2, pp. 419–425.
15. D. B. Sullivan et al., *Characterization of Clocks and Oscillators*, Technical Note 1337, National Institute of Standards and Technology, 1990.
16. D. A. Howe, D. W. Allan, and J. A. Barnes, Properties of signal sources and measurement methods, *Proc. 35th Annu. Symp. Freq. Control*, Electronic Industries Association 1981, Washington, DC, Philadelphia, PA: 1981, pp. A1–A47; also in Ref. 4.
17. T. E. Parker, Characteristics and sources of phase noise in stable oscillators, *Proc. 41st Annu. Symp. Freq. Control*, Philadelphia, PA: 1987, IEEE Catalog No. 87CH2427-3, pp. 99–110.
18. L. M. Nelson, C. W. Nelson, and F. L. Walls, Relationship of AM noise to PM noise in selected RF oscillators, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **41**: 680–684, 1994.
19. F. L. Walls, Correlation between upper and lower noise sidebands, to be published in *Proc. 1998 IEEE Int. Freq. Control Symp.*, Pasadena, CA, May 1998.
20. S. R. Stein, Frequency and time—their measurement and characterization, in E. A. Gerber and A. Ballato (eds.), *Precision Frequency Control*, New York: Academic Press, 1985, vol. 2, chap. 12.
21. D. W. Allan and J. A. Barnes, Modified Allan variance with increased oscillator characterization ability, *Proc. 35th Annu. Symp. Freq. Control*, Electronic Industries Assoc. 1981, Philadelphia, PA: 1981, pp. 470–474.
22. D. A. Howe, An extension of the Allan variance with increased confidence at long-term, *Proc. 1995 IEEE Int. Freq. Control Symp.*, San Francisco, CA: 1995, IEEE Catalog No. 95CH35752, pp. 321–329.
23. D. A. Howe, Methods of improving the estimation of long-term frequency variance, *Proc. European Frequency and Time Forum*, Swiss Foundation for Research in Microtechnology, Neuchâtel, Switzerland: 1997, pp. 91–99.
24. D. A. Howe and C. A. Greenhall, Total variance: A progress report on a new frequency stability characterization, *Proc. 29th Ann. Precise Time Time Interval PTTI Syst. Appl. Meet.*, Long Beach, CA: 1997, pp. 39–48.
25. F. L. Walls et al., Time-domain frequency stability calculated from the frequency domain: An update, *Proc. 4th Eur. Frequency Time Forum*, Swiss Foundation for Research in Microtechnology, Neuchâtel, Switzerland: 1990, pp. 197–204.
26. P. Lesage and C. Audoin, Characterization and measurement of time and frequency stability, *Radio Sci.*, **14** (4): 521–539, 1979.
27. P. Lesage and T. Ayi, Characterization of frequency stability: Analysis of the modified Allan variance and properties of its estimate, *IEEE Trans. Instrum. Meas.*, **IM-33**: 332–336, 1984.
28. C. A. Greenhall, Does the Allan variance determine the spectrum?, *Proc. 1997 IEEE Inter. Freq. Control Symp.*, Orlando, FL: 1997, IEEE Catalog No. 97CH36016, pp. 358–365.
29. F. L. Walls and S. R. Stein, *Servo techniques in oscillators and measurement systems*, Technical Note 692, Washington, DC: U.S. National Bureau of Standards, pp. 1–20, 1976.
30. J. Rutman and F. L. Walls, Characterization of frequency stability in precision frequency sources, *Proc. IEEE*, **79**: 952–960, 1991.
31. F. Vernotte et al., Oscillator noise analysis: multivariate measurement, *IEEE Trans. Instrum. Meas.*, **IM-42**: 342–350, 1993.
32. T. Walter, Characterizing frequency stability: A continuous power-law model with discrete sampling, *IEEE Trans. Instrum. Meas.*, **IM-43**: 69–79, Feb. 1994.
33. C. A. Greenhall, Estimating the modified Allan variance, *Proc. 1995 IEEE Int. Freq. Control Symp.*, San Francisco, CA: 1995, IEEE Catalog No. 95CH35752, pp. 346–353.
34. M. A. Weiss et al., Confidence on the modified Allan variance, *Proc. 9th Eur. Frequency Time Forum*, Besançon, France, 1995, pp. 153–165.

EVA S. FERRE-PIKAL

FRED L. WALLS

National Institute of Standards
and Technology

FREQUENCY STABILITY

Stable and spectrally pure signal generators have been widely employed in various fields of science and technology such as physics, high precision frequency standards, fundamental metrology, telecommunication systems, space missions, radars, and broadcasting. In addition to the inherent fluctuations in the output signal, almost all signal generators are influenced to some extent by their environment, such as changes in ambient temperature, supply voltage, magnetic field, barometric pressure, humidity, and mechanical vibration. These perturbations manifest themselves as noise in frequency or phase of the output signals and become the limiting factor in various applications. Therefore, it is of fundamental importance to characterize the frequency fluctuations in the output signal using a common measure as well as to reduce these fluctuations to an acceptable level.

Extensive research efforts have been devoted to the establishment of a general and common measure of frequency stability in the past 40 years. Tremendous progress has been achieved since 1955 through the development of high precision frequency standards, such as cesium beam clocks, optically pumped rubidium clocks, hydrogen masers, and quartz crystal oscillators. In the early 1960s the needs were clearly recognized for common parameters characterizing the frequency stability and for related measurement techniques. These parameters were required for at least two main purposes: The first purpose is to allow for meaningful comparisons between similar devices developed by different laboratories or between different devices in a given application; and the second one is to access application performance in terms of the measured oscillator frequency stability.

In 1964 a special symposium on frequency stability was organized by the National Aeronautics and Space Administration (NASA) and the Institute of Electrical and Electronics Engineers (IEEE) as an attempt to improve the situation (1). After this symposium, a Subcommittee on Frequency Stability was formed as a part of the Technical Committee on Frequency and Time of the IEEE Professional Group on Instrumentation and Measurement. In 1966 several members of this subcommittee contributed an original paper to a special issue of the *Proceedings of the IEEE* (2). In 1970 a report on the characterization of frequency stability was issued by the above mentioned subcommittee (3). Proposed definitions for the measure of frequency stability are the power spectral density in the frequency domain and the Allan variance in the time domain. Since then the concepts and methods proposed in this report have been widely accepted by many researchers and engineers in the field of time and frequency measurement.

In this article, a definition and a estimation procedure will be presented on the measure of frequency stability commonly employed. For more extensive and complete discussion, refer to the textbook and reviews on this subject (4,5).

BACKGROUND AND DEFINITION

Consider a sinusoidal signal generator whose instantaneous output voltage $u(t)$ may be written as

$$u(t) = [U_0 + \epsilon(t)] \sin[2\pi \nu_0 t + \phi(t)] \quad (1)$$

where U_0 and ν_0 are the nominal values of the amplitude and frequency, respectively. Throughout this article the Greek letter ν is used to stand for the signal frequency, while the Latin symbol f is used to denote the Fourier frequency in the representation of spectral densities. The parameter $\epsilon(t)$ and $\phi(t)$ in Eq. (1) represent the instantaneous amplitude and phase fluctuations, respectively. The instantaneous frequency $\nu(t)$ of the sinusoidal voltage is then expressed as the sum of a constant value ν_0 and variable term $\nu_v(t)$:

$$\nu(t) = \nu_0 + \frac{1}{2\pi} \frac{d\phi(t)}{dt} = \nu_0 + \nu_v(t) \quad (2)$$

Since we are dealing with stable oscillators, it is assumed that the magnitude of these fluctuations are much smaller than the nominal values; that is,

$$|\epsilon(t)| \ll U_0 \quad (3)$$

and

$$|\nu_v(t)| \ll \nu_0 \quad (4)$$

for substantially all time t . Oscillators with large frequency deviation are a subject of frequency modulation theory which is not treated in this article. Various types of oscillators are used in scientific and engineering fields, and their nominal frequencies cover a wide range of the spectrum—that is, from several hertz to several hundred terahertz (lightwave). For a general discussion of oscillators having a wide range of frequencies, it is useful to introduce normalization. The normalized frequency offset $y(t)$ from the nominal value is defined as

$$y(t) = \frac{\nu_v(t)}{\nu_0} = \frac{1}{2\pi \nu_0} \frac{d\phi(t)}{dt} \quad (5)$$

The quantity $y(t)$ is called the *fractional frequency offset*. Another useful quantity is the time integral of $y(t)$:

$$x(t) = \int_0^t y(t) dt = \frac{\phi(t)}{2\pi \nu_0} \quad (6)$$

and

$$y(t) = \frac{dx(t)}{dt} \quad (7)$$

The parameter $x(t)$ has the dimension of time and is proportional to instantaneous phase. This quantity was called the *phase-time* (Phasenzeit) by Becker (6).

From the definition of $y(t)$ given by Eq. (5), it is natural to assume that $y(t)$ should have a zero mean over the time of

observation, whereas this will not be the case for $x(t)$. However, most actual oscillators exhibit frequency drift with time as well as random variations. In performing a series of measurements over a long period of time, it is always possible to subtract the drift and the initial offset from the data.

There are two aspects in the analysis of measured results of $y(t)$, namely, time-domain analysis and frequency-domain analysis.

FREQUENCY DOMAIN

The behavior of $y(t)$ in the frequency domain is described by its power spectral density $S_y(f)$, which is defined as the Fourier transform of the autocorrelation function $R_y(\tau)$ given by

$$R_y(\tau) = \lim_{T \rightarrow \infty} \int_0^T y(t') y(t' + \tau) dt' \quad (8)$$

From the Wiener–Khinchine theorem, the power spectral density is obtained from the autocorrelation function as

$$S_y(f) = 4 \int_0^\infty R_y(\tau) \cos 2\pi f \tau d\tau \quad (9)$$

and inversely

$$R_y(\tau) = \int_0^\infty S_y(f) \cos 2\pi f \tau df \quad (10)$$

The power spectral density represents the fluctuation power of the fractional frequency offset $y(t)$ contained in a unit bandwidth as a function of the Fourier frequency. Its dimension is Hz^{-1} , since $y(t)$ and $R_y(\tau)$ are dimensionless.

The relations between the power spectral densities of various quantities are shown below.

Absolute frequency offset: $\delta\nu = \nu(t) - \nu_0 [\text{Hz}]$

$$S_{\delta\nu}(f) = \nu_0^2 S_y(f) [\text{Hz}] \quad (11)$$

Phase: $\phi(t)$ [radian]

$$S_\phi(f) = \frac{\nu_0^2}{f^2} S_y(f) [(\text{radian})^2/\text{Hz}] \quad (12)$$

Phase-time: $x(t) = \int_0^t y(t) dt$ [s]

$$S_x(f) = \frac{1}{(2\pi f)^2} S_y(f) [(\text{radian})^{-2}/\text{Hz}] \quad (13)$$

Angular frequency: $\omega = \dot{\phi} = d\phi/dt$ [radian/s]

$$S_\omega(f) = (2\pi \nu_0)^2 S_y(f) [(\text{radian})^2/\text{Hz}] \quad (14)$$

The above relations are very useful for converting measured data and for translating formula between various measures.

TIME DOMAIN

The measure of frequency stability in the time domain is based on the sample variance of the fractional frequency fluctuations. In actual measurements it is difficult to obtain the

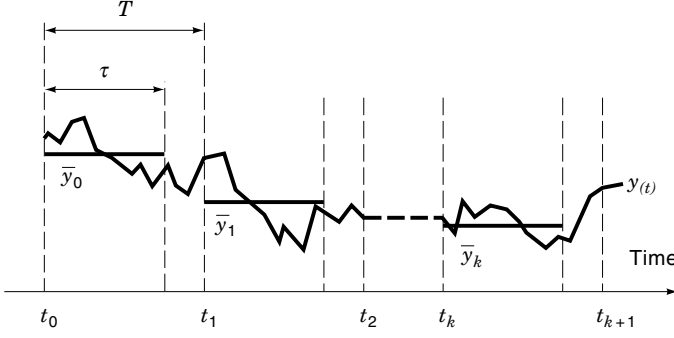


Figure 1. Measurement process for the calculation of sample variances. Here T and τ represent the repetition interval and duration of the measurement, respectively.

instantaneous sample of the fractional frequency offset $y(t)$. The results of frequency measurement are always in the form of sampled data of $\bar{y}_k$ of $y(t)$ averaged over a finite time interval τ , which is given by

$$\bar{y}_k(t_k, \tau) \equiv \frac{1}{\tau} \int_{t_k}^{t_k+\tau} y(t) dt = \frac{\phi(t_k + \tau) - \phi(t_k)}{2\pi \nu_0 \tau} \quad (15)$$

where $t_{k+1} = t_k + T$, $k = 0, 1, 2, \dots$, T represents the repetition interval for measurements of duration τ , and t_0 is arbitrary. The dead time between measurements is given by $T - \tau$. Figure 1 shows the measurement process for the sampled data $\bar{y}_k$.

The measure of frequency stability in the time domain is then defined in analogy to the sample variance by the relation

$$\sigma_y^2(N, T, \tau) \equiv \left\langle \frac{1}{N-1} \sum_{n=1}^N \left(\bar{y}_n - \frac{1}{N} \sum_{k=1}^N \bar{y}_k \right)^2 \right\rangle \quad (16)$$

where $\langle \rangle$ denotes the infinite time average. The quantity in Eq. (16) is called the *Allan variance* (7) and is dimensionless.

In many situations it is not correct to assume that the variance (16) converges to a meaningful limit as $N \rightarrow \infty$. In practice one cannot let N approach infinity, and it is known that some actual noise processes contain substantial fractions of the total noise power in the extremely low Fourier frequency range. Therefore it is important to specify particular N and T in order to improve the comparability of data.

The recommended definition of frequency stability is choosing $N = 2$ and $T = \tau$, which means no dead time between measurements. Expressing $\langle \sigma_y^2(N = 2, T = \tau, \tau) \rangle$ as $\sigma_y^2(\tau)$, the Allan variance may be written as

$$\sigma_y^2(\tau) = \left\langle \frac{(\bar{y}_{k+1} - \bar{y}_k)^2}{2} \right\rangle \quad (17)$$

In the practical situation, the true time average is not realizable and the estimate of $\sigma_y^2(\tau)$ should be obtained from a finite number of samples. It has been shown that when $\sigma_y^2(\tau)$ is estimated from the ensemble average of $\sigma_y^2(2, \tau, \tau)$, the average converges with increasing number of data even for noise processes that do not converge for $\langle \sigma_y^2(N, \tau, \tau) \rangle$ as $N \rightarrow \infty$. Therefore, $\sigma_y^2(\tau)$ has greater utility than $\langle \sigma_y^2(\infty, \tau, \tau) \rangle$ even though both involve an assumption of an infinite average. A widely used formula for

estimating $\sigma_y^2(\tau)$ experimentally is

$$\sigma_y^2(\tau, m) = \frac{1}{2(m-1)} \sum_{k=1}^{m-1} (\bar{y}_{k+1} - \bar{y}_k)^2 \quad (18)$$

where m represents the number of samples. The another advantage of the Allan variance with $N = 2$ and $T = \tau$ is the simplicity of computation from the measured data as shown by Eq. (18). In any case one should specify the number of samples used for estimation in order to avoid ambiguity and to allow for meaningful comparisons.

TRANSLATIONS BETWEEN MEASURES (FREQUENCY DOMAIN TO TIME DOMAIN)

The relation between the Allan variance and the power spectral density $S_y(f)$ was derived by Cutler and co-workers (3) and is expressed as

$$\langle \sigma_y^2(N, T, \tau) \rangle = \frac{N}{N-1} \int_0^\infty S_y(f) |H(f)|^2 df \quad (19)$$

where

$$|H(f)|^2 = \frac{\sin^2 \pi f \tau}{(\pi f \tau)^2} \left(1 - \frac{\sin^2 \pi r f \tau N}{N^2 \sin^2 \pi r f \tau} \right) \quad (20)$$

and

$$r = T/\tau \quad (21)$$

The parameter r represents the ratio of the time interval between successive measurements to the duration of the averaging periods. Equation (19) allows one to estimate the time-domain stability $\langle \sigma_y^2(N, T, \tau) \rangle$ from the frequency domain stability $S_y(f)$. Two assumptions were made in deriving Eq. (19). The first one is that $R_y(t_1 - t_2) = \langle y(t_1)y(t_2) \rangle$ exists; that is, $y(t)$ is stationary in the covariance sense. The second assumption is that

$$\langle y^2(t) \rangle = R_y(0) = \int_0^\infty S_y(f) df \quad (22)$$

exists. To satisfy the second assumption it is sufficient to assume that $S_y(f)$ is finite in the frequency interval $f_l < f < f_h$ and zero outside this interval; that is, there are lower and higher cutoff frequencies in the device and the measuring equipment. This condition is always satisfied in practical situations.

In the limit of $N \rightarrow \infty$, Eq. (19) reduces to

$$\lim_{N \rightarrow \infty} \langle \sigma_y^2(N, T, \tau) \rangle = \frac{1}{\pi \tau} \int_0^\infty S_y\left(\frac{u}{\pi \tau}\right) \frac{\sin^2 u}{u^2} du \quad (23)$$

$$u = \pi f \tau \quad (24)$$

In the special case of $N = 2$, Eq. (19) is written by

$$\langle \sigma_y^2(2, T, \tau) \rangle = \frac{2}{\pi \tau} \int_0^\infty S_y\left(\frac{u}{\pi \tau}\right) \frac{\sin^2 u \sin^2 ru}{u^2} du \quad (25)$$

By comparing Eqs. (23) and (24), it can be seen that the convergence on the lower limit is better for $N = 2$ because of the additional factor of $\sin^2 ru$. For the Allan variance with $N = 2$ and $T = \tau$, the translation is expressed as

$$\sigma_y^2(\tau) = 2 \int_0^\infty S_y(f) \frac{\sin^4 \pi f \tau}{(\pi f \tau)^2} df \quad (26)$$

TRANSLATIONS BETWEEN MEASURES (TIME DOMAIN TO FREQUENCY DOMAIN)

For general $\langle \sigma_y^2(N, T, \tau) \rangle$ no simple prescription is available for translation into the time domain. For this reason, $S_y(f)$ is preferred as a general measure of frequency stability, especially for theoretical work. For specific types of noise process discussed below, the Allan variance $\sigma_y^2(\tau)$ can be translated into the power spectral density $S_y(f)$.

OSCILLATOR NOISE MODEL

The types of noise observed on the output signal of actual oscillators can be suitably represented by the spectral density $S_y(f)$. It has been known empirically that a simple power-law model of the form

$$S_y(f) = h_{-2} f^{-2} + h_{-1} f^{-1} + h_0 + h_1 f + h_2 f^2 = \sum_{\alpha=-2}^2 h_\alpha f^\alpha \quad (27)$$

$$S_y(f) = 0 \quad \text{for} \quad f > f_h$$

can cover all actually known types of oscillators in the limit of drift elimination. In the above equation, h_α ($\alpha = -2, -1, 0, 1, 2$) is a constant. It is assumed that the measuring system has an ideally sharp upper cutoff frequency f_h . The individual terms have been identified by common names given in Table 1. Figure 2 shows the power spectral density $S_y(f)$ for five noise processes in Eq. (27). It can be seen that each noise process is clearly distinguishable from the slope.

It is easy to show the relationship between $S_y(f)$ defined above and $\sigma_y^2(\tau)$ by using the translation of Eq. (26). For every term of the form $h_\alpha f^\alpha$ ($\alpha = -2, -1, 0, 1, 2$) we have

$$\sigma_y^2(\tau) = \frac{2h_\alpha}{(\pi\tau)^{\alpha+1}} \int_0^{\pi\tau f_h} u^{\alpha-2} \sin^4 u du \quad (28)$$

where $u = \pi f \tau$. This relation may be expressed as

$$\sigma_y^2(\tau) = K_\alpha \tau^\mu \quad (29)$$

$$\mu = -\alpha - 1 \quad (30)$$

$$K_\alpha = \frac{2h_\alpha}{(\pi\tau)^{\alpha+1}} \int_0^{\pi\tau f_h} u^{\alpha-2} \sin^4 u du \quad (31)$$

Table 1. Common Names of Various Noise Terms

Random walk y	$\alpha = -2$	$\mu = 1$
Flicker y	$\alpha = -1$	$\mu = 0$
White y		
Random walk x	$\alpha = 0$	$\mu = -1$
Flicker x	$\alpha = 1$	$\mu = -2$
White x	$\alpha = 2$	$\mu = -3$

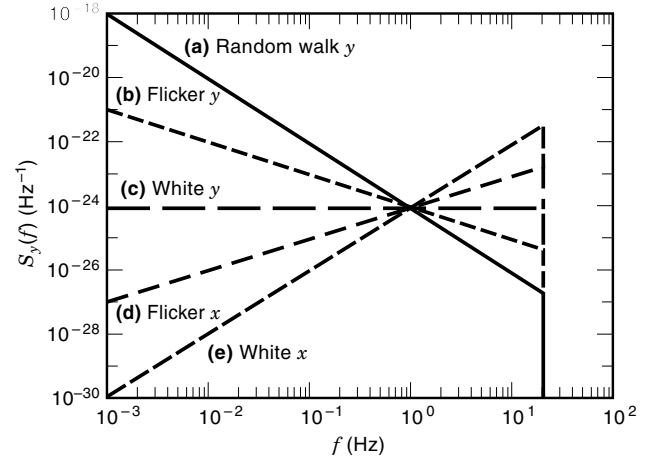


Figure 2. Power spectral density $S_y(f)$ for five noise terms in Eq. (27). A sharp upper cutoff frequency of $f_h = 20$ Hz is assumed. (a) Random walk frequency noise, (b) flicker frequency noise, (c) white frequency noise, (d) flicker phase noise, (e) white phase noise. $h_{-2} = h_{-1} = h_0 = h_1 = h_2 = 10^{-24}$.

For $\alpha < 1$ and $2\pi\tau f_h \gg 1$, K_α is independent of f_h and τ and becomes a constant due to the very rapid convergence of the integral. For $\alpha = 1$ and 2, the value of integral depends on f_h as well as on τ . The relations for general N and r are shown in Table 2. Figure 3 shows the dependence of $\sigma_y(\tau)$ on averaging time τ for five noise processes in the limit of $2\pi\tau f_h \gg 1$. It can be seen that the noise processes with $\alpha = 0, -1, -2$ are clearly distinguished from the slope of $\sigma_y(\tau)$. However, the slope of $\sigma_y(\tau)$ is almost the same for $\alpha = 1$ and $\alpha = 2$. As a consequence, the Allan variance $\sigma_y(\tau)$ is not useful for distinguishing flicker and white noise processes. For both types of noise, the dominant contribution to $\sigma_y(\tau)$ is frequency fluctuations at $f = f_h$ even for long measurement time. Therefore the determination of $\sigma_y(\tau)$ for some types of noise is dependent on the noise bandwidth and on the type of low-pass filter used in the measurement. The power expansion law of Eq. (27) has some physical meaning. Any practical oscillator contains a frequency-determining element (resonant circuit, quartz crystal resonator, atomic resonator, optical resonator) and a feedback loop. Any sources of noise have influences on the frequency or on the phase of the generated signal. Therefore, it is also useful to treat the noise in terms of phase fluctuations. Using Eq. (13) to transform Eq. (27) we can define the power spectral density of the phase-time fluctuations:

$$S_x(f) = \frac{1}{4\pi^2} \sum_{\alpha=-2}^2 h_\alpha f^{\alpha-2} \quad (32)$$

$$= \frac{1}{4\pi^2} (h_{-2} f^{-4} + h_{-1} f^{-3} + h_0 f^{-2} + h_1 f^{-1} + h_2)$$

The various types of noise characterized by α terms in Eqs. (27) and (32) usually dominate over some frequency range; in most cases, not all five terms are significant. The individual terms have been identified by common names given in Table 1.

Table 2. Stability Measure Conversion Chart

Frequency Domain (Power Spectral Densities)	$\sigma_y^2(\tau)$ [$N = 2, r = 1$]	Time Domain (Allan Variances) $\langle \sigma_y^2(N, T = \tau, \tau, f_h) \rangle$ [$r = 1$]	$\{\sigma_y^2(N, T, \tau, f_h)\}$
White x $S_y(f) = h_2 f^2$ $S_x(f) = h_2 / (2\pi)^2$ $2\pi f_h \tau \gg 1$	$h_2 \cdot \frac{3f_h}{(2\pi)^2 \tau^2}$	$h_2 \cdot \frac{N+1}{N(2\pi)^2} \cdot \frac{2f_h}{\tau^2}$	$h_2 \cdot \frac{N + \delta_k(r-1)}{N(2\pi)^2} \cdot \frac{2f_h}{\tau^2}$ $\delta_k(r-1) = 1, r = 1$ $= 0, \text{otherwise}$
Flicker x $S_y(f) = h_1 f$ $S_x(f) = h_1 / (2\pi)^2 f$ $2\pi f_h \tau \gg 1, 2\pi f_h T \gg 1$	$h_1 \cdot \frac{\frac{9}{2} + 3 \ln(2\pi f_h \tau) - \ln 2}{(2\pi)^2 \tau^2}$	$h_1 \cdot \frac{2(N+1)}{N\tau^2(2\pi)^2} \left[\frac{3}{2} + \ln(2\pi f_h \tau) - \frac{\ln N}{N^2 - 1} \right]$	$h_1 \cdot \frac{2}{(2\pi \tau)^2} \left\{ \frac{3}{2} + \ln(2\pi f_h \tau) + \frac{1}{N(N-1)} \times \sum_{n=1}^{N-1} (N-n) \cdot \ln \left[\frac{n^2 r^2}{n^2 r^2 - 1} \right] \right\}, r \gg 1$
White y (random walk x) $S_y(f) = h_0$ $S_x(f) = h_0 / (2\pi)^2 f^2$	$h_0 \cdot \frac{1}{2} \tau^{-1}$	$h_0 \cdot \frac{1}{2} \tau^{-1}$	$h_0 \cdot \frac{1}{2} \tau^{-1}, r \geq 1$ $h_0 \cdot \frac{1}{6} r(N+1) \tau^{-1}, Nr \leq 1$
Flicker y $S_y(f) = \frac{h_{-1}}{f}$ $S_x(f) = h_{-1} / (2\pi)^2 f^3$	$h_{-1} \cdot 2 \ln 2$	$h_{-1} \cdot \frac{N \ln N}{N-1}$	$h_{-1} \cdot \frac{1}{N(N-1)} \sum_{n=1}^{N-1} (n-n) [-2(nr)^2 \ln(nr) + (nr+1)^2 \ln(nr+1) + (nr-1)^2 \ln nr-1]$
Random walk y $S_y(f) = \frac{h_{-2}}{f^2}$ $S_x(f) = h_{-2} / (2\pi)^2 f^4$	$h_{-2} \cdot \frac{(2\pi)^2 \tau}{6}$	$h_{-2} \cdot \frac{(2\pi)^2 \tau}{12} \cdot N$	$h_{-2} \cdot \frac{(2\pi)^2 \tau}{12} [r(N+1) - 1], r \geq 1$

The types of noise commonly observed in oscillators are as follows:

1. *Additive Noise.* Thermal noise in amplifier is simply added to the signal. This type of noise appears as phase noise and is usually white with high cutoff frequency.
2. *Perturbing Noise.* Thermal and shot noise acting within the feedback loop appears as white frequency noise (random walk in phase).
3. *Modulating Noise.* This type of noise is caused by random variations of reactive parameters such as semiconductor

junction capacitors, frequency-determining parameters of resonators, and so on. These fluctuations are either inherent in the devices or due to environmental effect. Flicker x and y usually belong to this class.

MODIFIED ALLAN VARIANCE

To improve the relatively poor discrimination of the Allan variance $\sigma_y(\tau)$ against flicker ($\alpha = 1$) and flicker ($\alpha = 2$) phase noise, the modified Allan variance was introduced in 1981. The definition is based on (a) the algorithm developed by Snyder (8) for increasing the resolution of frequency meters and (b) the detailed consideration of Allan and Barnes (9).

It consists in dividing a time interval τ into n cycles of period τ_0 such that

$$\tau = n\tau_0. \quad (33)$$

As depicted in Fig. 4, for a given observation time interval of duration 2τ , there are n overlapping time intervals of duration

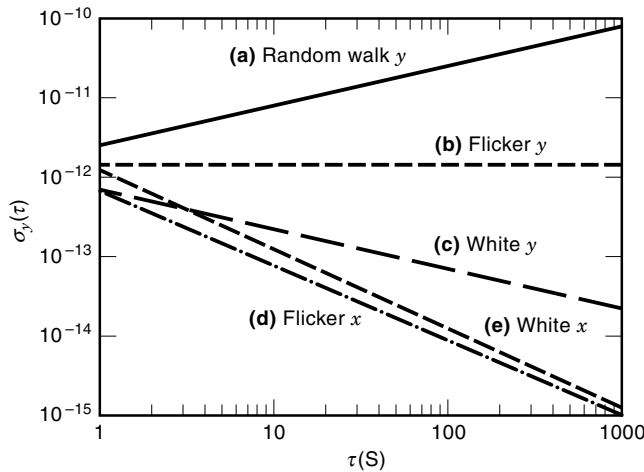


Figure 3. Square root of the Allan variance $\sigma_y(\tau)$ for five noise processes in Eq. (27) in the limit of $2\pi\tau f_h \gg 1$, where f_h represents the sharp upper cutoff frequency. (a) Random walk frequency noise $S_y(f) = h_{-2}f^{-2}$, (b) flicker frequency noise $S_y(f) = h_{-1}f^{-1}$, (c) white frequency noise $S_y(f) = h_0$, (d) flicker phase noise $S_y(f) = h_1f^{-1}$ with $f_h = 20$ Hz, (e) white phase noise $S_y(f) = h_2f^{-2}$ with $f_h = 20$ Hz. $h_{-2} = h_{-1} = h_0 = h_1 = h_2 = 10^{-24}$.

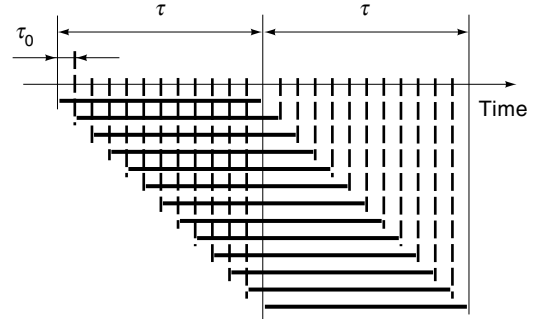


Figure 4. Measurement process for the calculation of the modified Allan variance.

τ . Allan and Barnes introduced the modified Allan variance such that

$$\text{Mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2} \left\langle \left[\frac{1}{n} \sum_{k=1}^n \left\{ \int_{t_0+(k+n)\tau_0}^{t_0+(k+2n)\tau_0} y(t) dt - \int_{t_0+k\tau_0}^{t_0+(k+n)\tau_0} y(t) dt \right\} \right]^2 \right\rangle \quad (34)$$

It can be seen from the above equation that the calculation of each statistical sample involved in the definition of $\text{Mod } \sigma_y^2(\tau)$ requires a signal observation of duration 3τ .

For $n=1$, the Allan variance and the modified one are equal.

$$\text{Mod } \sigma_y^2(\tau) = \sigma_y^2(\tau) \quad (35)$$

The power spectral density $S_y(f)$ is translated into the modified Allan variance $\text{Mod } \sigma_y^2(\tau)$ by the following relation:

$$\text{Mod } \sigma_y^2(\tau) = \frac{2}{n^2 \pi^2 \tau^2} \left[n \int_0^\infty \frac{S_y(f)}{f^2} \sin^4(\pi f n \tau_0) df + 2 \sum_{k=1}^{n-1} (n-k) \int_0^\infty \frac{S_y(f)}{f^2} \cos(2\pi f k \tau_0) \sin^4(\pi f n \tau_0) df \right] \quad (36)$$

The analytical expressions of the modified Allan variance for each noise term in Eq. (27) can be directly calculated by the above equation. These relations are summarized in Table 3 in the limit of $2\pi \tau f_h \gg 1$ (10).

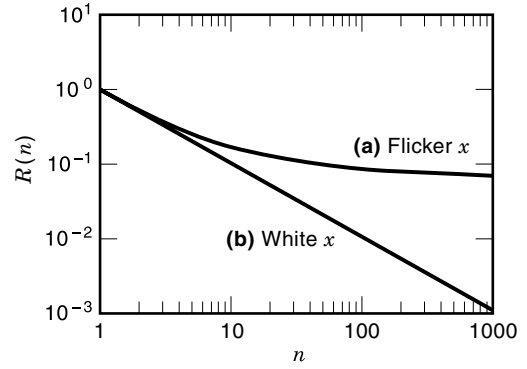


Figure 5. Dependence of the ratio $R(n)$ on n for (a) flicker and (b) white phase noise processes under the condition $2\pi \tau f_h \gg 1$.

It is useful for comparing the Allan variance with the modified one to define the ratio $R(n)$ as

$$R(n) = \frac{\text{Mod } \sigma_y^2(\tau)}{\sigma_y^2(\tau)} \quad (37)$$

Figure 5 depicts the variation of the ratio $R(n)$ with n , for flicker and white phase noise processes. It can be seen that for a large value of n , flicker and white phase noise processes have different dependencies. This property can be used to distinguish these two types of noise processes.

EXAMPLE OF THE ALLAN VARIANCE

Sinusoidal Frequency Modulation

Consider a frequency-modulated (FM) signal with the modulation frequency of f_m and maximum frequency deviation of $\Delta\nu_0$.

Table 3. Conversion Chart for the Modified Allan Variance

Frequency Domain (Power Spectral Densities)	Time Domain (Allan Variances) $\text{Mod } \sigma_y^2(\tau)$
White x $S_y(f) = h_2 f^2$ $S_x(f) = h_2 / (2\pi)^2$ $2\pi f_h \tau \gg 1$	$h_2 \cdot \frac{3f_h}{8\pi n \tau^2}$
Flicker x $S_y(f) = h_1 f$ $S_x(f) = h_1 / (2\pi)^2 f$ $2\pi f_h \tau \gg 1, 2\pi f_h T \gg 1$	$h_1 \cdot \frac{1}{4\pi^2 n^2 \tau^2} \left[3n \ln(2\pi f_h \tau) + \sum_{k=1}^{n-1} (n-k) \left\{ 4 \ln \left(\frac{n^2}{k^2} - 1 \right) - \ln \left(\frac{4n^2}{k^2} - 1 \right) \right\} \right]$
White y (random walk x) $S_y(f) = h_0$ $S_x(f) = h_0 / (2\pi)^2 f^2$	$h_0 \cdot \frac{n^2 + 1}{4n^2 \tau^2}$
Flicker y $S_y(f) = \frac{h_{-1}}{f}$ $S_x(f) = h_{-1} / (2\pi)^2 f^3$	$h_{-1} \cdot \frac{2 \ln 2}{n^2} \left[\frac{4n^2 - 3n + 1}{n^2} + \frac{1}{n^2 \ln 2} \sum_{k=1}^{n-1} (n-k) \left\{ \frac{n}{2} [(k+2n) \ln(k+2n) - (k-2n) \ln(2n-k)] + \frac{1}{2} (k+n)(k=2n) \ln(k+n) + \frac{1}{2} (k-n)(k+2n) \ln k-n + 3k^2 \ln k - k \left[(n+2k) \ln \left(k + \frac{n}{2} \right) - (n-2k) \ln \left k - \frac{n}{2} \right \right] \right\} \right]$
Random walk y $S_y(f) = \frac{h_{-2}}{f^2}$ $S_x(f) = h_{-2} / (2\pi)^2 f^4$	$h_{-2} \cdot \frac{2\pi^2 \tau}{3} \left(\frac{33}{40} + \frac{1}{8n^2} + \frac{1}{20n^4} \right)$

The normalized frequency offset $y(t)$ is given by

$$y(t) = \frac{\Delta\nu_0}{\nu_0} \sin(2\pi f_m t) \quad (38)$$

The power spectral density for this signal is expressed as

$$S_y(f) = \frac{1}{2} \left(\frac{\Delta\nu_0}{\nu_0} \right)^2 \delta(f - f_m) \quad (39)$$

where δ is the Dirac delta function. Substitution of Eq. (39) into Eq. (26) and (36) yields

$$\sigma_y^2(\tau) = \left(\frac{\Delta\nu_0}{\nu_0} \right)^2 \frac{\sin^4(\pi f_m \tau)}{(\pi f_m \tau)^2} \quad (40)$$

and

$$\text{Mod } \sigma_y^2(\tau) = \left(\frac{\Delta\nu_0}{\nu_0} \right)^2 \frac{\sin^6(\pi f_m n \tau_0)}{n^2 (\pi f_m n \tau_0)^2 \sin^2(\pi f_m \tau_0)} \quad (41)$$

respectively. The effect of sinusoidal FM in both cases is 0 when τ equals the modulation period $T_m = f_m^{-1}$ or one of its multiples, since the modulating signal is completely averaged away. The largest value of $\text{Mod } \sigma_y(\tau)$ occurs when τ is $T_m/2$ or one of its odd multiples. $\text{Mod } \sigma_y(\tau)$ falls n times faster than $\sigma_y(\tau)$ for sinusoidal FM.

Linear Frequency Drift

When linear frequency drift exists (i.e., $y(t) = dt$), no tractable model exists for the power spectral density $S_y(f)$. Direct calculation in the time domain using Eqs. (17) and (34) yields

$$\sigma_y(\tau) = \frac{d}{\sqrt{2}} \tau \quad (42)$$

and

$$\text{Mod } \sigma_y(\tau) = \frac{d}{\sqrt{2}} \tau \quad (43)$$

respectively. Thus linear frequency drift yields τ^{+1} law for both $\sigma_y(\tau)$ and $\text{Mod } \sigma_y(\tau)$.

OTHER MEASURES OF FREQUENCY STABILITY

A number of other measures have been proposed and used during the past 25 years. Each measure has some advantages and limitations compared with the well-established power spectral density and the Allan variance. The Hadamard variance (11) has been developed for high-resolution spectral analysis of $y(t)$ from measurements of $\bar{y}_k$. The high-pass variance (12) has been proposed through the transfer function approach and is defined by

$$\sigma^2(\tau) = \int_0^\infty S_y(f) |H(f)|^2 df \quad (44)$$

It was shown that the Allan variance can be estimated by high-pass filtering the demodulated phase noise without using counting technique. A band-pass variance (12) has also been

proposed to distinguish white and flicker phase noise processes. A filtered Allan variance (13) has been used to separate various noise processes.

BIBLIOGRAPHY

1. *Proceedings of the IEEE-NASA Symposium on Short Term Frequency Stability*, Goddard Space Flight Center, Greenbelt, MD, Nov. 23–24, 1964; Document NASA SP-80.
2. Special issue on frequency stability, *Proc. IEEE*, **54**: 101–338, 1966.
3. J. A. Barnes et al., Characterization of frequency stability, *NBS Technical Note* 394, US Government Printing Office, Washington, D.C., October 1970; also *IEEE Trans. Instrum. Meas.*, **IM-20**: 105–120, 1971.
4. P. Kartaschoff, *Frequency and Time*, London: Academic Press, 1978.
5. J. Rutman and F. L. Walls, Characterization of frequency stability in precision frequency sources, *Proc. IEEE*, **79**: 952–960, 1991.
6. G. Becker, Über die Begriffe Phase—Phasenzeit und Phasenwinkel bei zeitabhängigen Vorgängen, *PTB—Mitteilungen*, **81**: 348–352, 1971.
7. D. W. Allan, Statistics of atomic frequency standards, *Proc. IEEE*, **54**: 221–230, 1966.
8. J. J. Snyder, Algorithm for fast digital analysis of interference fringes, *Appl. Opt.*, **19**: 1223–1225, 1980.
9. D. W. Allan and J. A. Barnes, A modified “Allan variance” with increased oscillator characterization ability, In *Proceedings of the 35th Annual Frequency Control Symposium*, 1981, pp. 470–475.
10. P. Lesage and T. Ayi, Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate, *IEEE Trans. Instrum. Meas.*, **IM-33**: 222–225, 1971.
11. R. A. Baugh, Frequency modulation analysis with the Hadamard variance. In *Proceedings of the 25th Annual Frequency Control Symposium*, 470–475, 1981.
12. J. Rutman, Characterization of frequency stability: A transfer function approach and its application to measurement via filtering of phase noise, *IEEE Trans. Instrum. Meas.*, **IM-23**: 40–48, 1974.
13. J. Gros Lambert et al., A new “filtered Allan variance” and its application to the identification of phase and frequency noise sources, In *Proceedings of the 43rd Annual Frequency Control Symposium*, 326–330, 1989.

HIDEMI TSUCHIDA
Electrotechnical Laboratory

PHASE METERS

Phase meters measure a time-domain position of one periodic signal, including electromagnetic waves, with respect to another periodic signal of the same frequency, then relate the resulting time difference to the period to display it in terms of the phase angle.

Measurement techniques applied in phase meters depend upon the frequencies of the tested signals. In the range of lower frequencies, below several megahertz, the signals under test are acquired and directly processed to produce the phase reading. Analog or digital time-into-output reading conversion methods can be used.

Application of internally generated phase references improves measurement accuracy. With the recent development of fast real-time data acquisition systems, sampling methods are

used to compute wave parameters, like phase shift and magnitude, average value, and others.

In the higher frequency region, above several megahertz, the linear down-conversion of both tested signals is applied to reduce the operating frequency so lower frequency measurement techniques can be used. The frequency down-conversion employs sinusoidal heterodyning, persecuting or down counting, and coherent or periodic random sampling.

Besides stand-alone phase measuring devices, built-in phase meters are used as parts of impedance meters, power meters, frequency meters, network and waveform analyzers, as well as analog and digital oscilloscopes. Some high frequency wave analyzers and digital oscilloscopes use coherent or random sampling methods to reconstruct the wave images and display them on the screen before processing. Direct oscilloscope two-channel observation methods and indirect $x - y$ ellipsis display methods are discussed in Oscilloscopes.

Typical phase measurement is based on the assumptions that the investigated waves have the same periods and that they are sampled at the same points relative to the cycle start.

Many direct and indirect or conversion phase meters allow reduced random errors by accumulating many readings and computing average values.

Microwave six-port network analyzers (1,2) and homodyne four-port meters (3) are a separate group of phase meter related instruments which use scalar power or voltage readings to compute phase shift and other parameters of the networks.

REVIEW OF THE PHASE MEASUREMENT METHODS

In this section we review the basic methods of phase measurement which are applied in the majority of phase meters. The details related to the specific implementations are described in the later parts of the article.

As mentioned before, the oscilloscope $x - y$ method with Lissajous patterns is given in the article entitled Oscilloscopes in greater detail. The phase shift is determined from the following equation for signals of the same frequency

$$PH = \arcsin(A/B) \quad (1)$$

where PH is the phase shift, and A and B are as shown in Fig. 1.

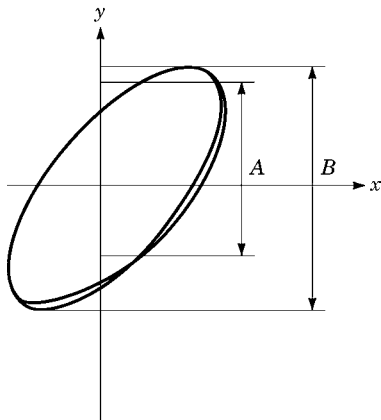


Figure 1. Phase measurement using Lissajous pattern for two signals of the same frequency.

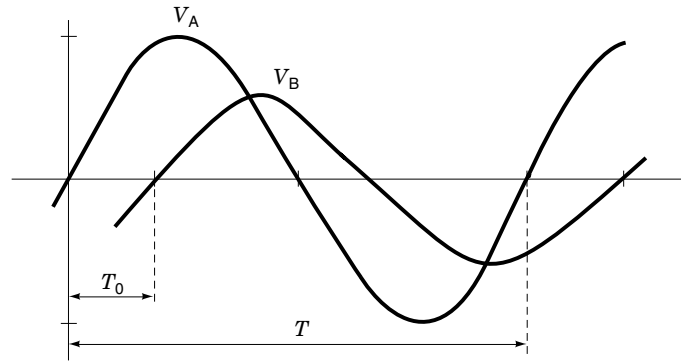


Figure 2. Phase measurement using sweep method.

The accuracy of this method is limited by the errors of the vertical and horizontal channels and by the error of reading from the oscilloscope screen. The total error may exceed 5%.

Another visual method, which could be digitized, is the sweep method. Two waves are displayed on the screen, as shown in Fig. 2. The phase reading involves measurements of the period and the time delay between the two waves. The following expression is used to determine the phase

$$PH = 360 \deg T_0/T \quad (2)$$

where T_0 is the delay between the waves and T is the period.

The visual method leads to the error which could be much more than several percent when the phase between the waves is small. Additional error is introduced due to channel misalignment and vertical wave offset. Some analog and digital oscilloscopes have digital signal processing packages which automate the measurement process, so the error of reading from the screen can be significantly reduced.

The wave offset error is compensated in some analog and digital phase meters by the proper technique described in (15). The phase measurement involves the process of conversion of the input waves into standardized rectangular signals, as shown in Fig. 3. The time shift preserved in the front edge differential signal is related to the entire period thus resulting in the phase shift.

The same differential signal, V_{DD} of Fig. 3 can be averaged to provide direct information about the phase shift. Proper calibration of the averaging device provides correct readings. The lead or lag information is indicated by additional logic circuitry described later in this paper.

Many waveform analyzers, including more complex digital oscilloscopes, accumulate the signal samples from the investigated signals and process them in order to determine various signal parameters. Among them, the phase shift is obtained from the computations.

GENERAL BLOCK DIAGRAMS OF DIRECT PHASE METERS

Figure 4 is a general block diagram of a direct phase meter system measuring phase without frequency down-conversion. Two input waves are applied to the inputs A and B. The waves are converted into trains of pulses of the same lengths and delivered to the processor, which determines the phase difference

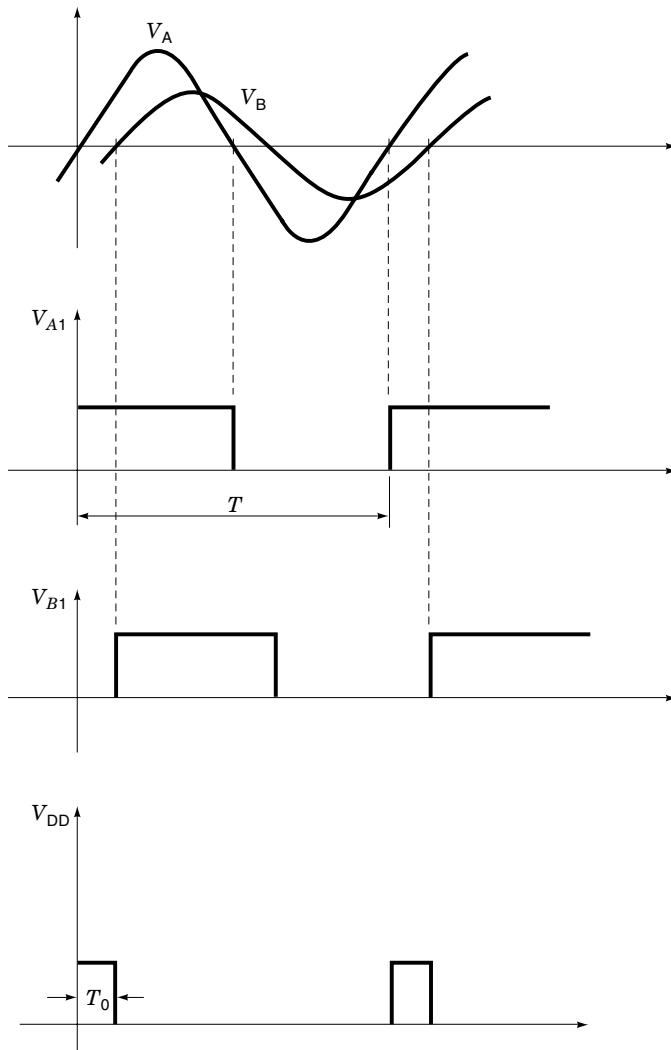


Figure 3. Phase measurement using wave amplitude standardization (for both analog and digital measurements).

in an analog or digital fashion. Connection of the frequency down-converters to both inputs greatly expands the frequency range of operation.

The heart of the direct phase meter is a phase-to-output variable converter whose major function is translating the phase difference of the investigated signals into a quantity like voltage, time ratio, or frequency. Figure 5 illustrates the first type of converter that changes input signal phase difference into

output voltage. The output voltage can be measured directly by a voltmeter or converted again into current and measurement by an ammeter. Analog-to-digital conversion process added to the converter of Fig. 5 changes it into a simple digital phase meter. The way the results are displayed depends upon the type of applied phase detector. Phase meters with two-quadrant, digital or unipolar type of a phase detector produce the results in terms of phase magnitude and lead/lag indication. Meters with a four-quadrant phase detector deliver bipolar readings thus indicating negative or positive phase shift when the positions of both waves are shifted more or less than 90° . Figure 6(a) shows a two-quadrant phase detector with Exclusive-NOR gate whose inputs are controlled by two binary signals formed from the input tested signal and reference signal. The gate output signal is filtered to yield the dc level proportional to the phase shift. The static transfer characteristic of Fig. 6(b) illustrates the relationship between the output voltage level and the phase difference between the input signals. The detector registers only the phase magnitude. Additional circuitry is necessary to indicate the phase polarity.

The four-quadrant phase detector of Fig. 7(a) includes two clocked D flip-flops. The binary input signals shifted in phase are applied to the clock inputs while the D inputs are high. The two output signals from the flip-flops are subtracted in a bipolar analog subtractor and then filtered by a low-pass filter. The detector exhibits inconsistent readings near -180° and 180° which is shown as sudden changes in the transfer characteristic of Fig. 7(b).

Another four-quadrant product detector and its static transfer characteristic are shown in Fig. 8. This type of a detector is often used in phase-lock-loop systems.

A classical digital meter shown in Fig. 9 uses two channel time interval measurement that combines the time shift between the waves and the period measurements. The reading is given as a ratio of these two. The meter could be a stand-alone unit or it could be a part of a frequency counter or an oscilloscope.

In some digital meters, the time shift between the waves can be initially converted into gate pulses whose duration is measured by a single channel time interval meter. The reading related to the period constitutes the measure of the phase shift (4). Figure 10 shows a phase-to-output variable converter in which an internal counter locks in to the input signal frequency producing a multiple frequency of that of the external signal. An internal counter counts the multiple-frequency pulses, and if the multiplication factor is 360 with a resolution of one degree, the phase shift is given directly in degrees. Larger multiplication factors, like 3600, offer better resolution. Application of the high-frequency clock whose frequency is much greater

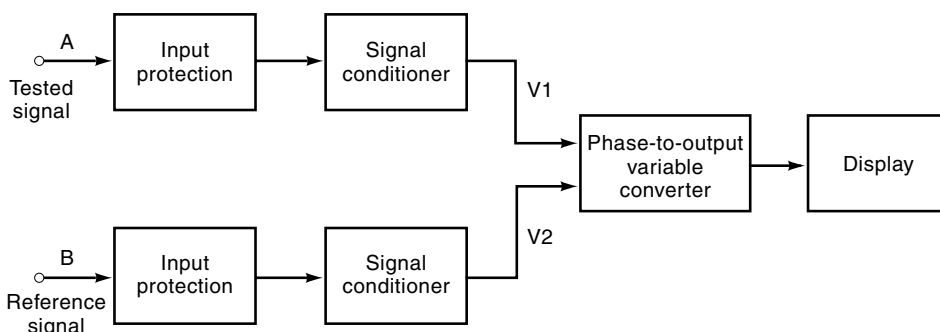


Figure 4. General block diagram of a phase meter.

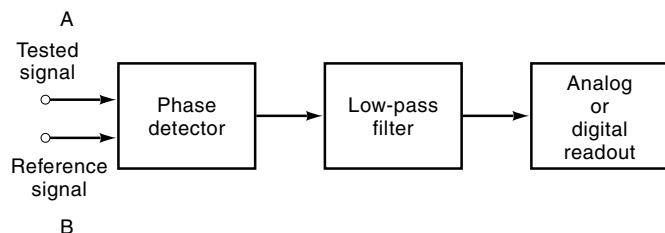


Figure 5. Phase-to-output variable converter with phase detectors.

than the input signal repetition frequency requires compatible high-frequency counters (5).

Real-time sampling technique which is applied in modern data acquisition systems allows computations of the phase shift between two signals. The two signals are applied to the samplers through anti-aliasing filters. The sampled signals are then converted by analog-to-digital converters and stored in processor memory for further processing. The phase shift is calculated from the samples located near zero-crossings for both signals. Basic phase error is related to uncertainty of these samples (6).

Complex impedance bridges and complex transmission coefficient bridges also measure phase shift with the help of precision phase shifters. In the impedance meters phase measurements are accomplished with reference to the calibrated phase shifter by adjusting and comparing the phase shifter and system attenuator readings for a known and unknown impedance by setting the output of a video detector to null. In the transmission coefficient bridges the signal is applied to the device under test and to the variable phase shifter. A general block diagram of such a meter is shown in Fig. 11, and a simplified diagram of a millimeter wave transmission coefficient bridge is shown in Fig. 12 (7).

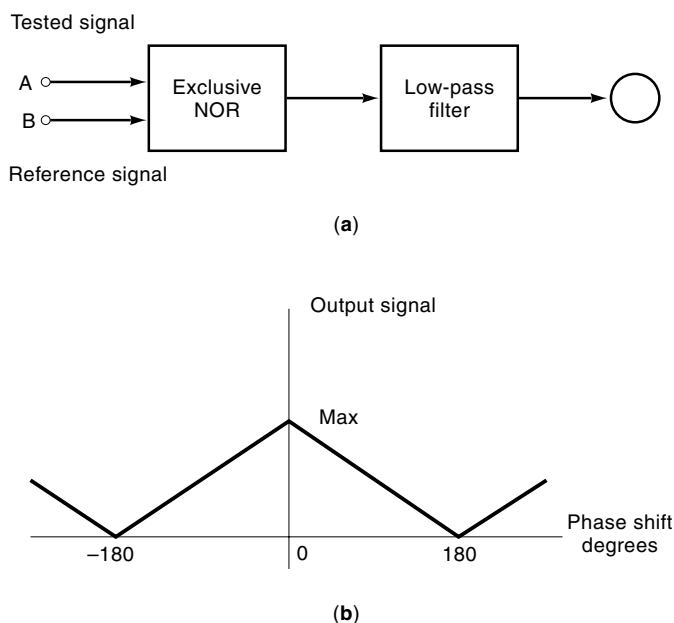


Figure 6. Two-quadrant phase detector. (a) Diagram. (b) Static transfer characteristic.

DOWN-CONVERSION TECHNIQUES USED IN PHASE METERS

Phase-to-Output Variable Converters previously described are limited in their operation frequency range by the speed of the input circuits, logic circuits in a counter, and circuits of a phase detector. Although the frequency counters can directly measure frequencies of signals reaching far more than several megahertz, time interval measurements are affected by large errors which reduce operating frequencies for phase measurements.

In order to extend the phase meter operating frequency range the following techniques are used:

1. Persecuting or down counting (8)
2. Sinusoidal heterodyne conversion (5,9)
3. Coherent sampling (10)
4. Random repetitive sampling (11,12,13)

All methods involve acquisition of many periods of the signal under test.

Application of the down-conversion technique in signal and network analyzers, impedance meters, and multichannel microwave receivers increases the frequency range of phase measurement to millimeter waves. The most critical issue in high-frequency tests related to wave separation is solved with the aid of directional couplers and hybrid-tee connectors.

Persecuting or Down Counting

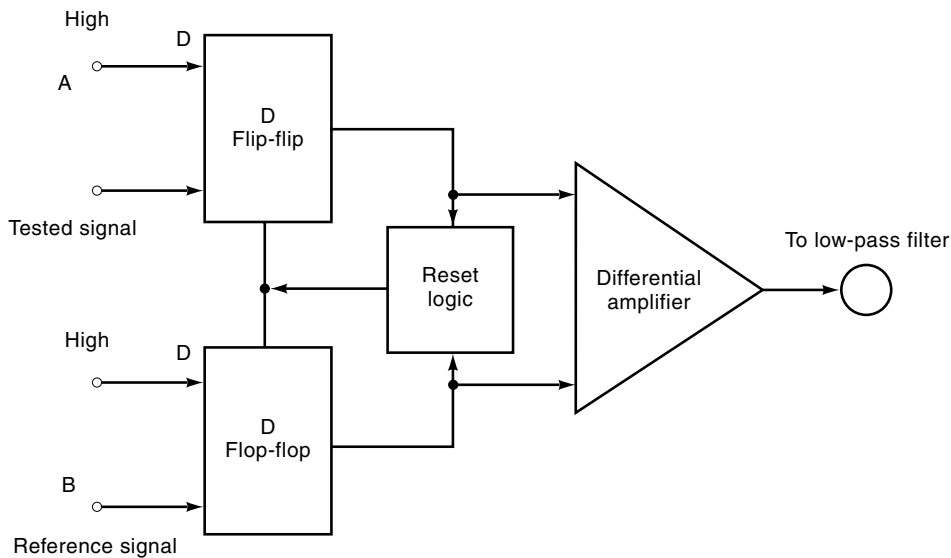
Two input signals, the signal under test and the reference signal, are applied to two identical frequency dividers. As a result, lower frequency signals are measured in analog or digital circuits previously described. This method has found applications in high-speed frequency counters which also have the capability of time interval measurements leading to phase determination.

Sinusoidal Heterodyne Conversion

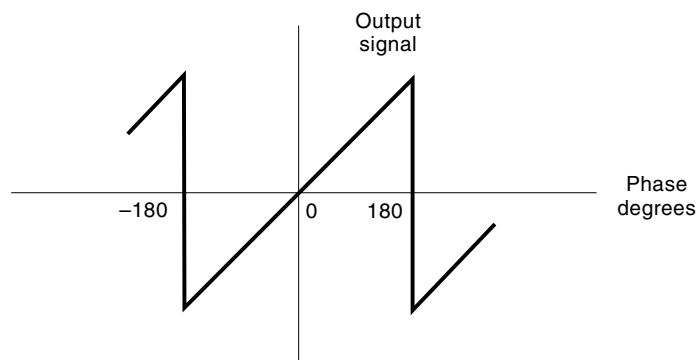
The heterodyne, single or double down-conversion process, reduces the frequency of the tested signals by mixing so they can be acquired and directly processed by lower frequency phase meters. The block diagram of a typical heterodyne phase meter is shown in Fig. 13. The system is designed to lock-in the voltage controlled oscillator to the incoming signal frequency of the reference channel so a signal of fixed intermediate frequency is produced. The voltage controlled oscillator also mixes with the other signal to reduce its frequency to the same intermediate frequency as that of the reference signal. Intermediate frequency signals are then processed by a classical low-frequency phase meter.

Coherent Sampling

The phase meter uses a fast sampling technique to down convert periodic high frequency signals to intermediate frequency signals. Initially, the internally generated sampling pulses are swept automatically until their frequency is locked to produce the intermediate frequency signals, as in Fig. 14. Two high-frequency input signals are sampled in the sampling gates which are switched on by very short sampling pulses. The samples are stored in analog memories (capacitors of



(a)



(b)

Figure 7. Four-quadrant phase detector. (a) Diagram. (b) Static transfer characteristic.

sample-and-hold circuits). The sampling rate is automatically adjusted so each sample occurs at an earlier relative point on the waves during a subsequent recurrence of the waves. The reconstructed waveforms after sampling resemble the original waves but they have much lower frequencies. The sampling pulse frequency is controlled by a phase-locked-loop system which maintains the desired intermediate frequency at the sampler output. The method is applied in vector voltmeters, network analyzers, and impedance meters to measure phase shift and the magnitude of two different signals.

Another solution, shown in Fig. 15, involves a regular subsequent process of sampling typically used in sampling wave analyzers and oscilloscopes. Two input signals are sampled in sampling gates by very narrow sampling pulses whose timing is coherently triggered by the waves of lower frequencies which are synchronous submultiples of the input signal frequencies. Subsequent samples are automatically delayed so the output waves look like time-transformed input signals. The sampling rates do not exceed several hundred kHz while the input signal frequencies could reach 20 GHz or more. The time-domain transformation factors are very high. For instance, if 10 GHz waves are tested and 1000 samples are taken to reconstruct a period, with the samples located 10 μ s apart, then the

time transformation factor is 10^8 . This allows us to determine the phase using a low frequency processor calculating signal parameters or a direct phase meter (14).

Random Repetitive Sampling

Random repetitive sampling is applied in modern digital scopes which can also measure time intervals and phase shifts. A typical block diagram of major parts of such a scope is shown in Fig. 16. Two analog signals applied to the inputs are converted into corresponding digital signals. The sampling process in analog-to-digital converters, including sample-and-hold blocks, is controlled by an acquisition processor. The analog signals are used to trigger the time base of the acquisition processor. The samples are not collected in a sequential order, and the processor determines their proper location in relation to the trigger signals synchronized with the input waves. The voltage and time coordinates of individual samples are stored in the processor memory. The waves can be displayed on the scope screen and all information about the waves is also available for further processing. The phase shift between the two tested signals is measured with measurement/storage modules, such as HP 54657A, HP 54658A, and HP 54659B.

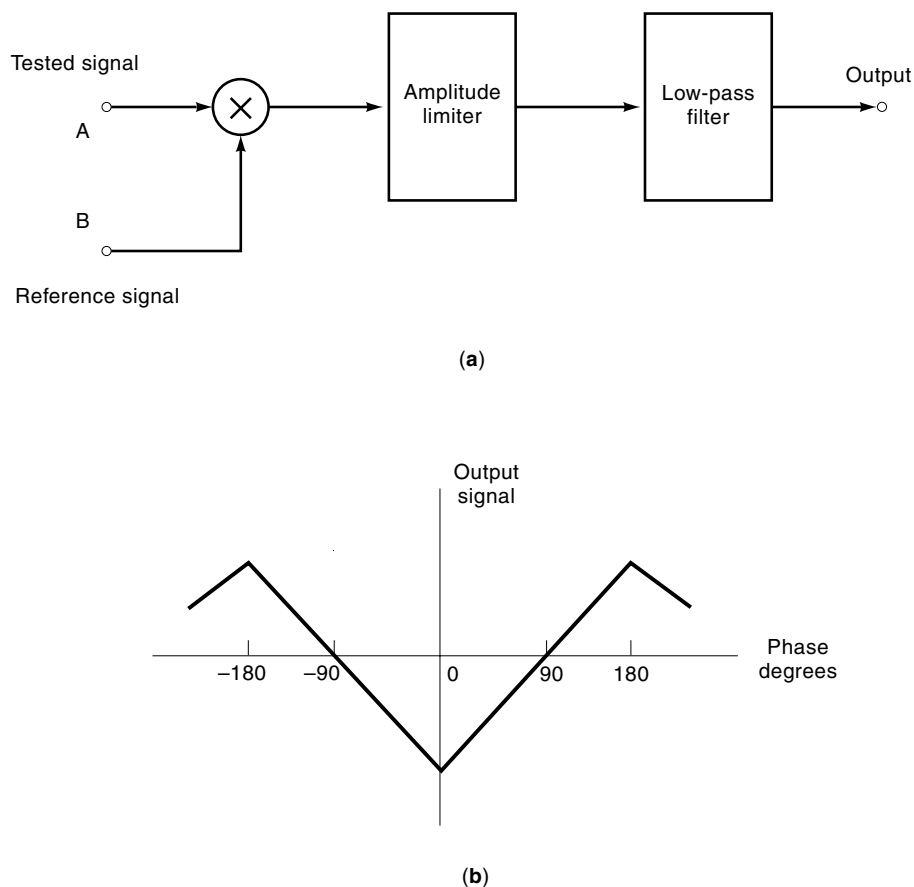


Figure 8. Four-quadrant analog phase detector. (a) Diagram. (b) Static transfer characteristic.

The measurement requires the display of a minimum of one cycle of the reference signal. The data acquisition time necessary to capture one period is usually larger than in coherent sampling units. The internal computer, together with the measurement/storage unit, can produce the calculated result of the phase shift. In automatic measurement mode, the phase is referred to as 50% of the transition level. A wave processor, also shown in Fig. 16, reconstructs the waves and displays them at its own rate. Processes of acquisition and wave reconstruction take a long time, but the enhanced signal processing capability of the system allows us to automate the measurement of phase shifts and other signal parameters.

EXAMPLES OF PHASE METERS AND OTHER PHASE MEASUREMENT UNITS

In this section, several examples of commercial phase measuring devices are discussed. Two typical stand-alone phase meters, a phase meter of a more advanced network analyzer, and a phase meter applied in a high-frequency impedance meter are included.

Example of Low-Frequency Direct Phase Meters

The first two units shown in Figs. 17 and 18 are two classical stand-alone phase meters designed by Feedback Instruments, Inc. (UK) (15). Both analog and digital meters have the same analog signal processors. In the analog meter the processor output signal which is proportional to the phase shift between the

input signals under test is filtered and directly measured by a dc meter. In the digital meter, the processor output signal is converted into its digital equivalent form and displayed. The meters use a smart compensation technique to correct errors due to input comparator or input wave offsets which are critical when the input signals have different amplitudes. The mark/space ratio-adjusting circuit fulfills this function. The analog meter operates up to 1 MHz, with a resolution of 1° and inaccuracy less than 2° . The digital meter operating within the same frequency range has a resolution and mid-frequency error of 0.1° .

Example of a Phase Meter Applied in Network Analyzers

Figure 19 shows the simplified block diagram of a typical network analyzer. The analyzer measures magnitude and phase

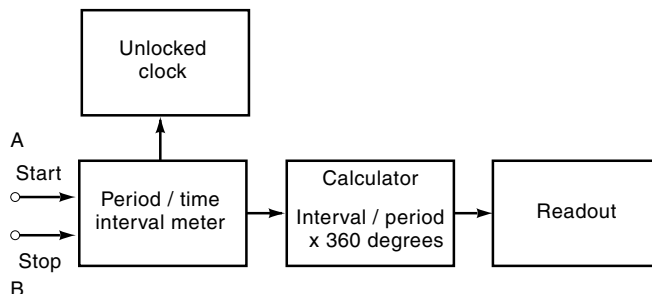


Figure 9. Phase-to-output variable converter with time interval meter.

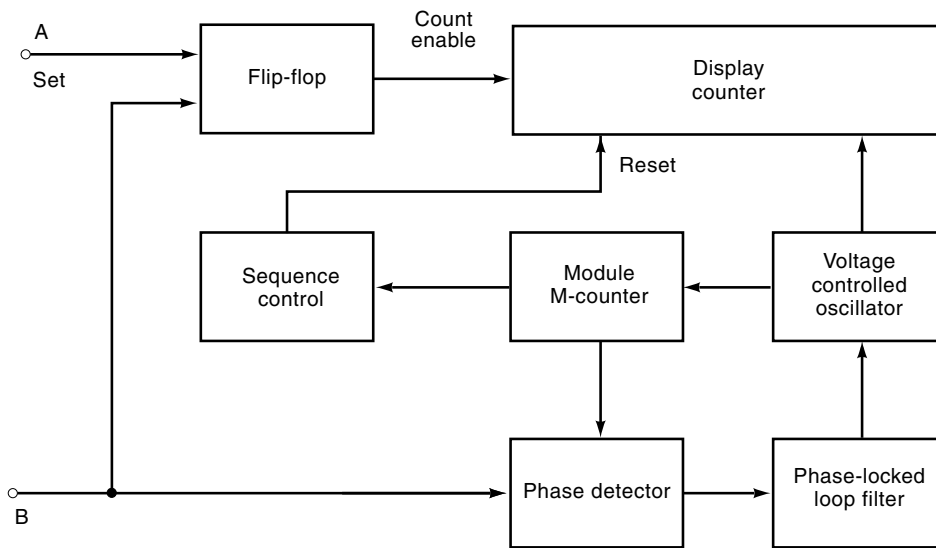


Figure 10. Counting phase-to-output variable converter with a frequency multiplier.

responses of high-frequency networks. The signals can be acquired by three receivers which have double frequency down conversion. After conversion the signals can be processed by analog quadrature detectors and split into real and imaginary parts to calculate phases and magnitudes after analog to digital conversion (Fig. 20). In some other solutions the signals can be sampled after frequency down-conversion, then converted from analog to their digital equivalents. The output signals of the analog-to-digital converters contain both amplitude and phase information. In the receivers, digital quadrature detectors are used to generate real and imaginary parts of the converted input signals. The signals are then filtered and processed by the fast math processor to yield magnitude and phase (Fig. 21). Such a solution is used in the HP 3577A Network Analyzer (9).

Example of Phase Meters Applied in Wave Analyzers

Many wave analyzers, including some modern digital oscilloscopes, combine features of coherent samplers and digital signal processors. Figure 22 shows a two-channel signal analyzer in which time transformed waves in an analog part of the analyzer are converted into digital signals and then processed in the digital processor. The results of processing, in terms of

wave parameters, like phase, magnitude, average value, or others are displayed. This type of circuit is used in the HP 71500A Microwave Transition Analyzer (16), and it is also applied in sampling digitizing oscilloscopes (13).

The details of another solution applied in wave analyzers and vector voltmeters involves the sampling technique and is described in Coherent Sampling in this article can be found in (10).

Example of a Phase Meter Applied in an Impedance Meter

Figure 23 shows a block diagram of a vector impedance meter. Two high-frequency signals, one proportional to the voltage across impedance, the other proportional to the current of the impedance, are converted to lower, intermediate frequency by means of coherent sampling. The coherent sampling process is controlled by a phase-locked-loop circuit supported by a search loop acting when the phase-locked-loop is not locked. The intermediate-frequency signals are processed by a low-frequency pulse phase detector to determine the phase. A voltage magnitude signal after demodulation defines the impedance magnitude.

ERRORS IN PHASE METERS

The errors affecting the accuracy of direct phase measurements originate in analog and digital parts of different phase measuring devices described before. The phase measuring devices have two or more channels. Misalignment of the channels, interference between adjacent channels, differences in delays, threshold levels, gains, differences in hysteresis loops of comparators, and noise introduce major errors. If analog readout is used the reading error is added to the previous errors. In digital time interval meter errors also involve the time base uncertainty and the count error (17,18,19). Frequency down-conversion introduces additional phase errors due to differences in converter channels. Low-frequency digital phase meters operating without built-in references can have their maximum errors below 0.1° maintaining very fine resolution, well below 0.1° (17).

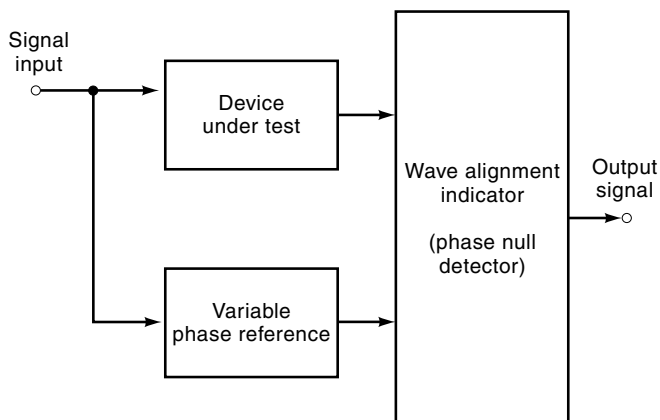


Figure 11. General block diagram of a referred direct phase meter.

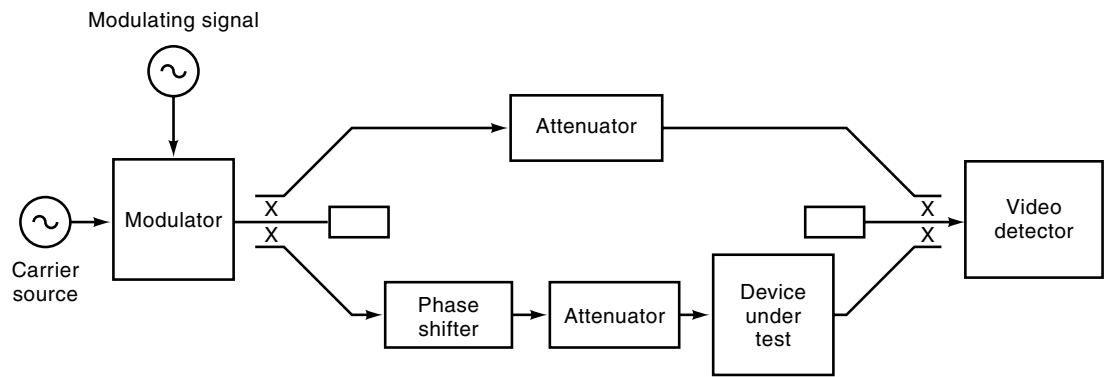


Figure 12. Simplified diagram of a referred complex transmission coefficient bridge.

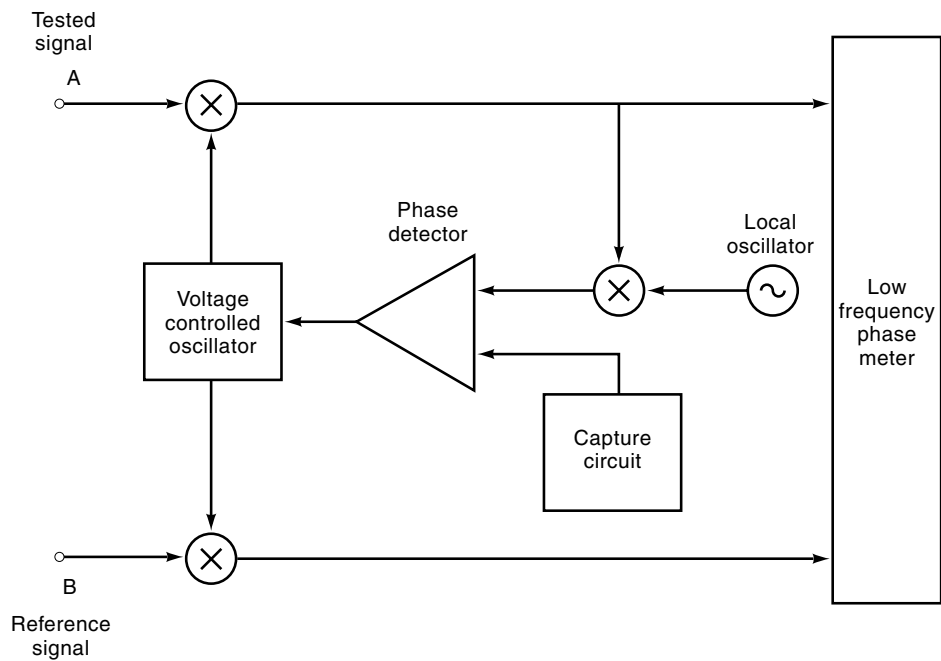


Figure 13. Simplified block diagram of a heterodyne phase meter.

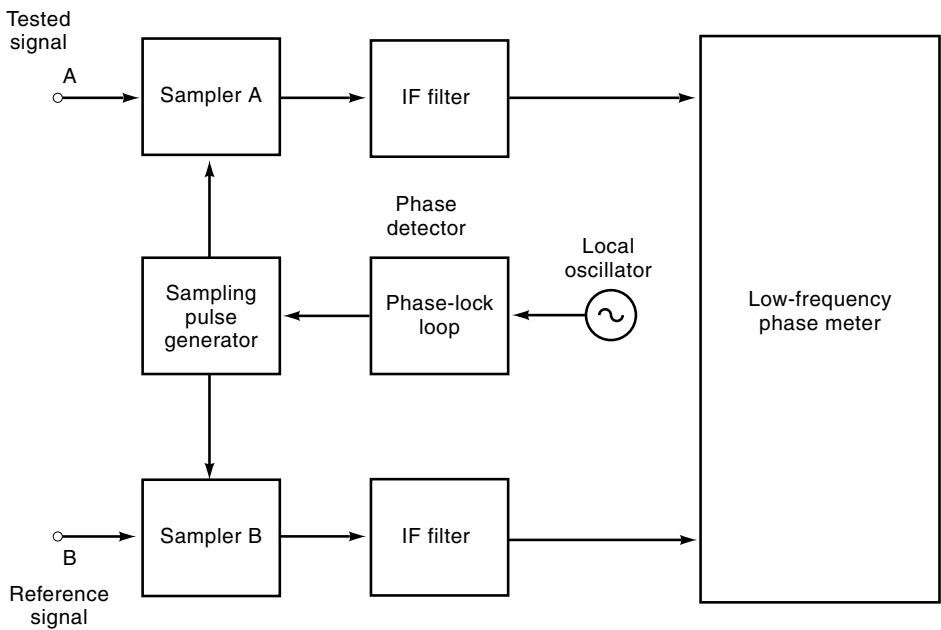


Figure 14. Simplified block diagram of a phase meter with a phase-lock loop coherent sampler.

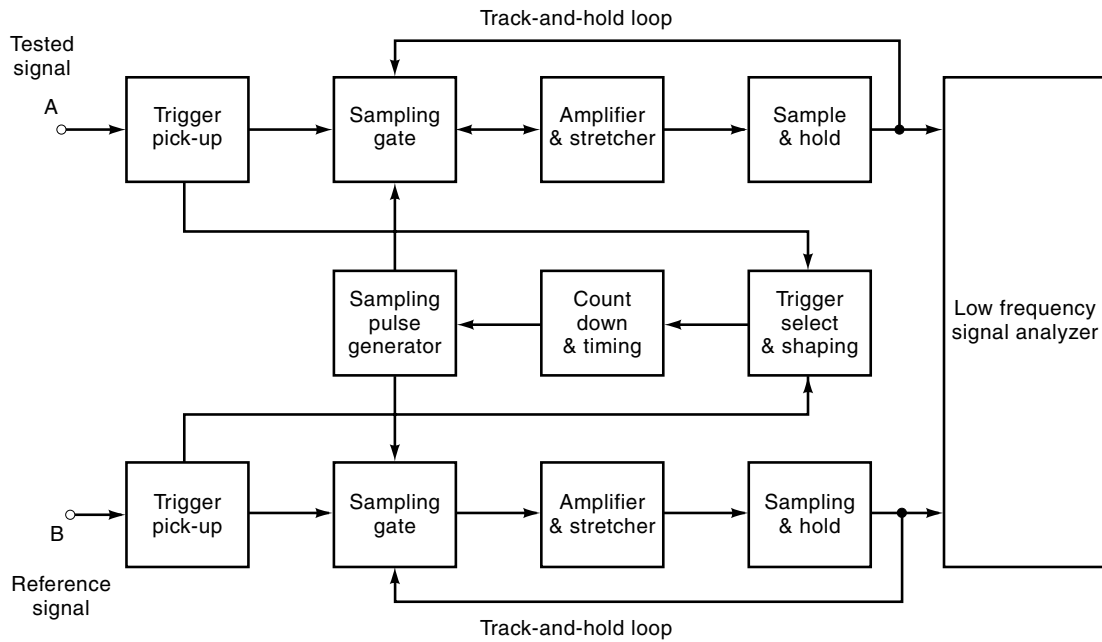


Figure 15. Simplified block diagram of a coherent sampler with a time base.

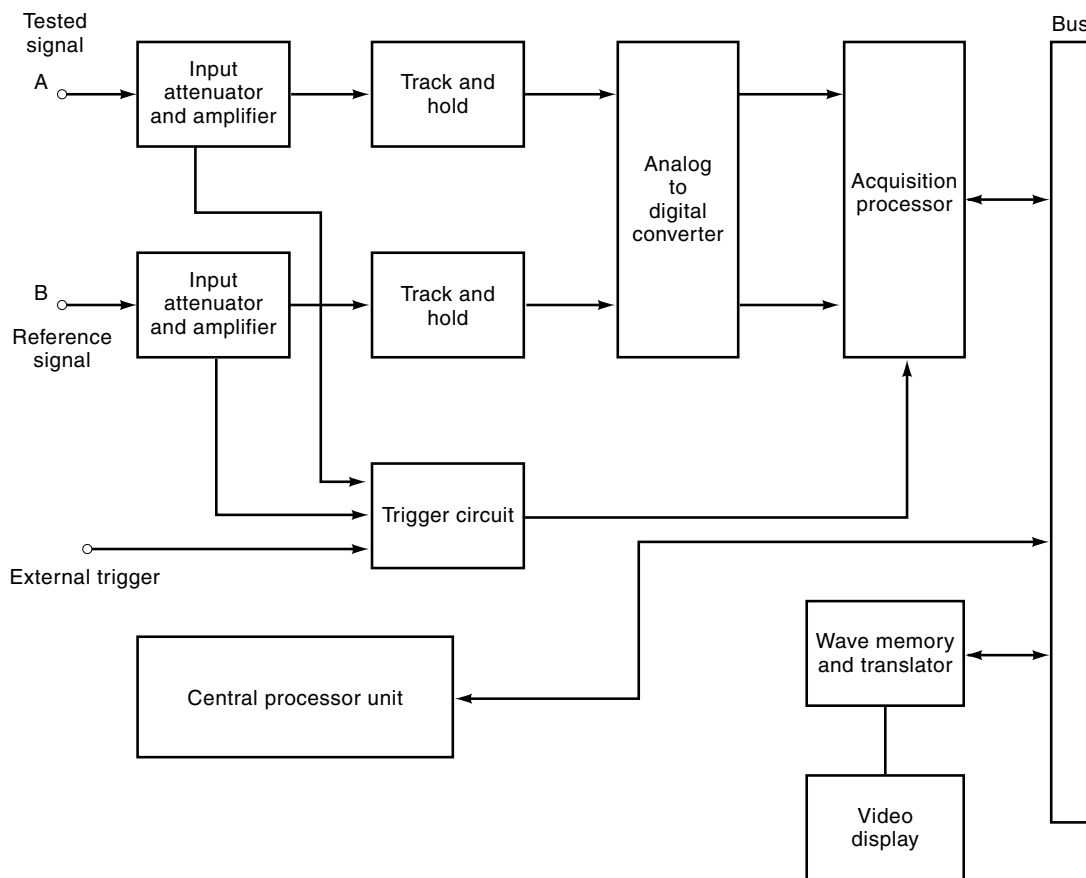


Figure 16. Simplified block diagram of a random repetitive sampling scope.

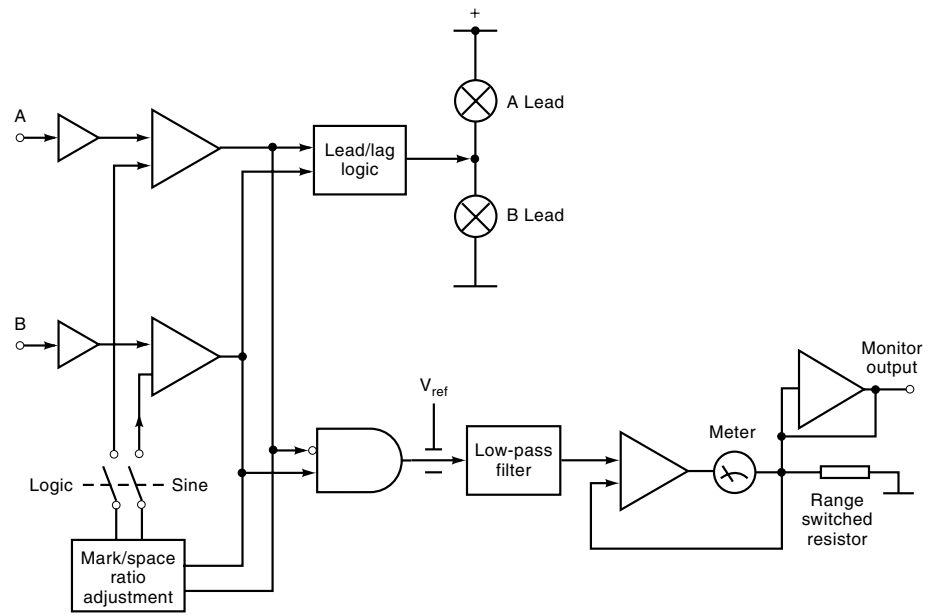


Figure 17. Simplified block diagram of Feedback Instruments Limited analog phase meter, APM 612 (Courtesy Feedback Instruments Incorporated).

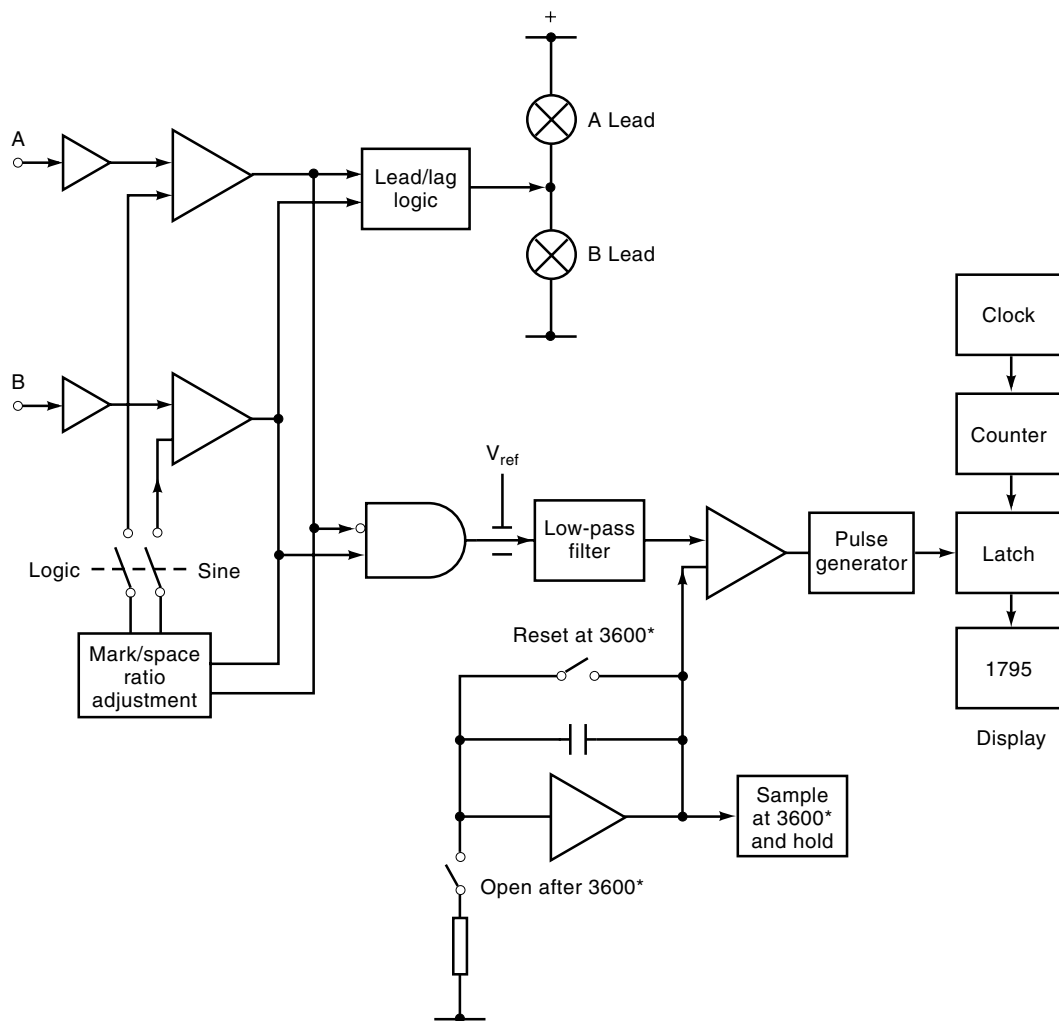


Figure 18. Simplified block diagram of Feedback Instruments Limited digital phase meter, DPM 609 (Courtesy Feedback Instruments Incorporated).

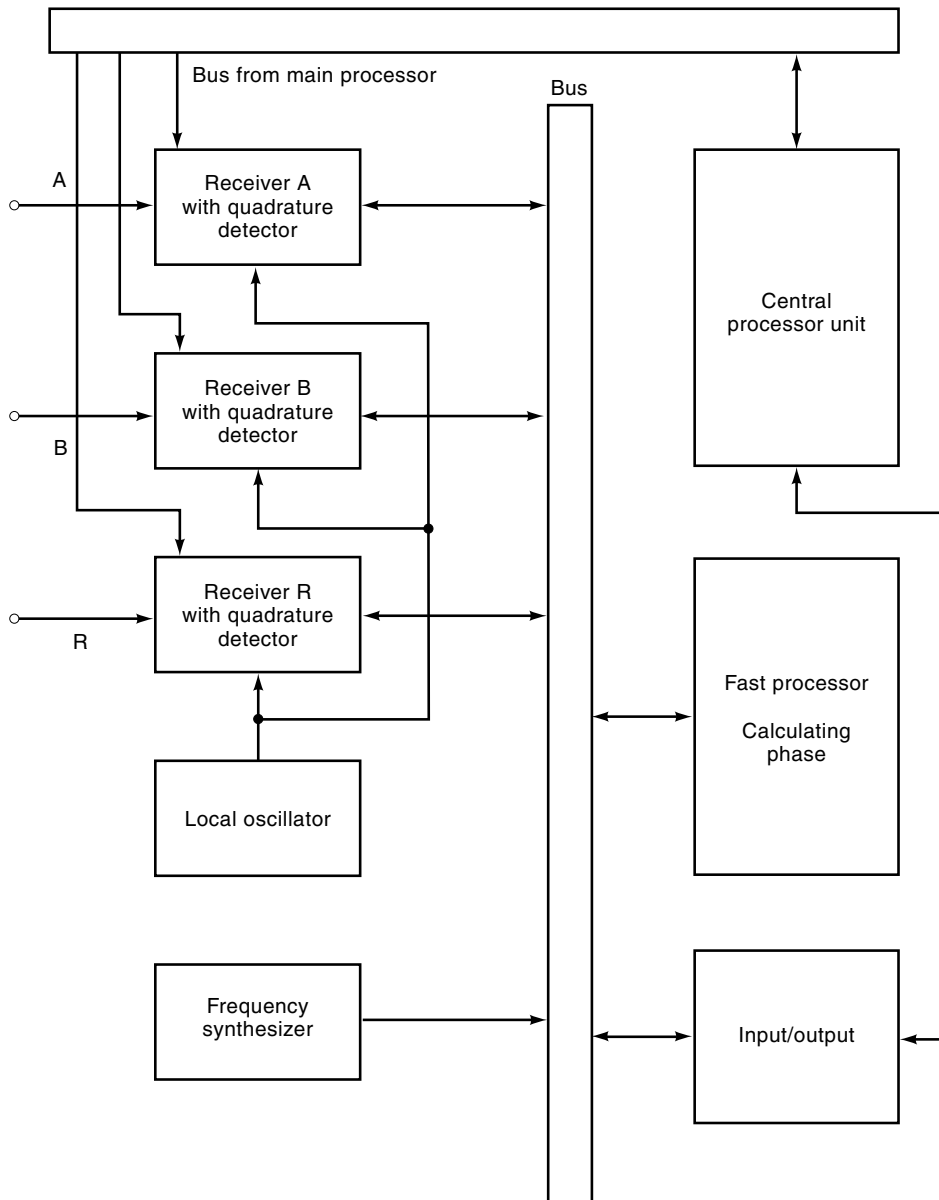


Figure 19. Simplified block diagram of a network analyzer with receivers and quadrature detectors.

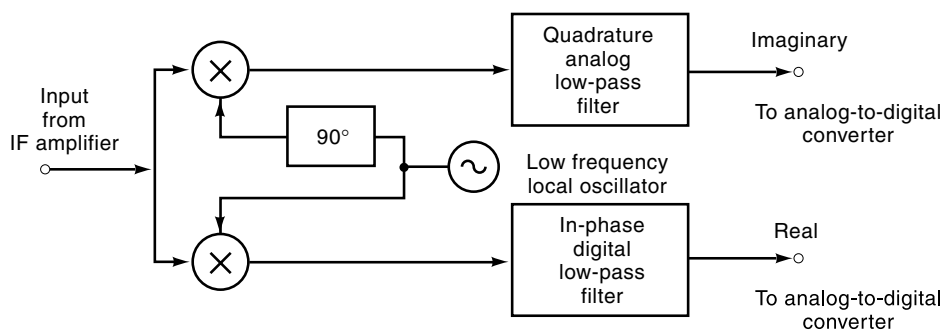


Figure 20. Quadrature analog detector recovering real and imaginary signal components.

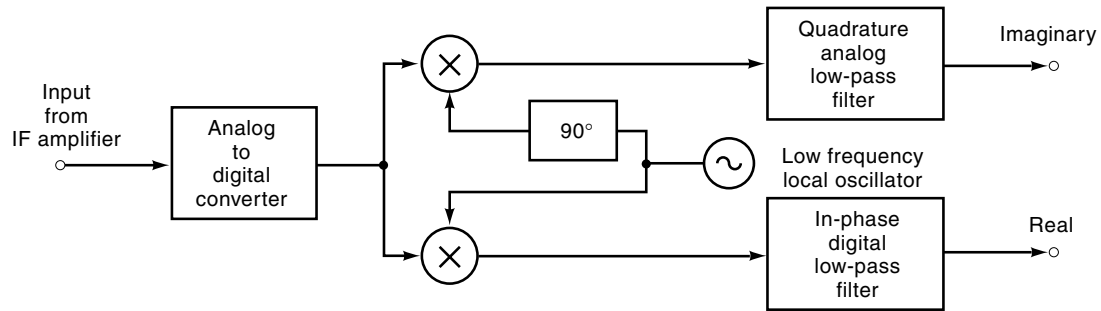


Figure 21. Quadrature digital detector recovering real and imaginary signal components.

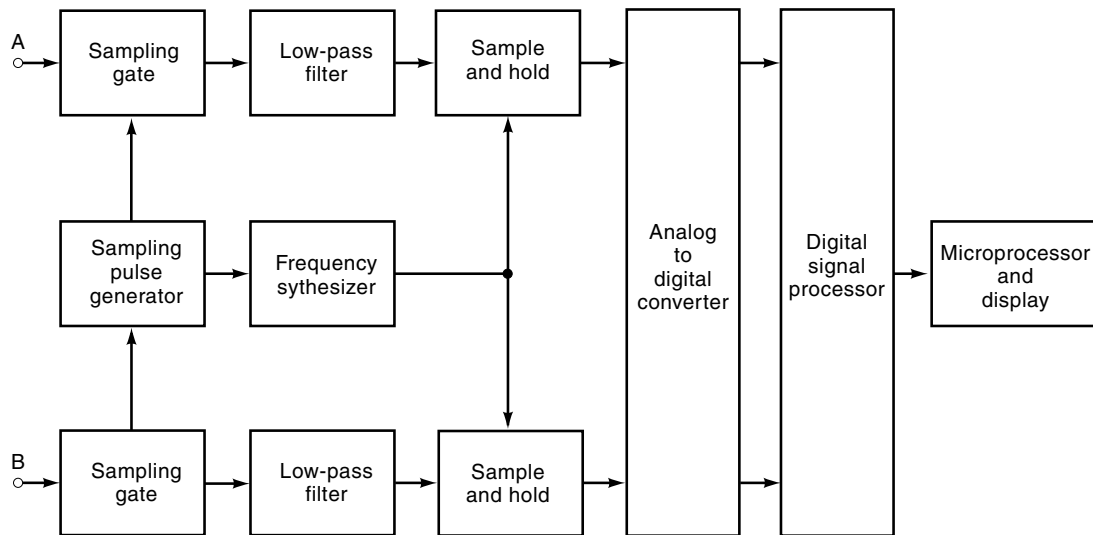


Figure 22. Simplified block diagram of a transition wave analyzer.

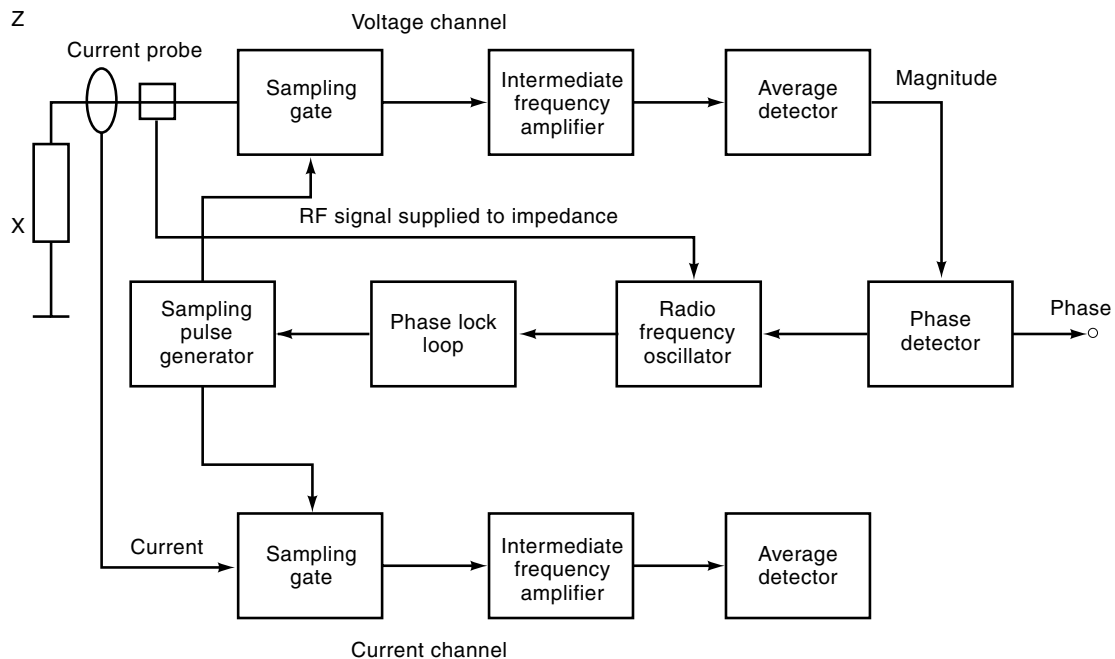


Figure 23. Simplified block diagram of an impedance meter with a phase meter unit.

Calibration of low-frequency phase meters requires application of high precision, digitally synthesized variable phase generators whose estimated uncertainties do not exceed 20 millidegrees (20). In microwave and millimeter wave regions, air-filled transmission lines of calibrated lengths help achieve significant accuracy. The length of such lines can be measured with laser interferometers yielding errors below 0.01° (17).

BIBLIOGRAPHY

1. L. Kaliouby, A new transformed Smith chart for real-time impedance measurement using the six-port concept, *IEEE Trans. Instrum. Meas.*, **36** (2): June 1987.
2. C. A. Hoer, The six-port coupler: a new approach to measuring voltage, current, power, impedance, and phase, *IEEE Trans. Microw. Theory Tech.*, **MTT-27**: November 1972.
3. K. T. Czarnecki, Wide-band homodyne method of measuring microwave circuits, *IEEE Trans. Instrum. Meas.*, **IM-34** (4): December 1985.
4. HP 54657A, HP 54658A. and HP 54659B Measurement/Storage Modules, User's Guide.
5. A. Dziadowiec, M. Lescure, and J. Boucher, A heterodyne low-level signal phase meter operating over 1 MHz to 300 MHz, *IEEE Trans. Instrum. Meas.*, **IM-33** (1): March 1984.
6. M. Fawzy Wagdy and M. S. P. Lucas, Errors in sampled data phase measurement, *IEEE Trans. Instrum. Meas.*, **IM-34** (4): December 1985.
7. Hughes Millimeter-Wave Products for 1987/88.
8. HP 531331A/132 Universal Counter, Operating Guide.
9. R. A. Witte and J. W. Daniels, An advanced 5 Hz to 200 MHz network analyzer, *HP Journal*, November 1984.
10. J. R. Zellers, An economical network analyzer for the 4 to 1300 MHz frequency range, *HP Journal*, October 1992.
11. HP 54600A and HP 54601A Oscilloscopes, User and Service Guide.
12. K. Rush and D. J. Oldfield, A data acquisition for a 1 GHz digitizing oscilloscopes, *HP Journal*, April 1986.
13. R. A. Witte, Low-cost, 100 MHz digitizing oscilloscope, *HP Journal*, February 1992.
14. Measurement Product Catalog, Beaverton, OR: Tektronix, 1996.
15. Digital Phase Meter, DPM 609, and Analog Phase Meter, APM 612, Feedback Instruments, Inc., Operating Manuals.
16. D. J. Ballo and J. A. Wendler, The microwave transition analyzer: a new instrument architecture for component and signal analysis, *HP Journal*, October 1992.
17. P. I. Somlo and G. W. Small, Phase-shift measurements in a standard laboratory environment, *Proc. IEEE*, January 1986.
18. Fundamentals of Microwave Frequency Counters, Hewlett Packard Application Note, AN 200-1.
19. Fundamentals of Time Interval Measurements, Hewlett Packard Application Note, AN 200-3.
20. B. A. Bell, Standards for waveform metrology based on digital techniques, *J. Res. NIST*, **95** (4): July–August 1990.

ANDREW RUSEK
Oakland University

TIME MEASUREMENT

The function of a time interval measurement device is to produce a quantitative measure of the length of a time interval at

the input of the device. Typically this interval is presented by short, logic level start and stop pulses that have a relation to some physical phenomena whose time behavior is to be studied by the device. Its output is given either by an analog quantity such as a voltage change in a capacitor or as a digital number, an output of a counter, for example (Fig. 1). Typically the final output of the device, independent of the method that is being used, is characterized by a number that gives the length of the time interval as a multiple of the least resolvable time interval of the device. This can be achieved by an analog-to-digital converter (ADC) in case of analog realizations, for example. The least resolvable time interval characterizing the measurement system being used can be considered the value of its least significant bit (LSB) analogically to ADCs. It defines the resolution of the device. Accordingly, a time-interval measurement device is often called a time-to-digital converter (TDC).

Time-to-digital converters have manifold applications. They are one of the critical components of a pulsed time-of-flight laser radar, for example, where the distance measurement accuracy is directly dependent on the time measurement accuracy. Pulsed time-of-flight laser radars can be used in geodesy and also in various industrial inspection applications (1,2).

Time interval counters can also be found in applications of time and frequency comparison. For example, each Global Positioning System (GPS) time transfer receiver uses a time interval counter to determine the time difference between the atomic clock on board a GPS satellite and a local clock. In frequency calibration applications, the time interval measurements obtained from an oscillator being tested and a reference oscillator are used to characterize the performance (such as the phase offset, frequency offset, and frequency drift of the oscillator being tested with respect to the reference oscillator (38,39,40).

Another application area for TDCs is found in nuclear science, for example, dE/E , time-of-flight, and positron annihilation studies (3,4). TDCs are also used in the calibration of test equipment for electronic circuits. Since in high-speed large-scale integration (LSI) circuit propagation delays are reduced to 100 ps or even below, high speed and accuracy are required in automatic test equipment systems (ATE). Test timing control with an accuracy to the order of tens of picoseconds is needed and, to maintain such timing performance, the calibration hardware requires even higher-performance TDCs with picosecond resolution and accuracy (5,6).

The most straightforward method to realize a time-to-digital converter is to have an electronic counter counting the pulses of an accurate oscillator during the start–stop time interval. The resolution of this method, as shown later, is defined by the period of the oscillator. This means that a resolution better than in the range of nanoseconds is difficult to achieve due to difficulties in the realization of high-frequency, high-performance oscillators. The purpose of this article is to describe this time-interval measurement method with its variations where its fundamental resolution limitation is overcome with some kind of

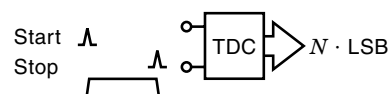


Figure 1. Input–output interfaces of a time-to-digital converter (TDC). LSB is the least significant bit, for example, 10 ps to 10 ns. $N \times \text{LSB}$ is the length of the start–stop interval.

interpolation and where accordingly picosecond-range time-interval measurement resolution can be achieved. Some of these interpolator structures, such as analog-to-digital converters or digital delay lines, can be used as stand-alone time-interval measurement units, too.

In the following sections, first some important performance parameters of a time-interval measurement device are given. Then the basic counting method is analyzed in more detail. After this, interpolation methods to increase the precision and accuracy of the counting method are described. Next the performance of one interpolation technique, the Nutt method, is analyzed in more detail. This method is selected as it is perhaps the most popular technique in high-performance time-to-digital conversion with many practical realizations. As part of the presentation some practical realization aspects are also discussed.

PERFORMANCE PARAMETERS IN TIME-INTERVAL MEASUREMENT

The main performance parameters used in connection with the time-interval measurement method are resolution, precision, and accuracy.

The term *resolution* is used for the smallest time interval that can theoretically be resolved by the TDC in a single measurement, that is, the quantization step (LSB). The term (single-shot) *precision* is used for the standard deviation (σ) of the measurement result distribution around the mean value (m) when a single time interval is measured repeatedly (Fig. 2). In a practical measurement, the single-shot precision is influenced, besides the quantization error, by nonidealities like jitter in timing signals and power-supply noise and can be used to estimate the smallest real time interval that can be resolved. Since precision is limited by quantization error and statistical error sources, it can usually be improved by averaging. (7).

While precision gives the statistical variation of the measurement result around a mean value, single-shot accuracy is affected both by this statistical variation and any systematic errors in the mean value. The statistical variation (σ) can be reduced by averaging but systematic errors cannot, that is, the precision of the average may be good but the accuracy may still be poor. Systematic errors include, for example, linearity and stability errors. Integral linearity error (INL) is the deviation of the input-output characteristics from the ideal, straight-line input-output characteristics (Fig. 2). Differential linearity

error (DNL) is the deviation of a single quantization step from the ideal value of 1 LSB. Stability of a TDC is defined as the sensitivity of its characteristics with temperature, supply voltage, time, and the like. Clearly, to measure the systematic errors reliably, the statistical variation should be reduced to a negligible level by averaging.

Other parameters that should be considered while choosing time-measurement techniques are, for example, the measurement range of a TDC and the conversion time. The range of the device defines the maximum time interval than can be measured or digitized. Conversion time is the time between the end mark of the input time interval and the moment when the measurement result is ready. Sometimes, especially with analog TDCs, a parameter called dead time that typically defines the period during which the system is incapable of accepting a new start after a registered start signal (conversion time plus possible recovery time) is also used. Dead time might be an important parameter especially in applications where the time intervals to be measured are randomly distributed, which often happens in nuclear measurements. For example, if the start detector efficiency is high and the stop efficiency low, respectively, a long dead time may result in reduced measurement efficiency since many of the start signals that trigger the time-interval measurement may not be followed by a valid stop signal. In this case it is important for the TDC to recover from the useless start pulse as quickly as possible (8).

COUNTING METHOD

Analysis of the Method

The counting method where clock pulses are counted during the input time interval is perhaps the most simple time measurement method (Fig. 3). Provided that the reference clock is accurate, a crystal oscillator, for example, the counter method has a wide linear range and good stability. In asynchronous measurement, that is, when the measurement begins in a random phase with respect to the clock, the maximum error in one measurement is $\pm T_{\text{clk}}$, where T_{clk} is the clock period. For each input interval the counter will measure either N or $N + 1$ counts, and for this binomial distribution it can be shown that for an input interval of $(Q + F) T_{\text{clk}}$, where Q is an integer and $0 \leq F < 1$, the expected value of the measurement result is $Q + F$ and the standard deviation is (9)

$$\sigma = \sqrt{F(1 - F)} \quad (1)$$

Thus, the measurement precision varies with the input interval and the worst case value for $F = 0.5$ is $0.5 T_{\text{clk}}$. For a 1 GHz clock, for example, the maximum single-shot error is ± 1 ns and the worst case value for σ is 500 ps.

To improve precision, averaging can be used in applications where several samples per measurement can be taken. In asynchronous measurement and for N_{av} samples the precision is (9)

$$\sigma = \frac{\sqrt{F(1 - F)}}{\sqrt{N_{\text{av}}}} \quad (2)$$

which is shown in Fig. 4.

In asynchronous measurement the phase of the input interval with respect to the clock is continuously distributed over

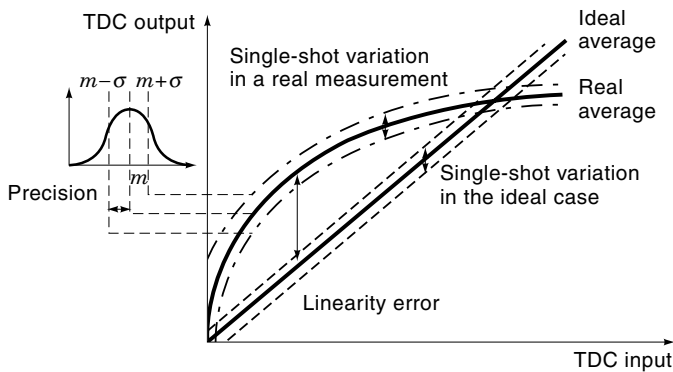


Figure 2. Performance parameters of a time-to-digital converter.

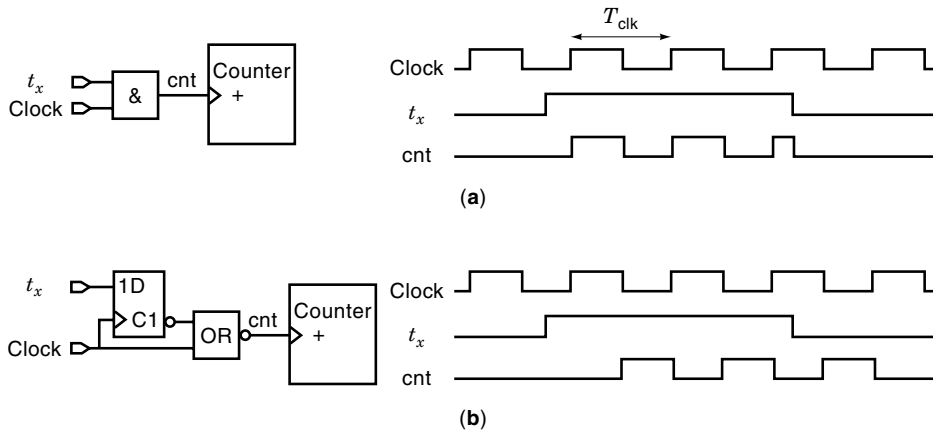


Figure 3. Basic counting method, where clock pulses are enabled to increment the counter during the input time interval. The gating logic of (b) is preferable to that in (a) because it allows only integral clock pulses to reach the counter. The method in (a) can produce short pulses that may or may not be detected by the counter.

the clock period or even several periods. Synchronous measurement is also possible. This can be realized if the start timing signal is repeated at a constant rate and if the oscillator of the system is locked to this rate. However, if the measurement is completely synchronous with respect to the clock, averaging does not improve precision. On the other hand, if the phase of the synchronous measurement can be controlled, it is possible to achieve a faster precision improvement rate than in asynchronous measurement. For example, if the phase of the start signal and oscillator has M discrete values evenly distributed within the clock period, the measurement precision can be improved in averaging with a rate proportional to $1/M$ (rather than to $1/\sqrt{M}$ as in synchronous measurement) but the improvement is limited by factor $1/M$. The synchronous method utilizing discrete phase steps is, however, more difficult to realize than the asynchronous method (9).

The basic counting method has good accuracy, since with a stable crystal oscillator reference the linearity and both the short-term and long-term stabilities are good. Measurement range can be increased simply by increasing the number of bits in the counter. With standard commercially available high-performance crystal oscillators the maximum frequency is limited to a few hundred megahertz, which corresponds to a single-shot resolution of several nanoseconds (10). With advanced clock-generation techniques single-shot resolutions of about 500 ps to 1 ns have been achieved. Here the idea typically is to multiply the clock frequency of a reference source by a

phase-locked loop which includes a divider and a voltage-controlled oscillator (VCO) in its loop. In integrated time-to-digital converter realizations a VCO based on a ring oscillator is especially attractive, since the measurement resolution can be further improved by utilizing all the clock phases of the ring oscillator (3,11).

Input Synchronization

In the counting method, the gating of the digitization clock may have a great impact on the measurement accuracy. If the input interval (asynchronous with respect to the clock) is directly used for gating the counter, short clock pulses appear frequently at the input of the counter [Fig (3a)]. These may or may not be long enough to increment the counter and cause an unpredictable error in the averaged result. Therefore, a synchronizer, with which the input interval is synchronized to the clock, is needed. An example is shown in Fig. 3(b), where only full-size clock pulses are allowed to the counter. It should be noted that even now there is a small possibility for a counting error if the setup time of the flip-flop is not fulfilled, which increases the propagation delay of the flip-flop and shortens the clock pulse. However, the probability to lose an entire clock pulse is small. More advanced synchronizing methods are presented in later sections.

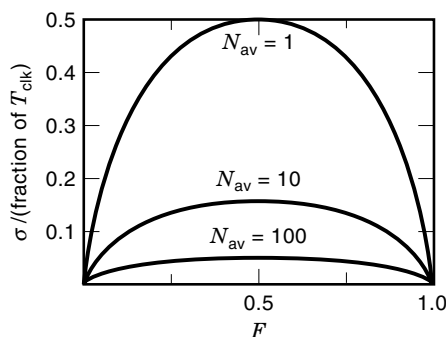
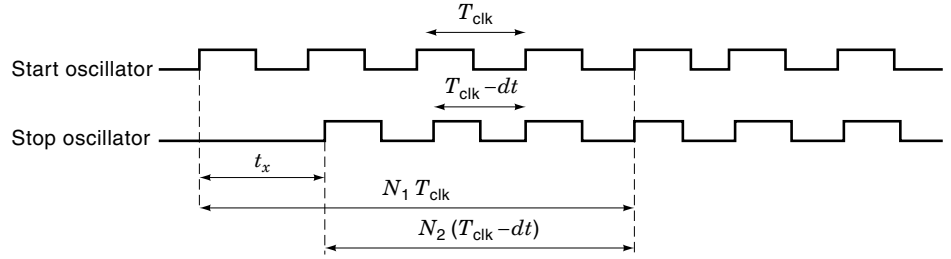


Figure 4. Precision of the counting method as a function of the fractional part F (by which the length of a time interval exceeds an integral number of clock periods) and number of averaged samples N_{av} .

INCREASING PRECISION AND ACCURACY

The precision of the counting method can be improved by averaging but at the cost of measurement time. In some applications, however, averaging is not possible due to the single-shot nature of the measured phenomena. Typically this is the case in the experiments made in nuclear physics. The aim of this section is to discuss the variations of the basic counting method that enable one to achieve an improved single-shot precision with realistic clock rates. These methods typically utilize accurate clock generators in connection with other time-interval measurement techniques, which enable a resolution markedly better than the clock period to be achieved. In a way these methods digitize the basic clock period so that the resolution of this digitization (interpolation) defines the system resolution.

Figure 5. Principle of the vernier method in time measurement. Two oscillators with different periods are triggered by the timing pulses. The counting is terminated at coincidence. The time interval t_x is effectively digitized with a resolution defined by the period difference.



Vernier Method

In the vernier method (12), two startable oscillators with slightly different frequencies are used to achieve an LSB equal to the difference of the oscillator periods (dt in Fig. 5). The start mark of the input interval enables the oscillator with a lower frequency $f_1 = 1/T_{\text{clk}}$, and the stop mark enables the oscillator with a higher frequency $f_2 = 1/(T_{\text{clk}} - dt)$. A counter records the pulses from the oscillators until, since $f_1 < f_2$, at some point the two oscillators will be in phase. Then, from Fig. 5, $t_x = (N_1 - N_2) T_{\text{clk}} + N_2 dt$. If $t_x < T_{\text{clk}}$, $N_1 = N_2 = N$ and $t_x = Ndt$. The maximum value of N_2 is T_{clk}/dt so the maximum conversion time (time from the stop mark to the coincidence) is $(T_{\text{clk}}/dt)(T_{\text{clk}} - dt)$. To measure input intervals longer than T_{clk} , two counters recording both N_1 and N_2 are needed; see Ref. (12).

Functionally the vernier method can also be considered a time-interval multiplier where the multiplication factor is T_{clk}/dt . When the multiplied time interval is digitized with a clock having a period T_{clk} , the effective resolution is equal to dt . The accuracy and the resolution of this method can be high if f_1 and f_2 are stable and dt is made small (typically 1% of the period).

In the dual vernier method (Fig. 6), the startable start and stop oscillators have an equal frequency $f_0 = 1/(T_{\text{clk}} + dt)$, and their phase crossover points with a third, continuously running oscillator with a frequency $f_{\text{clk}} = 1/T_{\text{clk}}$ are detected (13). Similar to the basic vernier method of Fig. 5, the LSB of the measurement is dt and the maximum conversion time is $(T_{\text{clk}}/dt)(T_{\text{clk}} + dt)$. Using the three counter values N_c , N_1 , and N_2 , the input interval can be calculated from

$$t_x = N_c T_{\text{clk}} + N_1(T_{\text{clk}} + dt) - N_2(T_{\text{clk}} + dt) \quad (3)$$

Good single-shot resolution can be achieved with the vernier techniques. In Ref. 13 the LSB is 20 ps with a clock frequency of 200 MHz, and in Ref. 6 the measured standard deviation is

2.3 ps in a 2 ns range with a clock frequency of 500 MHz. The time-measurement circuitry in Ref. 13 is a discrete implementation for a commercial counter, while the TDC in Ref. 6 is an integrated implementation but the start and stop clock sources are external. This circuit is used for tester timing calibration, where the start input is the reference clock and the stop input is a tester output.

Nutt Method

Analysis of the Method. A powerful method for measuring time intervals is to combine an accurate digital clock and an analog or digital interpolation circuit as shown in Fig. 7 (14). The input time interval is roughly digitized by counting the reference clock periods during this interval. The counter is enabled at the first clock pulse following the start mark, and disabled at the first clock pulse following the stop mark. The resulting time interval T_{12} is synchronized to the clock and is, therefore, accurately measured. The fractions T_1 and T_2 are digitized separately with interpolators to improve single-shot resolution. For an n -bit interpolator the LSB of the measurement is equal to $T_{\text{clk}}/2^n$ and the input interval t_x can be calculated from the following equation:

$$t_x = T_{12} + T_1 - T_2 = N_c T_{\text{clk}} + N_1 T_{\text{clk}}/2^n - N_2 T_{\text{clk}}/2^n \quad (4)$$

Note that if the system clock is asynchronous with respect to the time intervals to be measured, the length of time fractions T_1 and T_2 change randomly in a repeated measurement although their difference has only two discrete values (dt if the main counter result is $N_c - 1$ or $dt - T_{\text{clk}}$ if the main counter result is N_c). This means that in the averaged results the nonlinearities of the interpolators are also averaged in a way so that the accuracy of the system is not limited by them. Note also that the drifts of the interpolators tend to cancel as it is the difference $T_1 - T_2$ that counts in the final result.

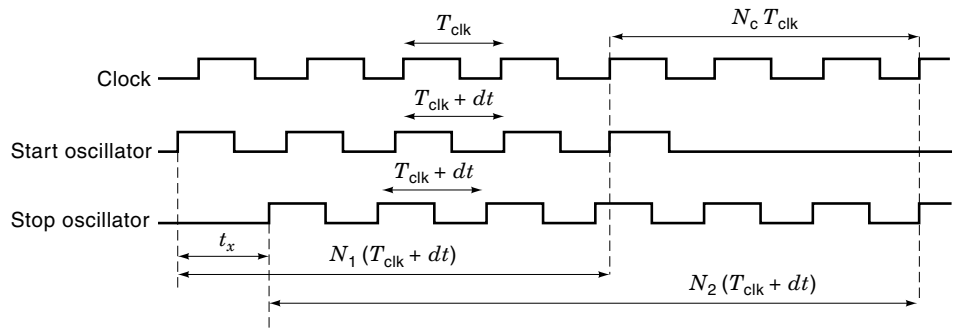


Figure 6. Principle of the dual vernier method in time measurement.

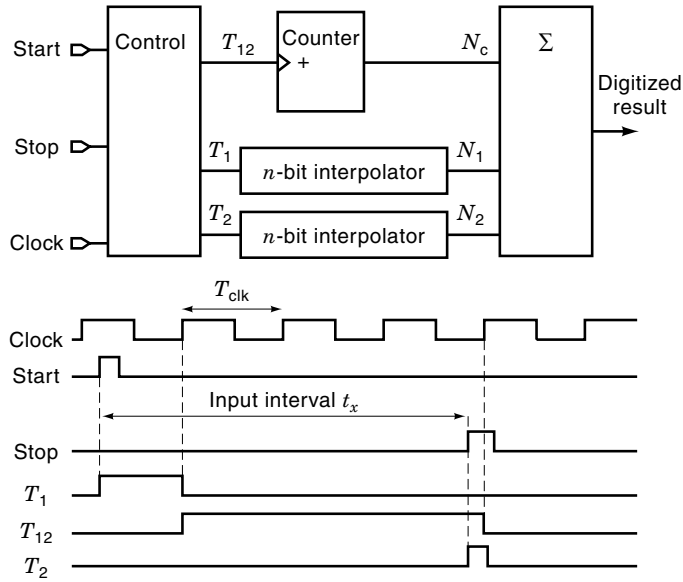


Figure 7. Block diagram and operating principle of a TDC based on the Nutt method. The start–stop time interval is digitized coarsely by the counter. In addition, time fractions from start and stop to the next following clock pulses, respectively, are digitized by interpolators.

The interpolators are generally based on analog time-to-voltage conversion or on digital delay lines. These structures can achieve good single-shot resolution in a limited dynamic range. Practically, they can be used for measuring time intervals from tens to some hundreds of nanoseconds with a resolution of 100 ps to 1 ns depending on the measurement range. The Nutt method can thus be considered a technique that combines the inherently good single-shot resolution of an analog time-interval measurement method such as time-to-amplitude conversion, for example, and the accuracy and wide linear range of the counting method.

Synchronization. For the Nutt method it is typical that the timing signals and the clock are deliberately asynchronous. This produces a synchronization problem in the generation of time intervals T_1 , T_{12} , and T_2 . In Fig. 8(a) is shown a simplified scheme of a control block of a TDC based on the Nutt method. The end mark of $T_1(T_2)$ is taken from the output of the flip-flop

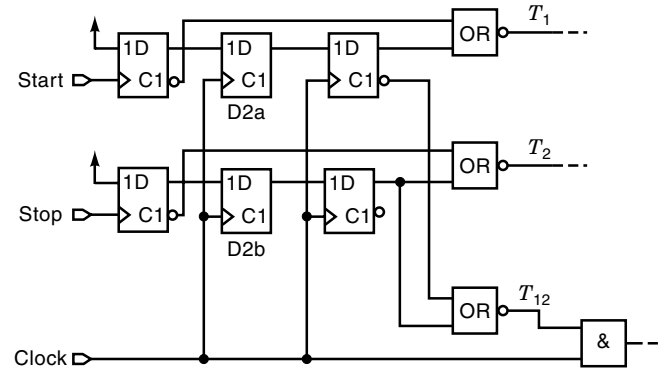


Figure 9. Improved synchronization of start and stop inputs in a TDC based on the Nutt method. Here the setup time violation in D2 does not produce error in T_1 and T_2 .

D2a (D2b). However, in asynchronous measurement the start (stop) pulse arrives at a random phase with respect to the reference clock. When the setup time requirement of the flip-flop D2a or D2b is not fulfilled, the propagation delay of that flip-flop will increase [Fig. (8b)] and in the extreme case the flip-flop will enter a metastable state. Thus, if a start (stop) pulse occurs near the rising clock edge, an erroneous measurement is possible. Unlike in the basic counting method, even a small excess delay affects the measurement result directly through T_1 and T_2 .

The probability of a synchronization error can be reduced by waiting before sampling the output of flip-flop D2a (D2b). In the scheme of Fig. 9 the measurement will be accurate if the flip-flop D2a (D2b) settles in less time than T_{clk} (15). This method reduces the number of errors very effectively (for a detailed analysis, see, e.g., Ref. 16, pp. 533–538). However, the linear measurement range required from the interpolators is doubled as compared to the scheme of Fig. 8(a).

The synchronization scheme of Fig. 8(a) can be used if averaging is used to improve resolution, and it is possible to detect and reject a single incorrect measurement. The scheme of Fig. 9 increases the measurement range required of the interpolators, which is a disadvantage in digital interpolators, where the integral nonlinearity due to delay element mismatch increases. In time-to-voltage conversion, a short time offset is an advantage, since the nonlinear part of the characteristics due to switching effects is then not used for measurement.

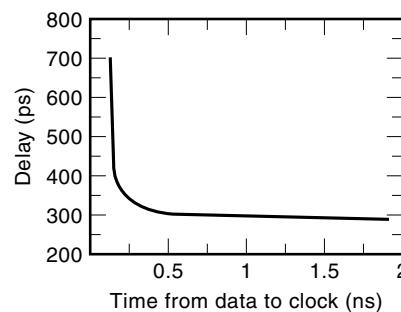
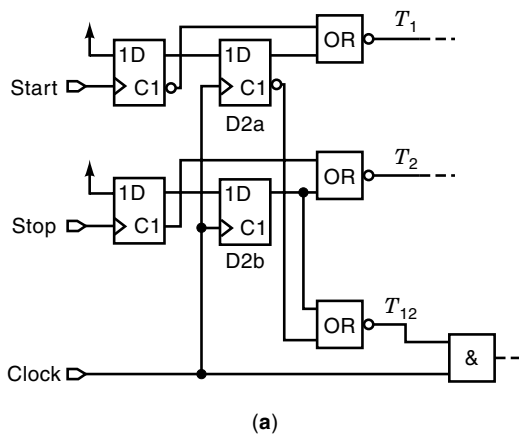


Figure 8. (a) Synchronization of start and stop inputs in a TDC based on the Nutt method. Problems may occur when the clock edge and output of first flip-flops occur within the setup time of D2. (b) Simulated propagation delay of a 0.8 μm current mode logic flip-flop.

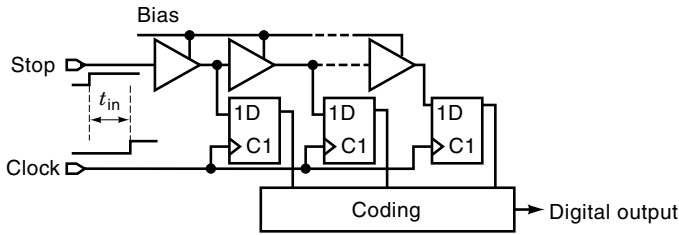


Figure 12. Delay line as a time-measurement unit. The output of the flip-flops is latched by the stop signal. The first flip-flop not set indicates the length of the start–stop interval.

Digital Interpolators. An example of a digital delay line is shown in Fig. 12. In a delay line, the time measurement unit is the propagation delay of a logic gate, usually of an inverter. In Fig. 12, the start mark of the input interval travels along the delay line. When the stop mark arrives, it stores the status of the delay line into the flip-flops. From these data, the input interval can be coded as a multiple of one gate delay. The delay of the element must be controllable to compensate for the effects of process variations and temperature and supply changes. As shown in Fig. 13, the control parameter can be the current (23,24) or the load of the element (5,25,26). The control voltages (p_{bias} , n_{bias} , bias) are commonly created in analog phase-locked loops or delay-locked control loops (23,25) but also digital control has been implemented (24,26).

However, if the propagation delay of a logic gate is directly used as the measurement unit, the achievable resolution is limited by the minimum gate delay, which, in turn, depends on the technology and operating temperature range. This limitation can be avoided by using the difference of two gate delays as the quantization step. The delay difference can be implemented with two delay lines having slightly different gate delays (27–30), but a more compact solution is the pulse-shrinking delay line presented in Ref. (31). In a pulse-shrinking delay line (Fig. 14) the time resolution depends on the difference between two delays of a single element. The input pulse t_{in} propagates in a chain of delay elements. The pulse shrinks by a constant amount in each element until it disappears entirely. The SR (set–reset) flip-flops are set by the propagating pulse until the pulse vanishes, after which the following flip-flops are left reset. The address of the first flip-flop not set is coded to the output. The propagation of the rising edge of the input pulse is slowed down by the current-starving transistor N_1 while the falling edge travels fast. Thus, in each element the pulse shrinks with an amount equal to the delay difference. The amount of

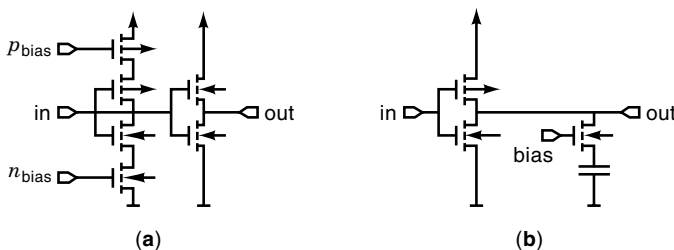


Figure 13. Controllable delay elements used in delay lines. In (a) the delay is controlled by adjusting the current of the inverter with the bias voltages n_{bias} and p_{bias} . In (b) the delay is controlled by varying the load seen by the inverter.

shrinking, that is, the LSB of the delay line is controlled by the bias voltage v_{bias} at the gate of the current starving transistor N_1 . V_{bias} is generated in the delay-locked loop, which stabilizes the length of the delay line to be equal to the reference clock period. Basically, the bias voltage is adjusted so that a pulse with a width of one clock period, which is equal to the maximum input time interval of the interpolator, just disappears in the last delay element.

Delay line interpolators based on inverter gates are easily implemented in a complementary metal oxide semiconductor (CMOS) process, have low power consumption, and can relatively easily be transferred to a low-supply-voltage environment (3,32). Also, the conversion time of a TDC based on delay-line interpolators is short, since the conversion time depends on the propagation delay of the delay element chain and the coding logic. The main factor limiting the performance is the nonlinearity caused by random mismatch, which determines the achievable single-shot resolution. LSB values below 100 ps are difficult to achieve. For example, with a clock frequency of 100 MHz, a 128-tap delay line is needed to achieve an LSB of 80 ps. The single-shot error due to the nonlinearity of the interpolators is then easily larger than 1 LSB, as can be seen from the measured nonlinearities in Refs. 3 and 33, which are 80 ps for a 32-tap delay line and 300 ps for a 100-tap delay line, respectively.

Note also that the delay-line method in which the difference of the two absolute delays defines the resolution is equivalent to the vernier method in which resolution was defined by the difference of the oscillator periods.

CHARACTERISTICS OF THE INTERPOLATION METHOD

Single-Shot Precision

Similarly to a counter, in asynchronous measurement (i.e., the start pulse arrives at a random phase with respect to the reference clock) the worst case single-shot precision of an ideal TDC based on the Nutt method is, according to Eq. (1), 0.5 LSB and the precision improves with averaging according to Eq. (2). The maximum single-shot quantization error is ± 1 LSB (if the interpolator results are truncated as is the case for delay lines and analog dual-slope converters) and the measurement result distribution is binomial (N or $N + 1$ counts). In practice, however, several error sources from the timing logic and interpolators deteriorate precision. In asynchronous measurement, the single-shot precision is often limited by gain error or, more generally, by the nonlinearity of the interpolators.

In Ref. 34 the effect of interpolator nonlinearity on the accuracy and precision of the TDC is analyzed. Assuming a general form $e(x)$ for the interpolator error (nonlinearity and/or gain error), using subscripts 1 and 2 for the start and stop interpolators, respectively, and by normalizing T_{clk} to 1, the TDC measurement error for input time interval $Q + F$ (Q is an integer and $0 \leq F < 1$) in a single measurement can be calculated as

$$\begin{aligned} e(x) &= e_1(x) - e_2(x + F), x + F < 1 \\ e(x) &= e_1(x) - e_2(x + F - 1), x + F \geq 1 \end{aligned} \quad (5)$$

The maximum single-shot measurement error is now equal to the maximum difference $e_1(x) - e_2(x)$, and the measurement result distribution is no longer binomial but more than two

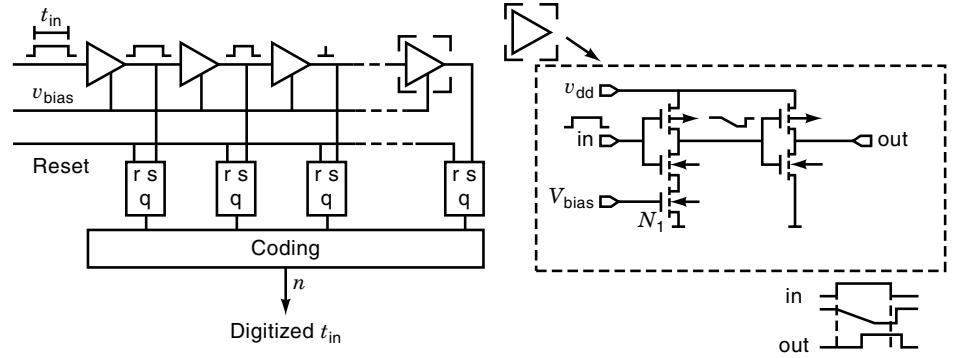


Figure 14. Pulse-shrinking delay line and schematic of the delay element.

results are possible. Since in asynchronous measurement x varies randomly and with equal probability between $0 \leq x < 1$, the mean value of the TDC measurement error is

$$\begin{aligned} m &= \int_0^1 e(x) dx = \int_0^1 e_1(x) dx - \int_0^{1-F} e_2(x+F) dx \\ &\quad - \int_{1-F}^1 e_2(x+F-1) dx \\ &= m_1 - m_2 \end{aligned} \quad (6)$$

where m_1 and m_2 are the mean values of the start and stop interpolator errors. Thus, the gain error or the INL of interpolators reduces to a constant bias error independent of the time to be measured. This is an extremely important result as it indicates that the error due to the interpolator nonlinearities is effectively averaged out in a repeated measurement due to the randomization process present in the asynchronous measurement. This is the source for the excellent linearity of the method.

The nonlinearities of the interpolators do have an effect on the precision of the system, however. The variance (σ^2) of the TDC measurement result is (34)

$$\begin{aligned} \sigma^2(x) &= \int_0^1 [e(x) - m]^2 dx \\ &= e_{1\text{rms}}^2 + e_{2\text{rms}}^2 - (m_1 - m_2)^2 - 2 \int_0^{1-F} e_1(x) e_2(x+F) dx \\ &\quad - 2 \int_{1-F}^1 e_1(x) e_2(x+F-1) dx \end{aligned} \quad (7)$$

Therefore, the standard deviation of the measurement result is a function of the fractional part (F) of the input time interval ($Q + F$). Since $0 \leq F < T_{\text{clk}}$, the measurement precision due to interpolator nonlinearity or gain error has now a period of T_{clk} instead of LSB.

To clarify the effect of interpolator gain or linearity error on precision, it is relatively easy to evaluate Eq. (7) if only gain error is considered. Assuming an ideal gain of K and an actual gain of k , the gain error of the interpolator can be calculated as

$$e(x) = x \left(\frac{k}{K} - 1 \right) \quad (8)$$

By making use of Eqs. (6) and (7) the mean and standard deviation of the measurement results can now be calculated. In

Fig. 15 are shown three examples of the results due to the combined effect of quantization error and $\pm 1\%$ gain error in 8-bit interpolators. For ideal interpolators the mean value of the TDC measurement error is zero ($m_k = 0$) and the maximum value of σ is $0.5 \times \text{LSB} \approx 0.002 T_{\text{clk}}$. The effect of gain error on precision can be clearly seen, since, assuming $T_{\text{clk}} = 10$ ns, for example, the worst-case σ value increases from the theoretical value of 20 ps to 60 ps [Fig. 15(c)].

Gain error, unlike nonlinearity, is relatively easy to cancel by calibrating the interpolators. It is also possible to measure the gain of both interpolators and use these values for calculation of the measurement results provided that the worst-case measurement resolution is adequate and that no overflow occurs. Nonlinearity of the interpolators can only be compensated by measuring the linearity error and using this data for correction of the measurement result. However, since this is technically more complicated than in the case of gain error, minimization of interpolator nonlinearity is important.

Also, jitter in the input timing signals, clock, and timing logic deteriorates precision with respect to the theoretical value. However, their effect is usually small compared to the nonlinearity of the interpolators. Timing jitter of the logic due to thermal, shot, or flicker noise is fairly small if a resolution of tens of picoseconds is adequate. According to simulations (35) and measurements (11), rms jitter seems to be 0.1% to 0.2% of the gate delay for both CMOS and bipolar elements. For 20 to 30 gates in series in the timing logic, the total rms jitter is only 1 ps to 3 ps for 200 ps to 300 ps gate delay. However, power-supply noise or other types of noise coupling can significantly increase jitter.

Linearity

In asynchronous averaging the linearity of a TDC based on the Nutt method is basically as good as the linearity of the counting method. Gain error or nonlinearity of the interpolators increases the single-shot error from the theoretical quantization error, but for averaged results their effect is a constant bias error offset independent of the time to be measured (as previously shown). Thus, interpolator gain error and nonlinearity have no effect on the linearity of the TDC, provided that enough samples are being averaged. If the interpolators were exactly identical, their errors would cancel completely (zero offset). In practice, the magnitude of the offset depends on the matching of the interpolators. If the nonlinearity is due to systematic errors, the interpolators usually have very similar nonlinearities and only a small offset error can be expected to remain.

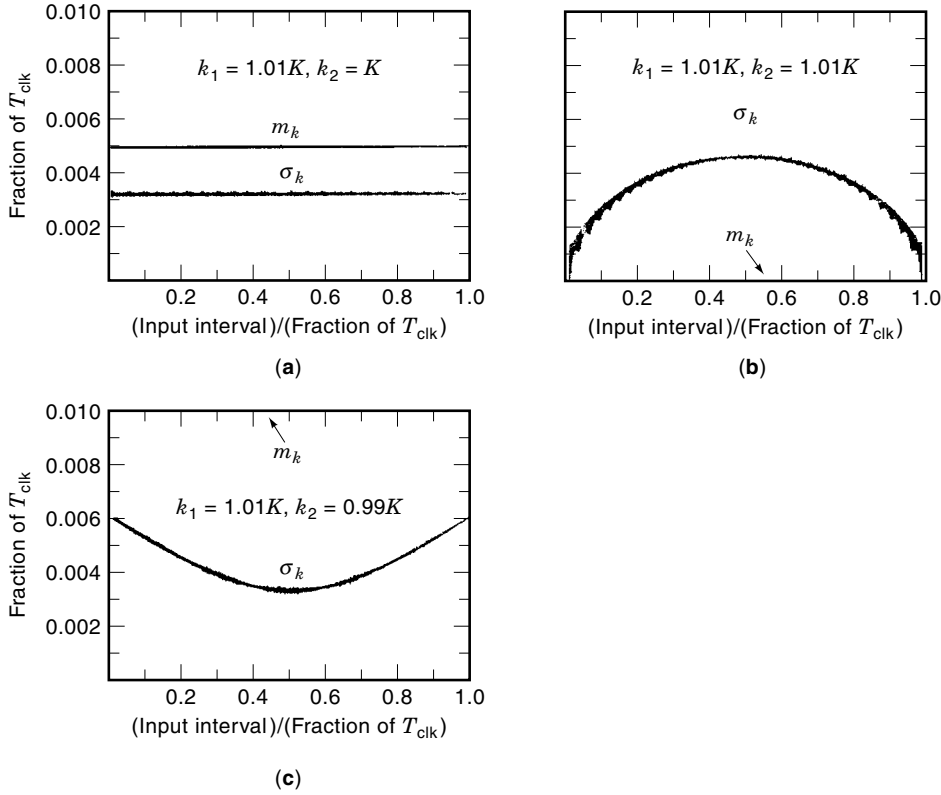


Figure 15. Simulated error of average (m_k) and standard deviation (σ_k) of the TDC result due to 8-bit quantization and $\pm 1\%$ interpolator gain error in asynchronous measurement.

However, if the nonlinearity is due to random mismatch (as in well-designed delay lines, for example), a similar cancellation cannot be assumed.

A prerequisite for the randomization of the systematic errors in averaging measurement is that the system clock and the time intervals are asynchronous, that is, the length of the time fractions T_1 and T_2 vary randomly in the operating range of the interpolator in a repeated measurement (although their difference has only two values). This is indicated by the simulation results of Fig. 16, which show the linearity error and standard deviation when Gaussian jitter with rms amplitudes $0.1 T_{\text{clk}}$ to $1 T_{\text{clk}}$ is added to the start input. It is clearly seen that jitter with a standard deviation of $1 T_{\text{clk}}$ effectively randomizes the measurement.

Stability

The stability error of a TDC based on the interpolation method can be divided into two components. The first is an offset-type

error that is independent of the input time interval and arises in the control logic (differences in delays) and interpolators (mismatch). If the start and stop signal paths in the control block and in the interpolators were identical, including the loading and layout, then, according to Eqs. (4) and (6), the measurement results would have zero offset. However, random mismatch cannot be avoided, and when this offset changes with temperature, for example, stability error is created. Methods for calibrating the TDC to compensate interpolator offset (and the stability error) are presented in Ref. 36. The effect of the mismatch in the interpolators can be minimized by using interpolators alternately in the start and stop channels, see Ref. 37.

The second error component arises from the temperature or time dependency of the reference oscillator. This gain error is dependent on the input interval. For example, a stability specification of ± 25 ppm inclusive of temperature dependency (0° to $+50^\circ\text{C}$) and aging corresponds to an error of $< \pm 25$ ps in a measurement range of $1 \mu\text{s}$.

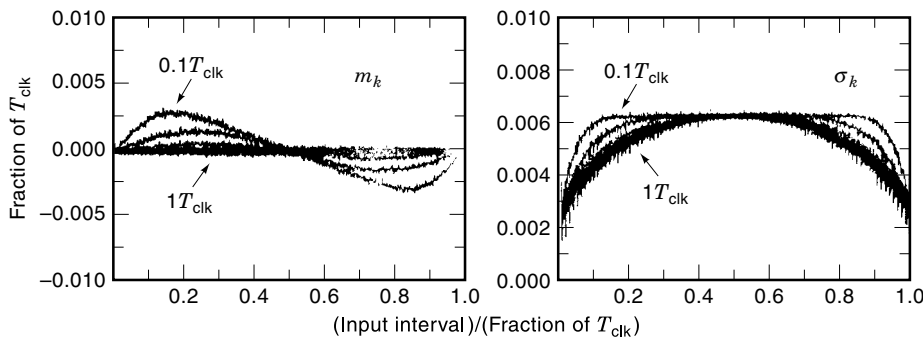


Figure 16. Simulated error of average (m_k) and standard deviation (σ_k) of the TDC result in synchronous measurement as a function of Gaussian jitter amplitude ($0.1 T_{\text{clk}}$ to $1 T_{\text{clk}}$).

Table 1. Some Commercially Available Frequency- and Time-Interval Counters

Manufacturer	Model No.	Time Base	Resolution	Range	Method
Hewlett Packard	HP 5314A	10 MHz	100 ns	250 ns–1 s	Counting
Hewlett Packard	HP 53131A		500 ps	1 ns–105 s	Counting
Hewlett Packard	HP 53132A		150 ps	1 ns–105 s	Counting
Hewlett Packard	HP 5370B		20 ps	–10 s–10 s	Digital Vernier
Hewlett Packard	HP 53310A		200 ps		
Hewlett Packard	HP E1725B		100 ps		
Hewlett Packard	HP E1740A		100 ps		
Tektronix	CMC251		10 ns	40 ns–1 s	Counting
Tektronix	73A541	10 MHz	100 ns	0–25 s	Counting
Standford Research	SR620	10 MHz	25 ps	±1000 s	Counting
Fluke	PM 6665		100 ns	100 ns– 2×10^8 s	Counting
Fluke	PM 6666	10 MHz	100 ns	100 ns– 2×10^8 s	Counting
Fluke	PM 6685	10 MHz	250 ps	6 ns–100 ms	Counting
Fluke	PM 6680B		250 ps	0 s– 10^{10} s	
Fluke	PM 6681		50 ps	0 s– 10^{10} s	
EG&G Ortec	457		<100 ppm $\times$ FS	50 ns–80 μ s	Analog TAC
EG&G Ortec	566		<100 ppm $\times$ FS + 5 ps	10 ns–2 ms	Analog TAC
EG&G Ortec	567		<100 ppm $\times$ FS + 5 ps	10 ns–2 ms	Analog TAC
EG&G Ortec	9308		<25 ps + 10μ l/l $\times$ FS	0 μ s–325 μ s	
Lecroy	2228A/2229	20 MHz	250 ps	0 ns–500 ns	Counting
Lecroy	4208	125 MHz	1 ns	±8.3 ms	Digital interpolation
Berkeley Nucleonics	B980	20 MHz	48.8 ps	±6871 s	Interpolation
Leti Dept. Syst.	Asic-circuit	12.5 MHz	1.3 ns	0.7 ns–8.7 μ s	Analog interpolation
AMSC	TMC1004	31.25 MHz	530 ps	~4 μ s	Del line
Racal Instruments	2251A	10 MHz	±1 ns	–2 ns– 8×10^5 s	Interpolation
Racal Instruments	2351	10 MHz	8 ps	2 ns–8 s	

In summary, it can be concluded that the nonidealities of the interpolators to large extent cancel out in averaged results. The difference between the nonidealities of the interpolators remains, but in many cases (systematic error source) it can be expected to be smaller than the nonideality itself. Furthermore, this error is constant for all input intervals so it can be corrected simply by subtracting it from the measurement results provided that it does not change with temperature, supply voltage, and the like. Single-shot accuracy, on the other hand, is affected by all nonidealities.

Commercial Time-Interval Counters. Table 1 lists certain universal frequency- or time-interval counters that are currently commercially available. This is not an exhaustive list of all useful types nor of all manufacturers. The column headed Time Base tells the used base frequency of the oscillator in the digital counting principle, if it is given. The precision gives the least significant digit in display or single-shot precision, whichever is presented in the manufacturers data sheet. In analog TACs the time range is usually selectable, in which case the minimum and maximum time ranges are given. The precision of the TAC is given as a function of the full-scale (FS) range.

BIBLIOGRAPHY

1. I. Kaisto et al., Laser radar based measuring systems for large scale assembly applications, *Proc. SPIE Int. Conf. Laser Dimens. Metrol.: Reent Adv. Ind. Appl.*, **2088**: 121–130, 1993.
2. K. Määttä, J. Kostamovaara, and R. Myllylä, Profiling of hot surfaces by pulsed time-of-flight laser range finder techniques, *Appl. Opt.*, **32** (27): 5334–5347, 1993.
3. Y. Arai and M. Ikeno, A time digitizer CMOS gate-array with 250 ps time resolution, *IEEE J. Solid-State Circuits*, **31** (2): 212–220, 1996.
4. E. Gerds et al., A CMOS time to digital converter with 2 level analog CAM, *IEEE J. Solid-State Circuits*, **29** (9): 1068–1076, 1994.
5. T. Otsuji and I. Narumi, A 3-ns range, 8-ps resolution, timing generator LSI utilizing Si bipolar gate array, *IEEE J. Solid-State Circuits*, **26** (5): 806–811, 1991.
6. T. Otsuji, A picosecond-accuracy, 700-MHz range Si-bipolar time interval counter LSI, *IEEE J. Solid-State Circuits*, **28** (9): 941–947, 1993.
7. Institute of Electrical and Electronics Engineers, *IEEE Standard Dictionary of Electrical and Electronics Terms*, New York: IEEE, 1984.
8. P. W. Nicholson, *Nuclear Electronics*, New York: Wiley, 1974.
9. Hewlett-Packard, *Time Interval Averaging*, Appl. Note 162-1, Hewlett-Packard.
10. R. W. Rhea, *Oscillator Design & Computer Simulation*, Englewood Cliffs, NJ: Prentice-Hall, 1990.
11. M. Loinaz and B. Wooley, A BiCMOS time interval digitizer based on fully-differential, current-steering circuits, *IEEE J. Solid-State Circuits*, **29** (6): 707–713, 1994.
12. D. Porat, Review of sub-nanosecond time-interval measurements, *IEEE Trans. Nucl. Sci.*, **NS-20** (1): 36–51, 1973.
13. D. Chu, M. Allen, and A. Foster, Universal counter resolves picoseconds in time interval measurements. *Hewlett-Packard J.*, **29** (12): 2–11, 1978.
14. R. Nutt, Digital time intervalometer, *Rev. Sci. Instrum.*, **39** (9): 1342–1345, 1968.
15. J. Kostamovaara and R. Myllylä, Time-to-digital converter with an analog interpolation circuit, *Rev. Sci. Instrum.*, **57** (11): 2880–2885, 1986.
16. J. Rabaey, *Digital Integrated Circuits*, Upper Saddle River, NJ: Prentice Hall, 1996.
17. I. J. Taylor and T. H. Becker, A high speed time-to-amplitude converter using integrated circuits, *Nucl. Instrum. Methods*, **99**: 387, 1972.

18. J. Kostamovaara and R. Myllylä, A time-to-amplitude converter with constant fraction timing discriminators for short time interval measurements, *Nucl. Instrum. Methods Phys. Res. A*, **239**: 568–578, 1985.
19. K. Määttä, J. Kostamovaara, and R. Myllylä, Time-to-digital converter for fast, accurate laser rangefinding, *SPIE Proc. Int. Congr. Opt. Sci. Eng.*, **1010**: 60–67, 1988.
20. A. Stevens et al., A time-to-voltage converter and analog memory for colliding beam detectors, *IEEE J. Solid-State Circuits*, **24** (6): 1748–1752, 1989.
21. J. Kalisz, M. Pawlowski, and R. Pelka, A multiple-interpolation method for fast and precise time digitizing, *IEEE Trans. Instrum. Meas.*, **35** (2): 163–169, 1986.
22. B. Turko, A modular 125ps resolution time interval digitizer for 10MHz stop burst rates and 33ms range, *IEEE Trans. Nucl. Sci.*, **NS-26** (1): 737–745, 1979.
23. Y. Arai, T. Matsumura, and K. Endo, A CMOS four-channel time memory LSI with 1-ns/b resolution, *IEEE J. Solid-State Circuits*, **27** (3): 359–364, 1992.
24. J. Dunning et al., An all-digital phase-locked loop with 50-cycle lock time suitable for high-performance microprocessors, *IEEE J. Solid-State Circuits*, **30** (4): 412–422, 1995.
25. M. Johnson and E. Hudson, A variable delay line PLL for CPU-coprocessor synchronization, *IEEE J. Solid-State Circuits*, **23** (5): 1218–1223, 1988.
26. P. Andreani et al., A digitally controlled shunt capacitor CMOS delay line, *Proc. 15th Norchip Semin.*, 1997, Vol. 1, pp. 166–171.
27. T. Rahkonen, J. Kostamovaara, and S. Säynäjäkangas, CMOS ASIC devices for the measurement of short time intervals, *Proc. IEEE Int. Symp. Circuits Syst.*, 1988, Vol. 2, pp. 1593–1596.
28. J. Genat, High resolution time-to-digital converters, *Nucl. Instrum. Methods Phys. Res. A*, **315**: 411–414, 1992.
29. C. T. Gray et al., A sampling technique and its CMOS implementation with 1 Gb/s bandwidth and 25 ps resolution, *IEEE J. Solid-State Circuits*, **29** (3): 340–349, 1994.
30. J. Christiansen, An integrated high resolution CMOS timing generator based on array of delay locked loops, *IEEE J. Solid-State Circuits*, **31** (7): 952–957, 1996.
31. T. Rahkonen and J. Kostamovaara, Pulsewidth measurements using an integrated pulse shrinking delay line, *Proc. IEEE Int. Symp. Circuits Syst.*, 1990, Vol. 1, pp. 578–581.
32. T. Rahkonen, E. Malo, and J. Kostamovaara, A 3-V fully integrated digital FM demodulator based on a CMOS pulse-shrinking delay line, *Proc. IEEE Int. Symp. Circuits Syst.*, 1996, Vol. 2, pp. 572–575.
33. J. Kalisz et al., Single-chip interpolating time counter with 200-ps resolution and 43-s range, *Proc. IEEE Instrum. Meas. Technol. Conf.*, 1996, Vol. 1, pp. 676–680.
34. J. Kalisz, M. Pawlowski, and R. Pelka, Error analysis and design of the Nutt time-interval digitizer with picocond resolution, *J. Phys. E: Sci. Instrum.*, **20**: 1330–1341, 1987.
35. T. Rahkonen, Circuit techniques and integrated CMOS implementations for measuring short time intervals, *Acta Univ. Ouluensis*, **C73**: 1993.
36. T. Rahkonen and J. Kostamovaara, The use of stabilized CMOS delay lines for the digitization of short time intervals, *IEEE J. Solid-State Circuits*, **28**: 887–894, 1993.
37. E. Räisänen-Ruotsalainen, T. Rahkonen, and J. Kostamovaara, A low-power time-to-digital converter, *IEEE J. Solid-State Circuits*, **30**: 984–990, 1995.
38. M. A. Lombardi, An introduction to frequency calibration: Part I., *Cal. Lab. Int. J. Metrology*, Jan.–Feb., 1996, pp. 17–28.
39. M. A. Lombardi, An introduction to frequency calibration: Part II., *Cal. Lab. Int. J. Metrology*, Mar.–Apr., 1996, pp. 28–34.
40. V. S. Zhang, D. D. Davis, and M. A. Lombardi, High resolution time interval counter, *27th Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1994, pp. 191–200.

JUHA KOSTAMOVARA
 ELVI RÄISÄNEN-RUOTSALAINEN
 KARI MÄÄTTÄ
 University of Oulu

CLOCKS IN TELECOMMUNICATIONS

The concept of time flow is related to the possibility of arranging events in an ordered sequence, that is, saying which of two events comes earlier (1). Then, a timescale is a system of labeling events with real numbers according to their sequence. The labeling law is quite arbitrary, given that the order of events is left unchanged, but for practical purposes it must use equal intervals (two time intervals are defined equal if equal processes take place during these two intervals) for its successive scale intervals (*uniform time scale*).

Clocks are devices able to measure the time, that is, to produce time markers together with identification of these markers (1). Under different forms, they are widespread in everyday life and penetrate a surprisingly wide range of applications. Beyond ruling human (and automated) activities worldwide, for example, clocks provide the timing of digital electronics in almost all of the commonly used electronic equipment and synchronize the nodes of telecommunications systems and networks. The performance of such systems heavily relies on the quality of the synchronization signals.

The oldest historical examples of clocks date back to 1400 B.C. and are a sun dial and a clepsydra (water clock) made by Egyptians. With several technical improvements, sun dials (based on measuring the rotation of a shadow with the sun) and clepsydrae (based on measuring the level of water in a vessel with a regulated flow of water as input or output) had been in use until the Middle Ages. On the other hand, early mechanical clocks date back to the thirteenth century, and the first pocket watches, based on a spring mechanism, were constructed in the fifteenth century. However, the true milestone was the invention of pendulum clocks due to Galileo and Huygens and the introduction of the swing wheel as an oscillating element. After that, until the introduction of electrical clocks and then of clocks based on quartz and atomic oscillators, mechanical clocks did not change substantially, at least in their operating principle, until today.

From a theoretical viewpoint, the operating principle of clocks of any kind consists of a generator of oscillations and an automatic counter of such oscillations. Although differing in capability, the oscillator can be based on *any* (pseudo-)periodic physical phenomenon. The swinging of a pendulum or a wheel in mechanical clocks, the vibration of atoms in a crystal around their minimum-energy position in quartz clocks, the radiation of specific quantic atomic transitions in atomic clocks are the best known examples because of their wide application, but are not the only ones. Recently, even the rotation period of pulsars, after some (although not trivial) data processing, has been used to design clocks of the highest precision, comparable to that of the best atomic clocks.

CLOCK SIGNALS AND THEIR TIMING RELATIONSHIPS

A clock is a device supplying a *timing signal*, or *chronosignal*, defined as a pseudoperiodic (ideally periodic) signal for controlling the timing of actions. Examples of applications are the timing of digital hardware systems (gates, chips, or boards), where the operation of different modules must be synchronized to ensure proper transfer of binary symbols, and the timing of telecommunications systems, where digital signals are multiplexed; transmitted, and switched.

Typical chronosignal waveforms are sine and square waves. A general expression describing a pseudoperiodic waveform which models the timing signal $s(t)$ at the output of clocks is given by (2,3)

$$s(t) = A(t) \sin \Phi(t) \quad (1)$$

where $A(t)$ is the instantaneous amplitude (in the following, without loss of generality, we assume $A(t) \cong A_0$), $\Phi(t)$ is the total phase, and the instantaneous frequency $\nu(t)$ is given by

$$\nu(t) = \frac{1}{2\pi} \frac{d\Phi(t)}{dt} \quad (2)$$

Obviously, in the ideal case, when the pseudoperiodic timing signal is periodic, the total phase increases linearly with time, that is,

$$\Phi(t) = 2\pi \nu_n t \quad (3)$$

where ν_n is the nominal frequency. In the actual case, other components may affect the total phase increase, including frequency offset, drifts, and purely random fluctuations. This topic is thoroughly discussed in the section Characterization and Modeling of Clocks.

A chronosignal fulfils its duty by triggering events, that is, timing the controlled process. From this point of view, a timing signal can be also modeled by a series of pulses spaced T apart, at special instants called *significant instants*. The timing signal triggers the controlled process at those instants. Suitable significant instants can be identified, for example, at the signal zero-crossing instants for ease of implementation.

Synchronous and Asynchronous Clock Signals

Two *synchronous* (from the Greek etymon $\sigma\upsilon\gamma\chi\rho\nu\nu\omicron\varsigma$, built by $\sigma\upsilon\nu = with$ and $\chi\rho\nu\nu\omicron\varsigma = time$) clock signals have the same frequency, at least on average, and a precisely controlled phase relationship (i.e., with phase offset $\Delta\Phi = \text{constant}$, at least on average). The expression “at least on average” points out that some zero-mean small fluctuations may be accepted as unavoidable in real systems.

Conversely, two clock signals are *asynchronous* if they are not synchronous.

Mesochronous Clock Signals

Two *mesochronous* (from the Greek etyma $\mu\acute{\epsilon}\sigma\omicron\varsigma = medium$ and $\chi\rho\nu\nu\omicron\varsigma = time$) clock signals are asynchronous timing signals which have the same frequency, at least on average, but no control on the phase relationship. It is worth noting that because the phase fluctuation function is proportional to the integral of the frequency fluctuation function, the phase error

$\Delta\Phi$ is not theoretically limited over an infinite time interval, even for small zero-mean frequency fluctuations (thus, $\Delta\Phi = \text{any value}$).

Plesiochronous Clock Signals

Two *plesiochronous* (from the Greek etyma $\pi\lambda\eta\sigma\acute{\iota}\omicron\varsigma = close$ and $\chi\rho\nu\nu\omicron\varsigma = time$) clock signals are asynchronous timing signals that have the same frequency values only nominally, but actually are different within a given tolerance range.

Heterochronous Clock Signals

Two *heterochronous* (from the Greek etyma $\acute{\epsilon}\tau\epsilon\rho\omicron\varsigma = different$ and $\chi\rho\nu\nu\omicron\varsigma = time$) clock signals are asynchronous timing signals that have different nominal frequencies.

Practical Examples

To give sound examples of the above abstract concepts, a phase-locked loop (PLL), locked to its reference, outputs a timing signal that is synchronous with the input signal because of the feedback control on the phase error between the signals. A frequency-locked loop (FLL), which is a feedback system operating like a PLL, but which instead controls the frequency error between the input and the output signals, outputs a signal that is mesochronous with the input. Two oscillators, even designed and built as equal by the same supplier, output plesiochronous timing signals because of unavoidable manufacturing tolerances. In telecommunications networks also, two digital signals that have the same nominal bit rate (e.g., two 2.048 Mb/s digital signals) are always plesiochronous unless generated by the same piece of equipment (driven, in this case, by a single clock). On the other hand, two digital signals with different rates (e.g., 2.048 Mb/s and 8.448 Mb/s signals) are heterochronous.

NETWORK SYNCHRONIZATION STRATEGIES

Network synchronization deals with the distribution of time and frequency over a network of clocks spread over a wide geographical area (4). The goal is to align the time and frequency scales of all of the clocks by using the communications capacity of the links interconnecting them (e.g., copper cables, fiber optics, radio links). In particular, network synchronization plays a central role in modern digital telecommunications (5,6) and has a determining influence on the quality of most services offered by the network operator to its customers.

Many intriguing examples of synchronizing a large number of oscillators are found in nature. Lindsey et al. (4) pointed out as one of the most spectacular ones the synchronous fireflies described by Buck and Buck (7). These fireflies flash their light organs at regular but individual and independent intervals if they are not close together. Though, if many of these insects are placed in a relatively close proximity, they synchronize their light organs until they flash in unison. Other biological examples are the synchronization of individual fibers in heart muscles to produce the familiar heartbeat or the resting and active periods of mammals, which exhibit rhythms.

During the last few decades, many different network synchronization strategies have been conceived, mainly for synchronizing the nodes of telecommunications networks, nowadays the most demanding application of network

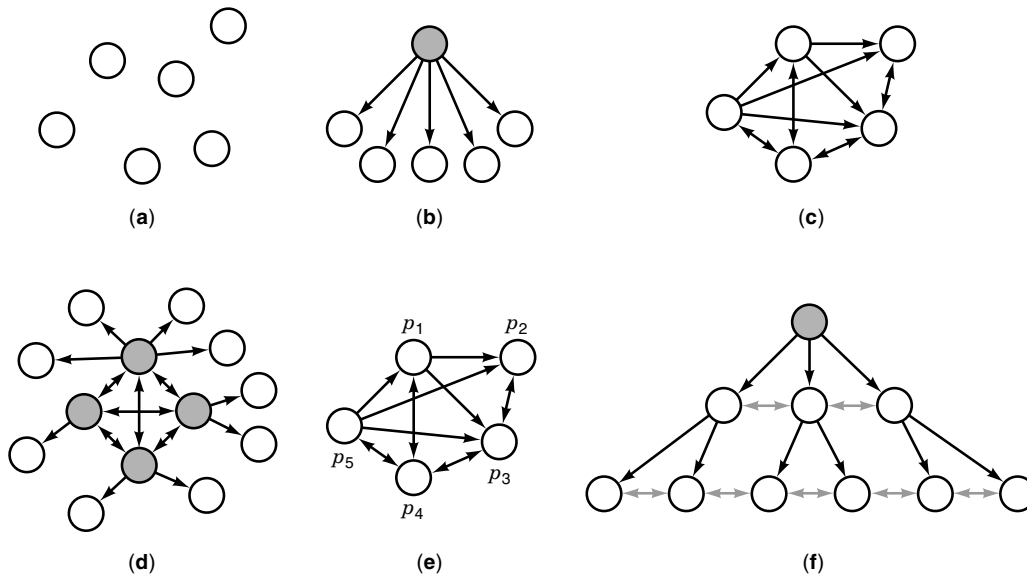


Figure 1. Network synchronization strategies. (a) Full plesiochrony; (b) Master-slave synchronization (despotism); (c) Mutual synchronization (democracy); (d) Mixed mutual/master-slave synchronization (oligarchy); (e) Hierarchical mutual synchronization (hierarchical democracy); (f) Hierarchical master-slave synchronization (hierarchical despotism).

synchronization. The main features of such strategies are now summarized briefly, pointing out for each one its sociopolitical analogy. Such an analogy may be helpful in understanding immediately the pros and cons of each strategy, because it may be more direct to assess a form for governing human beings rather than clocks.

Full Plesiochrony (Anarchy)

The full plesiochronous strategy is actually a *no-synchronization strategy*, that is, does not involve any synchronization distribution to the network of clocks. Each clock is independent of the others (*autonomous clocks*), as shown in Fig. 1(a), hence the expression synchronization *anarchy*.

Anarchy is the easiest form of government, but it relies on the good behavior of the single elements. Because of the lack of any timing distribution, the synchronization of the processes in different nodes is entrusted to the accuracy of the network's autonomous clocks, which therefore must feature excellent performance.

In the late 60s, this strategy was generally considered the most promising for synchronizing telecommunications networks because of the decreasing cost of atomic oscillators and the limited synchronization requirements envisaged then. Nevertheless, as the cost of such oscillators became stable and the new digital transmission and switching techniques demanded increasing timing performance, this strategy was eventually abandoned.

Master-Slave Synchronization (Despotism)

The principle of master-slave (MS) strategies is based on the distribution of the timing reference from a clock (*master clock*) to all of the other clocks of the network (*slave clocks*), directly or indirectly as shown in Fig. 1(b). Although the master clock must be an expensive high-precision oscillator (usually based on an atomic frequency standard), slave clocks can be much

less expensive. A slave clock is usually implemented as a PLL, mostly based on a quartz crystal oscillator and locked to the reference timing signal coming from the master. In PLLs, the output timing signal is kept synchronous with the input reference by a feedback control on the phase error between them.

The loop time constant of the PLL controls its filtering properties from the phase fluctuations of the input reference and of the internal oscillator. Further details can be found in the section "Characterization and Modeling of Clocks" and in Refs. 8–11. Therefore, an important step in designing a MS synchronization network is to design the loop time constants of the slave clocks in different nodes, according to specific requirements.

Despotism may look unethical, but it is certainly effective in ensuring very tight control on the slaves. A MS network is synchronous with the master clock and stable by definition. Therefore, MS-based strategies are currently the most widely adopted in many applications. Questions may arise, nevertheless, on what happens if the unique master fails. The hierarchical MS strategy has been conceived for this purpose.

Mutual Synchronization (Democracy)

Mutual synchronization is based on direct, *mutual control* among the clocks, so that the output frequency of each one is the result of the "suggestions" of the others, as shown in Fig. 1(c). Such pure democracy looks appealing. There are no masters and no slaves, but mutual cooperation exists, though the "discipline" of the mutually controlled elements is hard to guarantee.

Modeling the behavior of such networks or even ensuring the stability of the control algorithms (the control algorithm must damp the transient impairments, preventing them from propagating indefinitely) can be a very complex task (4,12,13). Thus networks so designed are quite expensive, but they are extremely reliable. Hence, till now, the field of application of

mutual synchronization has been limited mostly to special cases, for example, military networks.

Mixed Mutual/Master-Slave Synchronization (Oligarchy)

In this mixed solution, the mutual synchronization strategy is adopted for the main clocks of the network, and the MS strategy is adopted for the peripheral clocks, as shown in Fig. 1(d).

Oligarchy is a compromise that aims at mitigating the absolutism of despotism. Greater reliability is achieved, and the peripheral MS synchronization substantially simplifies the system control architecture compared with pure mutual synchronization.

Hierarchical Mutual Synchronization (Hierarchical Democracy)

Hierarchical mutual synchronization is a generalization of the democratic strategy. In hierarchical democracy, some count more than others.

All of the network nodes are given a relative weight p_i ($0 \leq p_i \leq 1$, $\sum_i p_i = 1$), as shown in Fig. 1(e). When all of the weights are equal, this strategy becomes pure mutual synchronization. When the weight of one clock is equal to one and all of the others are equal to zero, this strategy becomes a MS synchronization.

Hierarchical Master-Slave Synchronization (Hierarchical Despotism)

The Hierarchical Master-Slave (HMS) synchronization strategy is a variant of the pure MS strategy. A master clock synchronizes the slave clocks, directly or indirectly, and these are organized in two or more hierarchical levels [see Fig. 1(f)]. Protective mechanisms against link and clock failures are allowed through alternative transversal synchronization routes. If the master fails, another clock takes its place according to a hierarchical plan.

The HMS strategy is currently the most widely adopted for synchronizing modern digital telecommunications networks because of the excellent timing performance and reliability which is achieved at limited cost.

Mixed Plesiochronous/Synchronous Networks (Independent Despotic States)

Although most national administrations adopted the HMS strategy for synchronizing their national networks, locking all of them to a supranational timing reference is not feasible (it is worth noting that the GPS can take this role from a *technical* point of view, but for *political* reasons it is hardly accepted as first-choice reference by most national administrations). Therefore, the current arrangement is to have several synchronous HMS networks, each plesiochronous relative to the others. The political analogy is a set of independent despotic states.

CHARACTERIZATION AND MODELING OF CLOCKS

Designing such complex systems as networks of clocks or specifying the quality requirements of clocks suitable for a given application is not an obvious task. To face these issues effectively, first it is necessary to identify a proper mathematical model of the clock and of the timing signals generated and distributed. This section supplies this background knowledge. Simplified models of autonomous and slave clocks are described and the

basics of time and frequency stability characterization are provided.

Timing Signal Model

A general expression describing a pseudoperiodic waveform that models the timing signal $s(t)$ at the output of clocks is given by Eq. (1). A common and comprehensive model used to characterize $v(t)$, defined by Eq. (2), is given by [see the ideal case of Eq. (3)]

$$\begin{aligned} v(t) &= v_0 + v_d(t) + v_a(t) \\ &= v_n + \Delta v + \sum_{k=1}^{K-1} \frac{q_k t^k}{k!} + \frac{1}{2\pi} \frac{d\varphi(t)}{dt} \end{aligned} \quad (4)$$

The *nominal frequency* v_n (design goal) and the *starting frequency offset* Δv (also called *synchronization error*) make up the starting frequency v_0 of the oscillator:

$$v_0 = v_n + \Delta v \quad (5)$$

The term $v_d(t)$ is the deterministic (for a given oscillator) time-dependent component, modeling the *frequency drift* mainly caused by oscillator aging as a power series:

$$v_d(t) = \sum_{k=1}^{K-1} \frac{q_k t^k}{k!} \quad (6)$$

The coefficients q_i ($k = 1, 2, \dots, K-1$) are time-independent, random variables that are fixed for a given oscillator. The frequency drift in real clocks is caused by complex phenomena (14,15), but in practice, for simplicity, the previous summation is often truncated to the first term, so that

$$v_d(t) \cong q_1 t = D v_n t \quad (7)$$

where D is the *linear, fractional-frequency drift rate*.

Finally, the term $v_a(t)$ is the random time-dependent component

$$v_a(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt} \quad (8)$$

where $\dot{\varphi}(t)/(2\pi)$ and $\varphi(t)$ are stochastic processes, respectively, the *random frequency deviation* and the *random phase deviation* that model oscillator intrinsic phase-noise sources.

Basic Quantities

Two functions strictly related to $\dot{\varphi}(t)$ and $\varphi(t)$ are used in treating random frequency and time fluctuations: the *random fractional-frequency deviation* $y(t)$ and the *random time deviation* $x(t)$, defined as

$$\begin{aligned} y(t) &= \frac{1}{2\pi v_n} \frac{d\varphi(t)}{dt} \\ x(t) &= \frac{\varphi(t)}{2\pi v_n} \end{aligned} \quad (9)$$

These functions, together with the model and definitions provided in the previous section, have been widely adopted by

specialists since the 60s (2,16). More recently, needs in particular in the telecommunications field for designing synchronization equipment and networks led to the introduction of the following other basic functions, more oriented to the timing aspects of clocks.

The generated *time* function $T(t)$ of a clock is defined in terms of its total instantaneous phase as

$$T(t) = \frac{\Phi(t)}{2\pi\nu_n} \quad (10)$$

It is worth noting that for an ideal clock $T_i(t) = t$ holds, as expected. For a given clock, the *time error* function $TE(t)$ between its time $T(t)$ and a reference time $T_r(t)$ is defined as

$$TE(t) = T(t) - T_r(t) \quad (11)$$

Whereas the time error variation over an interval of duration τ starting at time t (i.e., the error committed by the clock in measuring an interval τ with respect to the reference clock) is called *time interval error* $TIE_t(\tau)$ and is defined as

$$\begin{aligned} TIE_t(\tau) &= [T(t + \tau) - T(t)] - [T_r(t + \tau) - T_r(t)] \\ &= TE(t + \tau) - TE(t) \end{aligned} \quad (12)$$

Now, it must be pointed out that $x(t)$ and $TE(t)$ have very similar definitions, but they differ in that the $TE(t)$ function takes into account both deterministic [the $\Delta\nu$ and $\nu_d(t)$ terms in Eq. (4)] and random phase-noise components, whereas $x(t)$ depends only on random components. Finding the deterministic components in the TE measured data may not be straightforward, and the result can depend greatly on the parameter estimation technique adopted (14,17,18).

Basic Concepts of the Quality of Clocks

Characterizing the *quality* of a clock (or equivalently of its timing signal) is one of the most complex and debated tasks in coping with the issues involved in practical applications of clocks. In the most common sense, one simple term is used to refer to clock quality: the *precision*, somehow denoting how close the timing signal of the clock under test is to a reference timing signal. From a more technical viewpoint, the quality of a clock is usually defined by two other basic terms: stability and accuracy. Although it must be recognized that researchers and engineers working in different fields may understand these two terms with subtly different meanings, the definitions provided in the following can be considered quite general.

The *stability* of a clock deals with measuring the random and deterministic variations of the instantaneous frequency (or of the time) of the output timing signal, compared to the nominal value (i.e., in practice, to a reference clock), over a given observation interval. With reference to the mathematical model of Eq. (4), the clock stability depends on the random phase noise $\varphi(t)$, the frequency offset $\Delta\nu$, and the frequency drift D (and all the coefficients q_i). The relative weights of such parameters in affecting stability depend on the observation interval. If it is short, the frequency drift is negligible.

When the observation interval is small, the expression *short-term* stability is commonly used. Otherwise the expression *long-term* stability applies. What should be the meaning of the word “small”, that is, where the boundary between short and

long term is, depends on the specific application. For example, in the time metrology field, it is common to consider observation intervals longer than one day as long-term, whereas in telecommunication applications observation intervals over 100 s are also definitely considered long term.

Since the 1960s, several quantities have been defined that aim at characterizing clock stability. Although they differ in applicability, they highlight distinct phenomena in the phase noise or they are more oriented to some application. A few details are given in the following subsections.

Accuracy, on the other hand, denotes the maximum frequency error $\Delta\nu_{\text{MAX}}$ compared to the nominal value, which may be measured in general over the entire clock life (e.g., twenty years), unless specified differently. It must be pointed out that accuracy also depends in principle on the previously mentioned parameters $\varphi(t)$, $\Delta\nu$, and the coefficients q_i , but in this case the observation interval is so long that in practice the only relevant quantities are the frequency offset and drift. Accuracy is usually expressed by the dimensionless ratio $\Delta\nu_{\text{MAX}}/\nu_n$ and is often measured in 10^{-6} units ($\mu\text{Hz}/\text{Hz}$), in engineering practice also called parts per million and abbreviated as ppm (not a SI unit).

Autonomous Clocks

An *autonomous clock* is a device for generating a timing signal, suitable for measuring time intervals, starting from a periodic physical phenomenon.

Examples of state-of-the-art autonomous clocks are atomic frequency standards (such as the rubidium or cesium-beam or hydrogen-MASER oscillators) and crystal quartz oscillators. Some of them (the rubidium and the quartz oscillators) may also be locked to a reference timing signal and thus work as slave clocks.

A simplified model of an autonomous clock is shown in Fig. 2. Here, according to the timing signal model of Eq. (4), the total output phase $\Phi(t)$ consists of three deterministic terms that model the phase generated by an ideal oscillator, the frequency offset $\Delta\nu$, and the frequency drift D , together with the random phase noise $\varphi(t)$.

The parameter $\Delta\nu$ decreases with more accurate clock calibration procedures (this is part of the production process). Such calibration procedures aim at making the clock generate an output frequency as close as possible to the nominal frequency. As for the frequency drift coefficient D , it is worth noting that it is practically null in cesium-beam oscillators, but it cannot be neglected in rubidium and quartz oscillators, at least in the most demanding applications. Finally, some details above the characteristics of the random phase-noise process $\varphi(t)$ are provided in the section Common Types of Clock Noise.

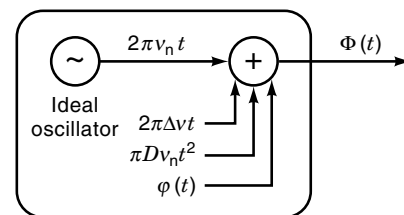


Figure 2. Simplified model of autonomous clock.

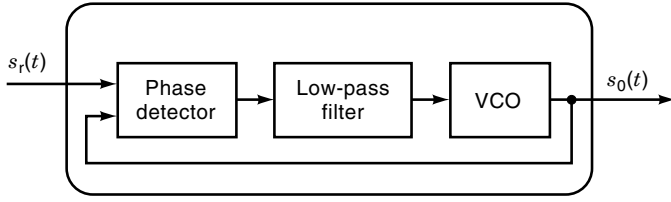


Figure 3. Scheme of the principle of a phase-locked loop (PLL).

Slave Clocks

A *slave clock* is a device for generating a timing signal, suitable for measuring time intervals whose phase (or much less frequently frequency) is locked to a reference timing signal at its input. They are usually implemented as PLLs (or FLLs).

Slave clocks are very widely employed (see Network Synchronization Strategies) in synchronization networks and in digital telecommunication equipment. Such a widespread application of slave clocks calls for a somewhat detailed description of their models and properties. Therefore, the next subsections deal with the main characteristics of PLLs and the operating modes of slave clocks in synchronization networks.

Phase-Locked Loops. A device that implements phase locking to a reference timing signal commonly uses loop architecture, based on the negative feedback principle, where the output signal keeps tracking the phase fluctuations of the input reference. Such a device is called *phase-locked loop* (PLL).

A scheme of the principle of PLL operation is depicted in Fig. 3. In this scheme, the following functional blocks are used:

- *phase detector*, which supplies a signal proportional to the phase error between the PLL output signal $s_o(t)$ fed back and the input reference $s_r(t)$;
- *low-pass filter (loop filter)*, whose task is canceling the high-frequency phase fluctuations at the output of the phase detector;
- *voltage-controlled oscillator (VCO)*, which supplies a periodic signal that has a frequency dependent on its input signal; the VCO, with null input signal, oscillates with angular frequency ω_F (VCO free-run angular frequency).

To summarize, the phase detector supplies a signal proportional to the phase error between $s_o(t)$ and $s_r(t)$. This signal is low-pass filtered and feeds the VCO to control its oscillation frequency and keep the phase error between $s_o(t)$ and $s_r(t)$ around zero.

Analysis of the PLL operation is very complex because of the nonlinear phase-detector block. A detailed analysis of its properties can be found in Refs. 8–10. Moreover, Ref. 11 glances at practical implementation. In this section, only the main features of PLL are overviewed.

Assuming that the phase error between $s_o(t)$ and $s_r(t)$ remains small (PLL locked to the reference), the system can be described by a linear model. Therefore, its input-output transfer function $H(s)$ between the input and output phase-modulation signals in the Laplace-transform domain can be evaluated. The transfer function $H(s)$ is typically of the second-order low-pass type (because the loop filter is typically first-order), with cut-off frequency B (PLL bandwidth). It is characterized by the natural angular frequency ω_n and the damping factor ζ and

thus has the form

$$H(s) = \frac{\phi_o(s)}{\phi_r(s)} = \frac{Cs + D}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (13)$$

where C and D are constants dependent on the particular PLL characteristics.

The PLL tracks the phase modulation of the reference signal $s_r(t)$ if the bandwidth of the phase-modulation signal is lower than B . Under static lock conditions (i.e., $s_r(t)$ is ideally periodic), the phase detector supplies a time-constant signal which makes the VCO oscillate at the same frequency of $s_r(t)$. The values ω_n and ζ are determined by the design of the PLL blocks, and they control the duration, amplitude, and shape of the transients in the PLL response $s_o(t)$ to disturbances in $s_r(t)$.

Noise in PLLs. Characterizing noise sources in a PLL is a very important topic that aims at evaluating slave clock performance. The linear model, depicted in Fig. 4, approximates PLL behavior by pointing out the main internal noise sources (a more thorough analysis is provided in Ref. 10). This model holds under the assumptions that the PLL is locked to the reference and that the noise amplitudes are small, compared to the signals.

In Fig. 4, ϕ_o and ϕ_r (rad) are the phase noises in $s_o(t)$ and $s_r(t)$, respectively, ϕ_{VCO} (rad) is the phase noise generated by the VCO, V_{DF} (V) is the tension noise produced by the phase detector and the loop filter, and $F(s)$ is the transfer function of the low-pass loop filter (all quantities in the Laplace domain).

By analyzing this model, considering one input at a time, the following transfer functions can be evaluated:

$$H(s) = \frac{\phi_o(s)}{\phi_r(s)} = \frac{K_0 K_d F(s)}{s + K_0 K_d F(s)} \quad (14)$$

$$H_A(s) = \frac{\phi_o(s)}{V_{DF}(s)} = \frac{K_0 F(s)}{s + K_0 K_d F(s)} \quad (15)$$

$$H_B(s) = \frac{\phi_o(s)}{\phi_{VCO}(s)} = \frac{s}{s + K_0 K_d F(s)} \quad (16)$$

The transfer functions $H(s)$ and $H_A(s)$ are low-pass, and $H_B(s)$ is high-pass (because the loop filter $F(s)$ is low-pass). Therefore,

- the phase noise in the reference signal $s_r(t)$ and the internal tension noise produced by the phase detector and the loop filter are low-pass filtered to the output signal $s_o(t)$;
- the internal phase noise produced by the VCO is high-pass filtered to the output signal $s_o(t)$.

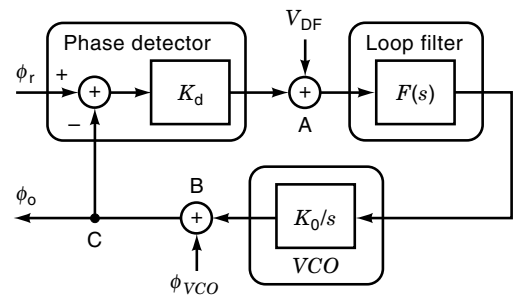


Figure 4. Linear model in the Laplace-transform domain of a phase-locked loop (PLL) with the main internal noise sources.

Key Parameters of PLLs. Four key parameters describe the performance of a PLL (here, the angular frequencies ω are considered, but the equivalent frequencies f can also be considered because $\omega = 2\pi f$):

- the *hold-in range* $\Delta\omega_{\text{HI}}$, defined as the maximum deviation of the frequency of the input signal $s_r(t)$ from the VCO free-run angular frequency ω_F within which the PLL can track slow (quasi-stationary) variations of the input frequency;
- the *pull-out range* $\Delta\omega_{\text{PO}}$, defined as the maximum deviation of the input frequency from ω_F within which the PLL can track fast variations (e.g., like steps) of the input frequency;
- the *lock-in range* $\Delta\omega_{\text{LI}}$, defined as the maximum deviation of the input frequency from ω_F within which the PLL can lock fast (i.e., faster than $1/\omega_F$ seconds) to the new input frequency;
- the *pull-in range* $\Delta\omega_{\text{PI}}$, defined as the maximum deviation of the input frequency from ω_F within which the PLL can lock in any time to the new input frequency.

Typically, $\Delta\omega_{\text{LI}} < \Delta\omega_{\text{PO}} < \Delta\omega_{\text{PI}} < \Delta\omega_{\text{HI}}$.

Operating Modes of Slaves Clocks. A slave clock may operate in the following three modes in a synchronization network:

- *Locked Mode.* The PLL tracks the input frequency, which is always within the hold-in and pull-out ranges. The lock mode may be *ideal*, if the reference signal is always available and stable, or *real* (stressed), if the reference is affected by impairments of various kinds.
- *Free-Run Mode.* If the reference signals fails, the clock works autonomously, supplying the free-run frequency ω_F of the internal oscillator (VCO) that works with null control tension. Generally, the nominal ω_F is designed equal to the nominal reference frequency.
- *Hold-Over Mode.* As in the free-run mode, if the reference signals fails, the clock works autonomously, supplying the frequency generated by the VCO. In this case, nevertheless, the last control tension value at the input of the VCO before the reference failure is maintained, so as to hold over the last output frequency value. More sophisticated clocks even store several subsequent samples of the VCO control tension, so that after entering the hold-over mode the VCO is controlled with variable tension extrapolated from the last data. Excellent hold-over accuracy can be achieved in this way, even under a substantial VCO nonlinear frequency drift.

Clock Stability Characterization

The background work on clock stability characterization is immense and broad in range. By the late 1960s, the pressure from clock manufacturers, time and frequency metrologists, and application engineers, all demanding a common set of frequency stability characterization parameters, led the IEEE to convene a committee to recommend uniform measures of frequency stability (2,16). Before and after this breakthrough, since the 1960s, several quantities have been defined aimed at characterizing clock stability. Although differing in ability, they highlight distinct phenomena in the phase, time, or frequency noise, or they are more oriented to some applications. A thorough and

detailed treatise on time and frequency stability characterization and measurement is beyond the scope of this article. Refs. 2, 3, and 17–24 provide further information about this topic.

Frequency-Domain Versus Time-Domain Characterization.

The characterization of clock stability is usually carried out by characterizing, by suitable analytical tools, the random processes $\varphi(t)$, $x(t)$, $\text{TE}(t)$, $y(t)$, or $v(t)$ [see Eq. (4), Eq. (9), and Eq. (11)].

Historically, a dichotomy became established between the characterization of such processes in the *Fourier frequency domain* and in the *time domain*. The inadequacy of measurement equipment strengthened the barriers between these two characterizations of the same noise process. Though these barriers are mainly artificial, nevertheless, it is not always possible to translate unambiguously from any quantity in one domain to any in the other.

Examples of stability measures in the frequency domain are the one-sided power spectral densities (PSDs, or more simply spectra) of phase, time, and frequency fluctuations, as functions of the Fourier frequency. On the other hand, the variances of the fluctuations averaged over an observation interval (e.g., the Allan variance), apart from the fact that they are evaluated starting from samples of time error or of instantaneous frequency, are examples of stability measures in the time domain as functions of the observation interval (time).

Clock Stability Characterization in the Frequency Domain: Power Spectral Densities. The most straightforward and intuitive way to characterize the stability of a timing signal $s(t)$ is to evaluate directly its two-sided spectrum, often denoted as $S_s^{\text{RF}}(f)$ (spectrum in radio frequency), or more precisely its low-pass representation. The variable f is the Fourier frequency. Positive values of f denote frequencies above ν_n . Negative values denote frequencies below ν_n . The spectrum $S_s^{\text{RF}}(f)$ is a continuous-time function proportional to the timing signal power per unit of bandwidth centered around f . The spectrum $S_s^{\text{RF}}(f)$ is definitely *not* a good tool to characterize clock frequency stability. Unfortunately, given $S_s^{\text{RF}}(f)$, it is not possible to determine unambiguously if the power at the various Fourier frequencies results from amplitude rather than phase fluctuations in the timing signal $s(t)$.

Commonly used stability measures in the frequency domain are, instead, the one-sided PSDs of $\varphi(t)$, $x(t)$, and $y(t)$, denoted as $S_\varphi(f)$ (rad^2/Hz), $S_x(f)$ (ns^2/Hz) and $S_y(f)$ (Hz^{-1}), respectively, because they describe the time and frequency stability characteristics directly. The following relationships hold:

$$\begin{aligned} S_y(f) &= \frac{f^2}{\nu_n^2} S_\varphi(f) \\ S_x(f) &= \frac{S_\varphi(f)}{(2\pi\nu_n)^2} \\ S_y(f) &= (2\pi f)^2 S_x(f) \end{aligned} \quad (17)$$

For a long time, the analog measurement of these PSDs in the frequency domain has been the main technique for studying the behavior of oscillators. More recently, the introduction of high-resolution digital instrumentation for the measurement in the time domain (time counters) made time-domain measurement more appealing in most applications. As a matter of fact, the

recent telecommunication standards recommend time-domain quantities evaluated starting from measured samples of time error as standard stability measures.

Clock Stability Characterization in the Time Domain: Telecommunications Standard Quantities. Although the previous frequency-domain characterization has proven very meaningful and complete in studying the behavior of oscillators, it is important to point out that the main concern in many modern applications, for example, in digital telecommunications, lies in controlling *time* deviations over given observation intervals (the buffer fill level in digital telecommunications equipment is proportional to the time error cumulated between the write and read clocks). Therefore, the time-domain stability quantities, basically a sort of prediction of the expected time and frequency deviations over an observation interval τ , are more oriented to this purpose.

Because of the prominent role of timing and synchronization in modern digital telecommunications and therefore the increasing interest in clock stability time-domain characterization far beyond the narrow circle of oscillator designers and time and frequency metrologists, this section focuses in particular on the time-domain quantities adopted for telecommunication standards to specify clock stability.

Among the several quantities defined in the literature for characterizing time and frequency stability, the following five in the time domain have been considered by telecommunication international standard bodies (25,26) for specifying timing interface requirements: the *Allan Deviation* (ADEV) $\sigma_y(\tau)$ square root of the Allan Variance (AVAR); the *Modified Allan Deviation* (MADEV) $\text{mod}\sigma_y(\tau)$ square root of the Modified Allan Variance (MAVAR); the *Time Deviation* (TDEV) $\sigma_x(\tau)$ square root of the Time Variance (TVAR); the *root-mean-square of Time Interval Error* (TIErms); and the *Maximum Time Interval Error* (MTIE). For the formal definitions of the first three quantities and the relevant theoretical background, the reader is referred to the works cited (2,3,19–24). The last two quantities, on the other hand, have been widely used by telecommunication engineers and look somewhat “exotic” to the traditional world of time and frequency metrology. A rather detailed analysis of their properties is available in Refs. 27 and 28.

For the practical purposes of telecommunications clock stability measurement (29), processes $\text{TE}(t)$ and $x(t)$ are considered synonymous [see Eqs. (9) and (11)]. Therefore, based on a sequence of N TE samples, defined as

$$x_i = x[t_0 + (i - 1)\tau_0] \quad i = 1, 2, \dots, N \quad (18)$$

that is measured with sampling period τ_0 over a measurement interval $T = (N - 1)\tau_0$ starting at an initial observation time t_0 , the following five standard estimators have been defined by the ITU-T (25) and ETSI (26) bodies:

$$\text{ADEV}(\tau) = \sqrt{\frac{1}{2n^2\tau_0^2(N-2n)} \sum_{i=1}^{N-2n} (x_{i+2n} - 2x_{i+n} + x_i)^2} \quad (19)$$

$$n = 1, 2, \dots, \left\lfloor \frac{N-1}{2} \right\rfloor$$

$$\text{MADEV}(\tau)$$

$$= \sqrt{\frac{1}{2n^4\tau_0^2(N-3n+1)} \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2}$$

$$n = 1, 2, \dots, \left\lfloor \frac{N}{3} \right\rfloor \quad (20)$$

$$\text{TDEV}(\tau)$$

$$= \frac{\tau}{\sqrt{3}} \text{MADEV}(\tau)$$

$$= \sqrt{\frac{1}{6n^2(N-3n+1)} \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2}$$

$$n = 1, 2, \dots, \left\lfloor \frac{N}{3} \right\rfloor \quad (21)$$

$$\text{TIErms}(\tau) = \sqrt{\frac{1}{N-n} \sum_{i=1}^{N-n} (x_{i+n} - x_i)^2} \quad n = 1, 2, \dots, N-1 \quad (22)$$

$$\text{MTIE}(\tau) = \max_{1 \leq k \leq N-n} \left[\max_{k \leq i \leq k+n} x_i - \min_{k \leq i \leq k+n} x_i \right]$$

$$n = 1, 2, \dots, N-1 \quad (23)$$

where $\tau = n\tau_0$ is the *observation interval* and $\lfloor z \rfloor$ denotes “the greatest integer not exceeding z .”

Common Types of Clock Noise

Experimental measurements on clocks may exhibit a wide variety of types of noise, either generated by physical processes intrinsic to the oscillator hardware or caused by external phenomena, such as environmental perturbations, mechanical vibrations, residual ripples in the power supply, signal coupling via power supplies and ground paths, and electromagnetic interference.

Power-Law Noises

In the frequency domain, the model most frequently used to represent the output phase noise measured on clocks is the so-called *power-law model* (3). In terms of the one-sided PSD of $x(t)$, this model is expressed by

$$S_x(f) = \begin{cases} \frac{1}{(2\pi)^2} \sum_{\alpha=-4}^0 h_{\alpha+2} f^\alpha & 0 \leq f \leq f_h \\ 0 & f > f_h \end{cases} \quad (24)$$

where the coefficients h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} are device-dependent parameters [the reason for the subscript $\alpha + 2$, for $\alpha = -4, -3, -2, -1, 0$, is that, historically, the coefficients h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} have been used in defining the power-law model in terms of $S_y(f)$], and f_h is an upper cutoff frequency, dependent mainly on low-pass filtering in the oscillator and in its output buffer amplifier. This clock upper cutoff frequency is

usually in the range of 10 kHz to 100 kHz in precision frequency sources (30).

The five noise types of this model are White Phase Modulation (WPM) for $\alpha = 0$; Flicker Phase Modulation (FPM) for $\alpha = -1$; White Frequency Modulation (WFM) for $\alpha = -2$; Flicker Frequency Modulation (FFM) for $\alpha = -3$; and Random Walk Frequency Modulation (RWFM) for $\alpha = -4$. All of the stability quantities cited are sensitive, according to different laws, to the presence of these noises in the timing signal (3,27).

According to this model, the PSD $S_x(f)$, plotted on a log-log diagram, results approximately in a broken line composed of straight segments, one per each noise type and each having a slope equal to the corresponding power α . In the time domain, on the other hand, the random realizations of the noise process of each single type have characteristic trends which can be recognized at a glance by an experienced eye. To give an idea, Fig. 5(a) through (e) shows sample realizations of TE affected, respectively, by the five types of power-law noise, simulated according to the procedure outlined in Ref. 31. Each realization is made of 4096 random TE samples, normally distributed and having a spectrum that obeys the law of Eq. (24) according to the value of α specified.

These noise processes may be from different physical causes (17) and, on a particular oscillator, they may all be recognized

or some may not. The main features and the origin of each of the five types of power-law noise are now summarized as follows:

Random Walk Frequency Modulation (h_{-2}/f^4). Difficult to measure as close to the carrier (the ideal timing signal). It is mostly ascribed to environmental effects. If RWFM noise dominates, then it is likely that frequent perturbations like mechanical shocks or temperature variations cause random shifts in the oscillation frequency.

Flicker Frequency Modulation (h_{-1}/f^3). The causes of this type of noise are not fully understood, but they are ascribed mostly to the physical resonance mechanism of an active oscillator or to phenomena in the control electronic devices. FFM noise is commonly recognized in high-quality oscillators but can be hidden by WFM or FPM noise in lower quality oscillators.

White Frequency Modulation (h_0/f^2). This is a type of noise commonly recognized in passive-resonator frequency standards, based on a slave oscillator (mostly a quartz) locked to a resonance of another device. Cesium-beam and rubidium standards feature a dominant WFM noise.

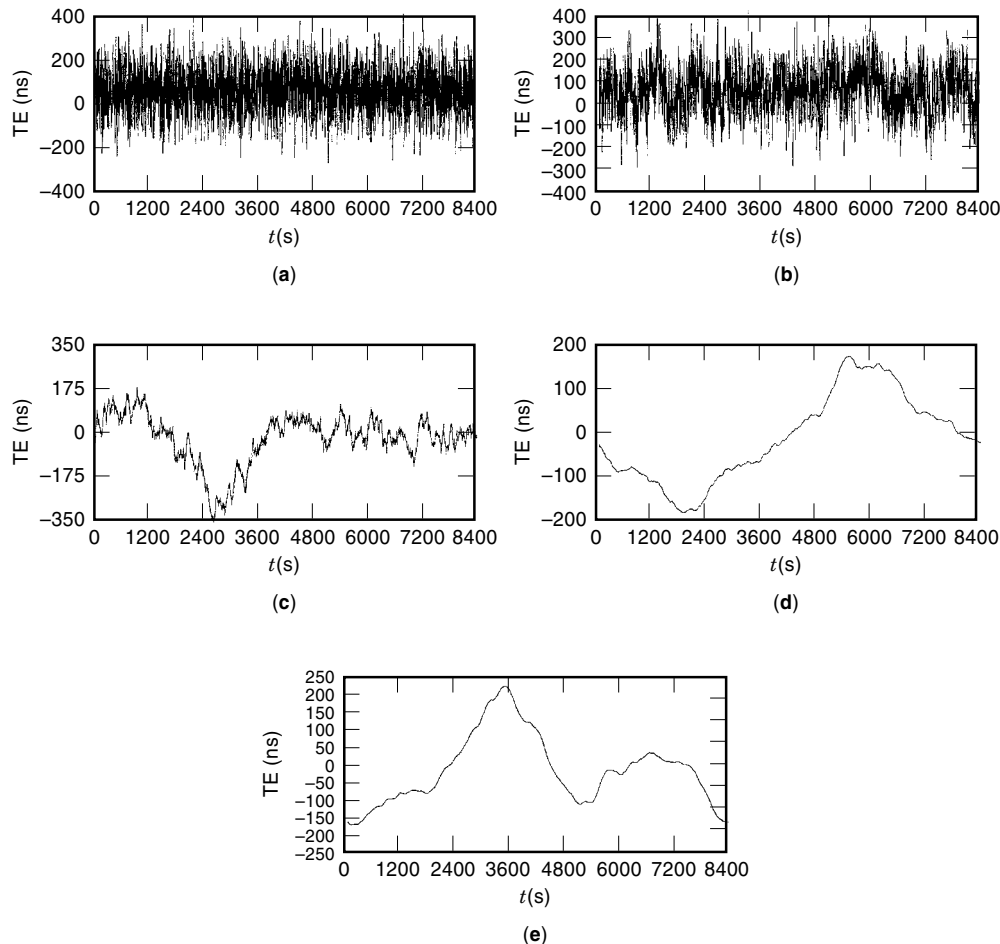


Figure 5. Simulated realizations of TE affected by the five types of power-law noise. © 1995 IEEE (32).

Flicker Phase Modulation (h_1/f^1). Although it may be related to a physical resonance mechanism of the oscillator, this noise is added mostly by noisy electronics, especially in the output amplification stages and in the frequency multipliers.

White Phase Modulation (h_2). This noise has little to do with the clock resonance mechanism, but it is added mainly by noisy electronics. In the past, this type of noise was often negligible in high-quality clocks, featuring very low-noise output stages. Nowadays, conversely, the spreading of clocks based on digital control electronics (such as the Digital PPLs, DPLLs) made the WPM noise the most commonly found in several applications, for example on telecommunications clocks, which are maybe the most impressive case of wide application of high-precision frequency sources. The substantial WPM noise in DPLLs is caused by the quantization error in the phase-lock loop, which produces a broadband white noise in the output timing signal. Moreover, WPM is the test-bench background noise caused by the trigger and quantization errors of time counters in the digital measurement of the TE.

Periodic Noise. Though the power-law model has proved very general and suitable for describing most measurement results, yet other types of noise may result in experimental measurements. Periodic noises are quite common. They may be typically caused by 50/60 Hz ac power line interference, diurnal temperature variations, or sensitivity to acoustic or mechanical vibrations, but they may be also due to intrinsic phenomena, such as special frequency control algorithms in DPLLs.

An example of real periodic noise is provided in Fig. 6, which shows the PSD $S_x(f)$ estimated from the TE data ($N = 79000$ samples spaced $\tau_0 = 23$ ms over a total measurement period of $T = 1800$ s) measured on a telecommunications equipment [synchronous digital hierarchy (SDH) STM-16 Line Terminal] slave clock in the synchronized clock configuration (25) according to the procedure outlined in Ref. 29. This PSD exhibits several discrete terms (spikes) in the frequency domain at harmonic frequencies of about 1 Hz, superposed on a power-law noise.

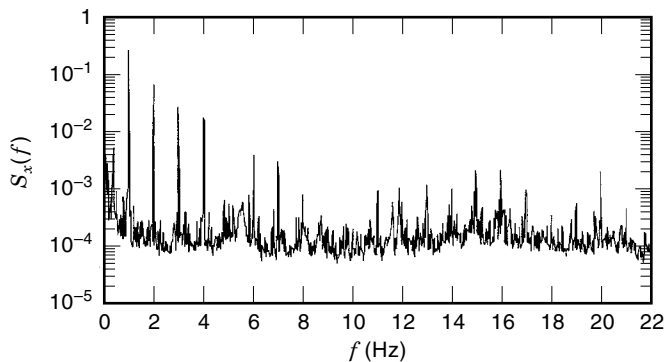


Figure 6. Experimental results featuring periodic noise. This graph shows the PSD $S_x(f)$ measured on a telecommunications equipment slave clock (SDH STM-16 Line Terminal). Several spikes are evident at harmonic frequencies of about 1 Hz. © 1997 IEEE (30).

TECHNOLOGY OF CLOCKS

Clocks are implemented by exploiting a physical mechanism of resonance generating a (pseudo-)periodic signal. In early clocks the oscillation was mechanical (e.g., pendulum clocks). Nowadays in high-precision clocks, on the other hand, the oscillation of atoms around their minimum-energy position in a quartz crystal or the atomic transition between two quantum energy levels is exploited. This section simply summarizes the physical operating principles of high-precision clocks. For a closer look at the physics and technology of quartz and atomic oscillators and their performance, the reader is referred to Refs. 32–37 and to the wide bibliography cited by the articles of Ref. 32.

Quartz Crystal Oscillators

Quartz oscillators are based on the piezoelectric effect discovered by P. Curie in 1880 (Nobel prize, 1903): a mechanical strain in the crystal yields an electrical field and vice versa. Thus a crystal oscillator (XO) is based on exciting a quartz crystal with a periodic electric signal at the resonance frequency (10 kHz to 100 MHz). The resulting resonator has a quality factor quite high (10^3 to 10^6) and, used in a feed-back circuit, allows generating a timing signal featuring excellent short-term stability (in particular over observation intervals smaller than one second). Because it performs well and is inexpensive, it is widely employed in most electronic equipment, for example, as voltage-controlled crystal oscillator (VCXO) in PLLs.

Temperature-Compensated Crystal Oscillators. The main problem with a plain XO is that its natural frequency depends on aging (around 10^{-7} /day in plain models) and on the temperature (typical values are on the order of $10^{-7}/^\circ\text{C}$ or above). To overcome the latter problem, temperature-compensated crystal oscillators (TCXOs) implement feedback control on the oscillation frequency, based on measuring the crystal temperature. Such a device achieves frequency stability of 10^{-7} over a temperature interval from 0° to 50°C . Through digital control, more sophisticated models achieve frequency stability on the order of 10^{-8} in the temperature interval from 0° to 70°C .

Oven-Controlled Crystal Oscillators. Far better than compensating temperature variations with a feedback control is to insulate the oscillator thermally and to make it work in a constant-temperature oven. Such clocks are called oven-controlled crystal oscillators (OCXOs). Thus frequency stability values exceeding 10^{-9} /day are achieved. State-of-the-art OCXOs, based on a double-oven temperature control and on a special technology of crystal excitation, may even achieve frequency stability on the order of 10^{-10} /day.

Performance and Characteristics of Crystal Oscillators. Table 1 summarizes some typical performance data and characteristics of quartz oscillators (for further data see Refs. 35–37). Performance is expressed in terms of the short-term stability (Allan variance) over one second, of the linear drift D (in parts $\Delta\nu_{\text{MAX}}/\nu_n$ per year), of the accuracy expected over one year ($\Delta\nu_{\text{MAX}}/\nu_n$) and of the temperature sensitivity. Moreover, the warm-up time is expressed in terms of the time needed to achieve the accuracy specified between brackets.

Table 1. Typical Performance Data and Characteristics of Commonly Available Quartz Oscillators

	XO	TCXO	OCXO
Short-term stability $\sigma_y(\tau = 1 \text{ s})$		1×10^{-9}	1×10^{-13} to 2×10^{-11}
Linear drift D	$>1 \times 10^{-6}/\text{year}$	$5 \times 10^{-7}/\text{year}$	$5 \times 10^{-9}/\text{year}$
Accuracy, 1 year	2×10^{-6} to 1×10^{-5}	2×10^{-6}	1×10^{-8} to 1×10^{-7}
Temperature sensitivity	$>1 \times 10^{-7}/^\circ\text{C}$	$>5 \times 10^{-7}$ (-55° to 85°C)	1×10^{-9} (-55° to 85°C)
Warm-up time		10 s (to 1×10^{-6})	1 hour (to 1×10^{-8})
Lifetime (performance guaranteed)	10 years to 20 years	>5 years	10 years to 20 years

Atomic Frequency Standards

The operating principle of passive atomic frequency standards is based on achieving a resonance frequency maximizing the number of atomic transitions between two quantum energy levels A and B (characterized by different magnetic moments of one unpaired electron). The energy difference $E_B - E_A$ between the two states is proportional to Planck's constant h and to the frequency ν_0 of the electromagnetic radiation that excites the transition, according to Bohr's law

$$E_B - E_A = h\nu_0 \quad (25)$$

A feedback control, aiming at maximizing the number of atomic transitions from A to B, then allows the system to synchronize on the frequency ν_0 (in the microwave range) within the frequency width given by Heisenberg's uncertainty principle. Hence, the accuracy of the output timing signal is determined solely by the atomic physical properties of the element adopted, that is, on fundamental constants which do not depend on space and time (within known relativistic effects).

Cesium-Beam Frequency Standard. The scheme of the principle of a basic cesium-beam frequency standard is shown in Fig. 7. An oven with a few grams of the isotope ^{133}Cs effuses cesium atoms uniformly distributed among 16 quantum energy levels, and the resonator is based on the transition between the levels characterized by magnetic moments $F = 4$ (level A) and $F = 3$ (level B). The nonhomogeneous magnetic field in magnet 1 (polarizer) deflects only the level B

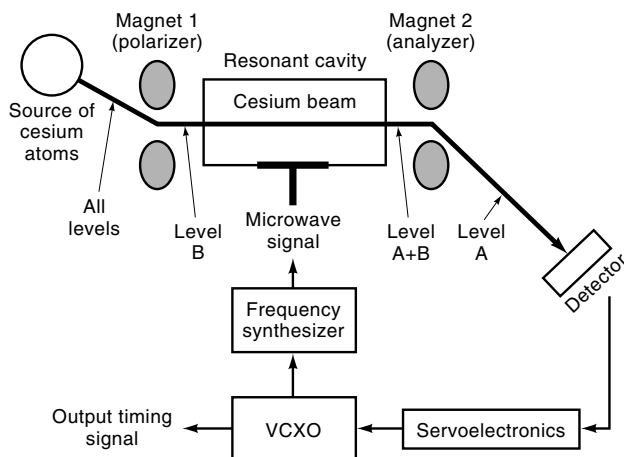


Figure 7. Scheme of the principle of a cesium-beam frequency standard.

atoms through the resonant cavity, where a microwave signal at frequency $\nu_0 = 9.192631770 \text{ GHz}$ stimulates their transition to level A. The nonhomogeneous magnetic field in magnet 2 (analyzer) deflects only the level A atoms to a detector. Thus the atom flux detected is proportional to the transition probability from level B to level A. Finally, the signal output by the detector is used to control a quartz VCXO, from which the microwave radiation and the output timing signal are synthesized, aiming at maximizing the number of transitions. Therefore the VCXO is locked to the atomic resonance frequency. The excellent short-term stability of the quartz oscillator is coupled with the excellent long-term stability of the atomic resonator.

Hydrogen MASER Frequency Standard. The operating principle of a hydrogen MASER (Microwave Amplification by Stimulated Emission of Radiation) frequency standard is based on the stimulated emission of electromagnetic radiation at the frequency ν_0 corresponding to the transition of hydrogen atoms between states that have magnetic moments $F = 1$ and $F = 0$. As shown in Fig. 8, the hydrogen atoms in the state $F = 1$ are deflected by a nonhomogeneous magnetic field of intensity 1 T into a storage bulb surrounded by a high- Q microwave resonant cavity, which is exposed to another internal magnetic field of intensity 10^{-7} T designed to resonate at the transition frequency $\nu_0 = 1.420450751 \text{ GHz}$. Then the hydrogen atoms decay to state $F = 0$ by bouncing inside the bulb, and microwave electromagnetic radiation at frequency ν_0 , stimulated by the magnetic field, is emitted and detected by an antenna. Finally, a quartz VCXO is kept locked to this signal to produce the output timing signal.

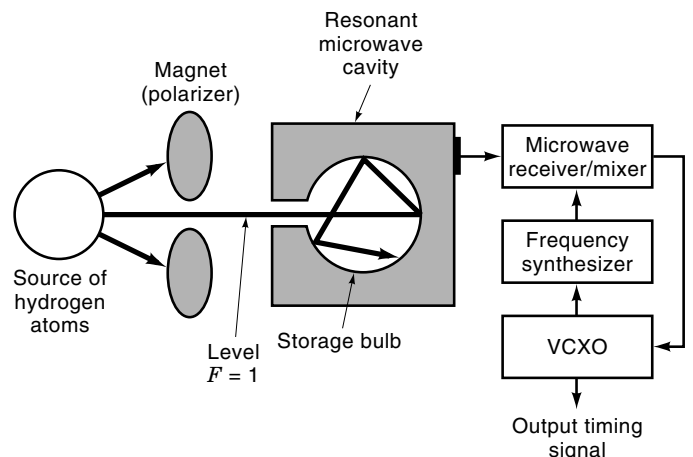


Figure 8. Scheme of the principle of a passive hydrogen MASER frequency standard.

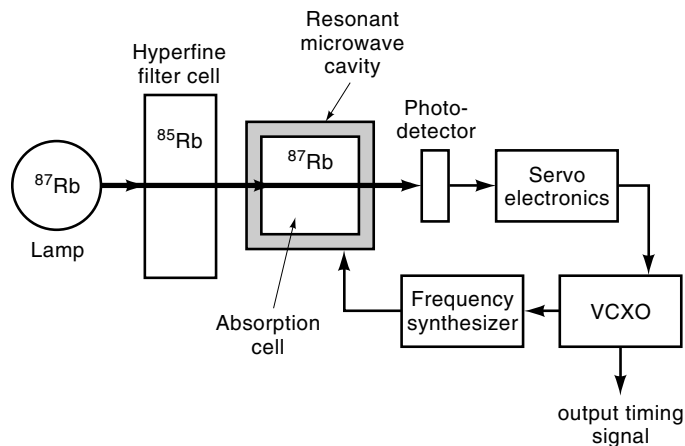


Figure 9. Scheme of the principle of a rubidium gas-cell frequency standard.

Rubidium Gas-Cell Frequency Standard. This frequency standard is based on the transition of atoms of ^{87}Rb between the base levels characterized by magnetic moments $F = 1$ and $F = 2$. A scheme of the principle of a rubidium gas-cell frequency standard is shown in Fig. 9. Light from a lamp filled with ^{87}Rb is filtered through a cell (hyperfine filter) containing ^{85}Rb vapor before it excites ^{87}Rb gas atoms in a cell (absorption cell) inside a resonant microwave cavity. The filter cell allows only the light spectrum component that excites the level $F = 1$ to reach the absorption cell, where this level is thus depopulated as ^{87}Rb atoms absorb the light radiation and migrate to the upper level and then down to both the base levels $F = 1$ and $F = 2$ (optical pumping). If the level $F = 1$ is depopulated by light absorption, the cell becomes transparent, but microwave radiation at $\nu_0 \cong 6.834682613$ GHz is applied to the atoms and excites the transition again to level $F = 1$. In resonance, when the frequency of the applied microwave radiation is exactly ν_0 , the signal at the photodetector shows a minimum. Thus the VCXO is driven accordingly to keep light absorption at its maximum.

The center frequency generated may deviate considerably (10^{-9}) from the theoretical value because of several kinds of environmental causes. For this reason, Rb frequency standards are not suited as *primary* standards, but need to be calibrated against cesium-beam or hydrogen masers, though their short-term stability is usually much better than that of Cs standards.

Performance and Characteristics of Atomic Frequency Standards. Table 2 summarizes typical performance data and characteristics of atomic frequency standards (for further data, see Refs. 35–37). As in Table 1, performance is expressed in terms of the short-term stability (Allan variance) over one second, of the linear drift D (in parts $\Delta\nu_{\text{MAX}}/\nu_n$ per year), of the accuracy expected over one year ($\Delta\nu_{\text{MAX}}/\nu_n$) and of the temperature sensitivity. Moreover, the warm-up time is expressed in terms of the time needed to achieve the accuracy specified between brackets.

The Global Positioning System

The Global Positioning System (GPS) is not actually a *clock*, but rather a complex system of clocks and satellites. Because it supplies, beyond the positioning service, the most accurate

timing reference signal available worldwide, it is treated in this section as it can be considered, broadly speaking, a sort of *superclock*.

GPS is a satellite radio system providing continuous and real-time three-dimensional position, velocity, and time information to suitably equipped users anywhere on the surface of the earth. Born essentially as a navigation and positioning tool, it is used also as a pure time reference (e.g., in telecommunications).

The NAVSTAR system operated by the US Dept. of Defense is the first GPS system available to civilian users. It has been in design since 1973 to take the place of the older navigation system LORAN-C and was completed in 1994, when the last satellites were launched. The Russian GPS system, called GLONASS, is very similar to NAVSTAR. These systems consist of three segments, the space, the control, and the user segments.

This section provides a quite general overview of GPS. A starting point for further details on GPS and its applications are Refs. 38 and 39. Nice and thorough overviews, moreover, are available on the Internet web, in particular at the addresses (40,41). Official pages on the web are provided by the US Air Force on NAVSTAR (42) and by the Russian Space Forces on GLONASS (43), together with almost real-time information (e.g., satellite status reports).

Space Segment. The space segment consists of a set of satellites (constellation), equipped with Cs or Rb atomic clocks, controlled by earth, and transmitting radio signals which can be received by user equipment.

The NAVSTAR constellation is composed of 24 satellites on 6 orbital planes. The satellites operate in 20200 km circular orbits with an inclination angle of 55° and a period of 12 h. The spacing of satellites in orbit is arranged so that a minimum of four satellites are in view to users worldwide at any time. Using spread-spectrum techniques, each satellite broadcasts a pair of L-band RF signals, that is, L1 (1575.42 MHz) and L2 (1227.6 MHz). The L1 signal carries a Precise (P) code and a Coarse/Acquisition (C/A) code, whereas L2 carries only the P code. The P code is normally encrypted so that only the C/A code is available to civilian users.

On the other hand, the GLONASS constellation is composed of 24 satellites on three orbital planes. The satellites operate in 19100 km circular orbits with an inclination angle of 64.8° and a period of 11 h and 15 min. Each satellite transmits on two L frequency groups (the L1 group centered on 1609 MHz and the L2 group centered on 1251 MHz) on a unique pair of frequencies. Unlike NAVSTAR, the GLONASS signals carry both a P code and a C/A code. The P code is encrypted for military use and the C/A code is available for civilian use.

Control Segment. The control segment is composed of all of the GPS facilities on the ground that monitor and control the satellites. The NAVSTAR system consists of five monitor stations, a master control station (MCS), and unlink antennas. The monitor stations (located at Colorado Springs, Hawaii, Ascension Island, Diego Garcia, and Kwajalein) passively track all GPS satellites in view, collect ranging data from each satellite, and send these raw data in real time to the MCS for processing. The MCS (located at Falcon Air Force Base, Colorado) evaluates clock and ephemeris changes in the satellites and uploads such correction data with other maintenance commands to each satellite once or twice each day by the ground-based

Table 2. Typical Performance Data and Characteristics of Commonly Available Atomic Frequency Standards

	Rubidium Oscillator	Cesium-Beam Oscillator	Hydrogen MASER Oscillator
Short-term stability $\sigma_y(\tau = 1 \text{ s})$	2×10^{-11} to 5×10^{-12}	$\sim 10^{-11}$	1×10^{-13} to 1×10^{-12}
Linear drift D	5×10^{-11} /year to 5×10^{-10} /year	0	$< 10^{-13}$ /year to 5×10^{-13} /year
Accuracy, 1 year	5×10^{-10}	2×10^{-11} to 1×10^{-14}	$\sim 10^{-13}$
Temperature sensitivity	$1 \times 10^{-11}/^\circ\text{C}$	$1 \times 10^{-13}/^\circ\text{C}$	$1 \times 10^{-14}/^\circ\text{C}$
Resonance frequency ν_0	6.3846826128 GHz	9.192631770 GHz	1.42040575177 GHz
Warm-up time	2 min to 5 min (to 5×10^{-10})	30 min (to 3×10^{-12})	24 hours (to 1×10^{-12})
Lifetime (performance guaranteed)	5 years to 10 years	5 years	3 years
Basic wear-out mechanism	Rb lamp and cavity life (15 years)	Cs beam tube (5 years)	Ion pumps and H_2 source depletion, cavity resonance shift (7 years)
Portability and intended location	Space, air, ground	Ground and lab. oriented	Definitely ground and lab. oriented
Average size	8 cm $\times$ 8 cm $\times$ 12 cm	50 cm $\times$ 30 cm $\times$ 14 cm	50 cm $\times$ 30 cm $\times$ 50 cm
Weight	0.5 kg to 2 kg	10 kg to 30 kg	> 50 kg
Power consumption	10 W to 50 W	25 W to 50 W	70 W to 100 W

uplink antennas (located in Ascension Island, Diego Garcia, and Kwajalein).

User Segment. A GPS receiver calculates its four-dimensional position (space and time) on the basis of the radio signals received from at least four satellites in view. GPS receivers come in many different sizes, shapes and price ranges, according to their performance and the application for which they are intended. For positioning, for example, inexpensive palm-receivers determine their positions with 95% accuracy (i.e., the value of two standard deviations of radial error from the actual antenna position under specified test measurement conditions) on the order of 100 m, whereas highly sophisticated receivers for military applications may achieve even subcentimetric resolution.

Use of GPS as Timing Reference. Because of the MCS control, GPS provides a timing signal traceable to the Universal Time Coordinated (UTC). Its main problem is very poor short-term stability (95% time accuracy is guaranteed to be 340 ns for the NAVSTAR civilian standard positioning service) together with the possibility of transitory local signal unavailability. Therefore, GPS timing supplies are equipped with a slave quartz oscillator to couple the excellent short-term performance of the latter with the long-term accuracy of GPS. One of the most important applications of GPS receivers as timing supplies is in telecommunications networks to synchronize the nodes of networks spread over long distances.

Clocks in Digital Telecommunications Networks

As mentioned previously, clocks and network synchronization play a central role in modern digital telecommunications (5,6) because they have a determining influence on the quality of the services provided. Clocks used in telecommunications are mainly primary reference clocks and slave clocks in synchronization networks, equipment clocks, and real-time clocks. The next subsections give some detail on this topic.

Architecture of Synchronization Networks for Digital Telecommunications Networks

As mentioned previously in the section Network Synchronization Strategies, among the several network synchronization

strategies conceived, the HMS strategy is the most widely adopted worldwide for synchronizing modern digital telecommunications networks (and it is indicated as the standard choice by Ref. 44). The master-slave architecture is organized hierarchically in at least two levels:

- at level 0, one (or more, for reliability) master *Primary Reference Clock* (PRC) generates the reference signal by running in autonomous mode;
- at the lower levels, slave clocks are synchronized by the signals coming from the upper level (thus traceable to the PRC) and synchronize the clocks of the lower level.

Such slave clocks are called *Synchronization Supply Unit* (SSU) or *Stand-Alone Synchronization Equipment* (SASE) in Europe and in the ITU-T Recommendations. In North America they are commonly known as *Building Integrated Timing Supply* (BITS). These clocks are deployed in telecommunications offices to distribute their timing signal to all of the equipment in the building.

Each piece of equipment in a telecommunications network is provided with an *Equipment Clock* (EC) that distributes the timing to all of the cards and modules of the apparatus. The task of a synchronization network is to synchronize all the equipment clocks in the telecommunications network exactly.

The timing is carried from one clock to another of the synchronization network through *synchronization trails*, which may be direct digital links or even chains of SDH/SONET equipment clocks (44). Direct digital links are usually primary-level signals of the plesiochronous digital hierarchy (PDM) that is the European 2.048 Mb/s (E1) or the North-American/Japanese 1.544 Mb/s (T1), in some cases through a multiplexer-demultiplexer chain. Alternatively, they may be SDM/SONET signals (46) on optical fibers. Within an office, on the other hand, timing is usually distributed as analog 2.048 MHz signals.

Clocks in Synchronization Networks

According to their different roles in a synchronization network, PRCs and SASEs pose different requirements to designers. They are specified, respectively, by the new ITU-T Recommendation G.811 and G.812, which substitute for the previous ones, denoted by the same number, in the Blue Book series.

PRCs are autonomous clocks which are masters of the entire network (level 0). Therefore, they must have the highest long-term accuracy and stability. *Cesium-beam clocks* are well suited to this and therefore are adopted as a national reference.

SASEs are slave clocks designed to filter effectively the synchronization impairments in the incoming reference and have very good or even excellent stability in the hold-over mode, should all of the references fail. They pose different requirements according to their level in the synchronization network. At level 1, *rubidium clocks* are usually adopted, because of their excellent long-term stability coupled with lower cost compared to cesium clocks. At level 2 and below, *quartz OCXO clocks* are mostly adopted. The loop time constant of SASEs is usually on the order of 100 s to 10000 s.

GPS receivers are often used as additional reference at the input of SASEs (as noted previously, many national Administrations do not accept GPS as a first-choice primary reference for political reasons, because it is controlled by a foreign government), especially in very wide area networks to face the issue of timing transfer over long distances.

Equipment Clocks

ECs are designed to run mainly in the slave mode and take their reference either from a synchronization network or from another EC (e.g., the previous EC in a chain). Therefore, on the one hand, their long-term stability in hold-over is specified more loosely than for PRCs or SASEs. On the other hand, their noise filtering capabilities meet different requirements according to the specific type of equipment and network architecture.

As a first example, the ECs of user primary digital multiplexers, which are located at the lowest level of the synchronization hierarchy, are synchronized by an external reference or an incoming digital multiplex signal but do not distribute their timing to any other equipment. Their clocks are usually inexpensive quartz oscillators (XO) with poor accuracy in free run (just better than 50 $\mu\text{Hz}/\text{Hz}$).

The clocks of digital switching exchanges meet tighter requirements because they need to control the slips in the input digital signals by a suitable network synchronization strategy (usually HMS). Most commonly, such exchanges are equipped with high-precision OCXOs whose loop time constant is set to at least 100 s (1000 s is common) to filter out a substantial amount of phase noise on the input reference.

Finally, SDH Equipment Clocks (SECs) are the ECs which must comply with the tightest and best specified requirements. The ITU-T Recommendation G.813 (46) is devoted to their specification, especially as for their output noise and noise filtering capabilities. SECs are usually TCXOs, with a free-run accuracy better than 4.6 $\mu\text{Hz}/\text{Hz}$ and loop bandwidth on the order of 1 Hz to 10 Hz.

Real-Time Clocks

Real-time clocks are a very special kind of equipment clock, strictly related to network and equipment management. Modern telecommunications equipment performs quite sophisticated management functions (e.g., configuration and fault management, performance monitoring) through a local terminal or a remote control center which may even manage several networks of different kinds.

Management messages sent to the local terminal or to the remote control center must include the key information, the

date and time (with one second resolution) on which the fault, the data error, the configuration change, etc. occurred, to allow the network manager to correlate events reported from different nodes and after variable delays. This information is given in each node by the local real-time clock.

It is essential that the real-time clocks of the whole network be synchronized, otherwise it would become impossible to correlate meaningfully different messages (which may be *numerous* and come from different parts of the network) under a common label (i.e., the actual event that happened or the issue that the network manager is interested in knowing). Synchronizing the real-time clocks of a network is a very special issue of network synchronization.

The network synchronization addressed to SASEs and EC has the target of minimizing time error fluctuations among the clocks and mostly achieves deviations not greater than 10 ns or 100 ns, regardless of the starting phase offset. This implies that *synchronous timing signals* (e.g., sine waves) are distributed.

The network synchronization addressed to real-time clocks, on the other hand, is conceptually different: the *absolute time* information (e.g., 23 Dec 1998, 01.32.04 AM) is needed, but a time alignment error of a few ms can be allowed. Here, synchronization is often distributed through software messages carrying the date-and-time message to the network nodes, according to suitable algorithms, as reported, for example, in Refs. 47–49.

BIBLIOGRAPHY

1. G. M. R. Winkler, Timekeeping and its applications, *Adv. Electron. Electron Phys.*, **44**: 1977.
2. J. A. Barnes et al., Characterization of frequency stability, *IEEE Trans. Instrum. Meas.*, **IM-20**: 105–120, 1971.
3. J. Rutman and F. L. Walls, Characterization of frequency stability in precision frequency sources, *Proc. IEEE*, **79**: 952–960, 1991.
4. W. C. Lindsey et al., Network synchronization, *Proc. IEEE*, **73**: 1445–1467, 1985.
5. P. Kartaschoff, Synchronization in digital communications networks, *Proc. IEEE*, **79**: 1019–1028, 1991.
6. J. C. Bellamy, Digital network synchronization, *IEEE Commun. Mag.*, **33** (4): 70–83, 1995.
7. J. Buck and E. Buck, Synchronous fireflies, *Sci. Amer.*, May, 1976.
8. F. M. Gardner, *Phaselock Techniques*, New York: Wiley, 1967.
9. R. E. Best, *Phase Locked Loops*, New York: McGraw-Hill, 1984.
10. V. F. Kroupa, Noise properties of PLL systems, *IEEE Trans. Commun.*, **COM-30**: 2244–2252, 1982.
11. P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge, England: Cambridge University Press, 1980, Sects. 9.28–9.33.
12. J. P. Moreland, Performance of a system of mutually synchronised clocks, *Bell Syst. Tech. J.*, **50** (7): 2449–2464, 1971.
13. W. C. Lindsey and J. H. Chen, Mutual clock synchronization in global digital communication networks, *Eur. Trans. Telecommun.*, **7** (1): 25–37, 1996.
14. J. A. Barnes, The measurement of linear frequency drift in oscillators, *Proc. 15th Annu. Precise Time Time Interval PTTI Meet.*, 1983.
15. R. L. Filler and J. R. Vig, Long term aging of oscillators, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **40**: 387–393, 1993.
16. IEEE Stand. 1139, *IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology*, approved October 20, 1988.

17. D. A. Howe, D. W. Allan, and J. A. Barnes, Properties of signal sources and measurement methods, *Proc. 35th Annu. Freq. Control Symp.*, 1981.
18. S. R. Stein, Frequency and time—their measurement and characterization, in E. A. Gerber and A. Ballato (eds.), *Precision Frequency Control*, New York: Academic Press, 1985, vol. 2, chap. 2, pp. 191–232.
19. D. W. Allan, Statistics of atomic frequency standards, *Proc. IEEE*, **54**: July, 1966.
20. L. S. Cutler and C. L. Searle, Some aspects of the theory and measurement of frequency fluctuations in frequency standards, *Proc. IEEE*, **54**: February, 1966.
21. W. C. Lindsey and C. M. Chie, Theory of oscillator instability based upon structure functions, *Proc. IEEE*, **64**: 1652–1666, 1976.
22. P. Lesage and C. Audoin, Characterization of frequency stability: Uncertainty due to the finite number of measurements, *IEEE Trans. Instrum. Meas.*, **IM-22**: 157–161, 1973.
23. P. Lesage and T. Ayi, Characterization of frequency stability: Analysis of the modified Allan variance and properties of its estimate, *IEEE Trans. Instrum. Meas.*, **IM-33**: 332–336, 1984.
24. L. G. Bernier, Theoretical analysis of the modified Allan variance, *Proc. 41st Annu. Freq. Control Symp.*, 1987.
25. ITU-T Rec. G.810, *Definitions and Terminology for Synchronisation Networks*, Geneva, 1996 and later editions.
26. ETSI Draft ETS 300 462, *Generic Requirements for Synchronisation Networks*, Bonn, 1996.
27. M. Carbonelli, D. De Seta, and D. Perucchini, Characterization of timing signals and clocks, *Eur. Trans. Telecommun.*, **7** (1): January–February, 1996.
28. S. Bregni, Measurement of maximum time interval error for telecommunications clock stability characterization, *IEEE Trans. Instrum. Meas.*, **IM-45**: 900–906, 1996.
29. S. Bregni, Clock stability characterization and measurement in telecommunications, *IEEE Trans. Instrum. Meas.*, **46**: 1284–1294, 1997.
30. F. L. Walls and A. De Marchi, RF spectrum of a signal after frequency multiplication: Measurement and comparison with a simple calculation, *IEEE Trans. Instrum. Meas.*, **IM-24**: 210–217, 1975.
31. S. Bregni, L. Valtriani, and F. Veghini, Simulation of clock noise and AU-4 pointer action in SDH equipment, *Proc. IEEE GLOBECOM '95*, Singapore, 1995.
32. Special Issue on Time and Frequency, *Proc. IEEE*, **79** (7): July, 1991.
33. E. A. Gerber and A. Ballato (eds.), *Precision Frequency Control*. New York: Academic Press, 1985.
34. J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency Standards*, Bristol: Adam Hilger, 1989, Vol. 1.
35. J. Vig, *Quartz Crystal Resonators and Oscillators—A Tutorial*, Tech. Rep. SLCET-TR-88-1, Fort Monmouth, NJ: U.S. Army Electron. Technol. Devices Lab., 1992.
36. *Precision Time and Frequency Handbook*, Irvine, CA: Ball Corp., Efratom Time and Frequency Products, 1993.
37. *Handbook Selection and Use of Precise Frequency and Time Systems*, Geneva: ITU-R, 1997.
38. I. A. Getting, The global positioning system, *IEEE Spectrum*, **30** (12): 36–38, 43–47, 1993.
39. E. Kaplan, *Understanding GPS: Principles and Applications*, London: Artech House, 1996.
40. J. T. Beadles, *Introduction to GPS Applications*, 1997. Available: URL: <http://ares.redsword.com/gps/>
41. P. H. Dana, *Global Positioning System Overview*, Department of Geography, University of Colorado at Boulder, 1999. Available: URL: <http://www.colorado.edu/geography/gcraft/notes/gps/gps.f.html>
42. US Air Force, NAVSTAR GPS HOMEPAGE. Available: URL: <http://www.laafb.af.mil/SMC/CZ/homepage/>
43. Russian Space Forces, *Welcome to the Coordinational Scientific Information Center of the Russian Space Forces!* Available: URL: http://mx.iki.rssi.ru/SFCSIC/SFCSIC_main.html
44. ITU-T Rec. G.803, *Architectures of Transport Networks Based on the Synchronous Digital Hierarchy SDH*, Geneva, 1993, §6.
45. IRU-T Revised Rec. G.707, *Network Node Interface for the Synchronous Digital Hierarchy SDH*, Geneva, 1996.
46. ITU-T Rec. G.813, *Timing Characteristics of SDH Equipment Slave Clocks SEC*, Geneva, 1996.
47. J. Levine, An algorithm to synchronize the time of a computer to universal time, *IEEE/ACM Trans. Network.*, **3**: 42–50, 1995.
48. S. R. Jefferts et al., Two-way time and frequency transfer using optical fibers, *IEEE Trans. Instrum. Meas.*, **IM-46**: 209–211, 1997.
49. D. L. Mills, Precision synchronization of computer network clocks, *ACM/SIGCOMM Comput. Commun. Rev.*, **24** (2): 28–43, 1994.

STEFANO BREGNI
Politecnico di Milano

NOISE AND DISTORTION

ELECTRIC NOISE MEASUREMENT

The pursuit of extremely sensitive electronic measurements is a common goal among modern scientists and engineers. The limit to the signal energy that may be detected is a result of the noise energy that competes with the signal in a detection scheme. Thus, the measurement and accurate determination of the noise energy is of crucial importance when designing or assessing any signal detection system. Noise may be defined as an unwanted disturbance superimposed upon a useful signal, which tends to obscure the signal's information content. Electronic noise limits the signal detection sensitivity of a wide range of systems such as radar, ground based and satellite communication, cellular phones, and guidance and tracking systems, as well as ultrasensitive physical measurements such as radio astronomy and gravitational wave detection.

Many different types of electronic noise have been studied and characterized. Some types are fundamental in nature and can be derived from basic physical principles. Other types of noise may be classified as technical and are the result of the electronic configuration of the read-out system of the detection or measurement process. To reduce the interference from fundamental noise sources, the physical principle should be understood so that the particular detection scheme may be optimized to minimize the noise. To reduce the interference from technical noise calls for good electronic design that maintains the linearity and stability of the electronic read-out across the dynamic range of operation and the required frequency band.

A common figure of merit that is used when detecting a signal is known as the signal-to-noise ratio. It is defined as the ratio of the signal power to the noise power over the frequency band of the detection system. Filtering techniques exist to maximize the signal-to-noise ratio and are known as optimum filtering. This technique relies on knowing the form of the signal and biases the detection data at the frequencies in which the signal is strongest. However, this article focuses on the common types of electronic noise and the measurement and characterization techniques. Thus, the ways of maximizing the signal-to-noise ratio via filtering techniques will not be discussed, for further information on these methods see Refs. 1–4.

To measure the noise characteristics in an electronic system accurately, we usually use a device that can be referred to as a *null instrument*. Classic devices include a bridge circuit based on a mixer, the Dicke microwave radiometer, and the Michelson optical interferometer. These devices nominally have zero output and are highly sensitive to changes in the power level or phase delay; thus, they are able to make very sensitive measurements of either the power or phase fluctuations in the system. Modern applications of these instruments to state-of-the-art physical experiments include the study of the cosmic microwave background radiation (COBE project) (5), as well as the use of advanced microwave bridge techniques and large-scale optical interferometers for the search for gravitational waves from astrophysical sources (NIOBE, LIGO, and VIRGO projects) (6,7).

Electric noise measurements may be done in the presence or absence of a carrier signal. Modern communication systems make use of carrier signals to transmit information. For

example, mobile telephone systems typically operate at a frequency around 1 GHz. The voice frequencies modulate the carrier frequency as shown in Fig. 1, then the signal is transmitted through the atmosphere by the carrier at 1 GHz. At the receiving end, the signal is demodulated back to audio frequencies. The audio signal competes with the noise in the system, and as long as the signal power is much larger than the noise power, the signal can be heard at the receiver end. There is more than one way to characterize the noise in such a system. When there is a carrier signal present, it is common to measure the noise with respect to the carrier power as a ratio. This means just a relative measurement with respect to the carrier is made, which is comparatively easy to calibrate. If no carrier is present, then it is necessary to measure the absolute noise energy present. This requires a bit more effort in calibration. It is usual to equate the noise energy E_N to an effective noise temperature T_N via Boltzmann's constant, $E_N = k_B T_N$. For a system in equilibrium, this gives the fundamental relationship between the associated temperature and the energy. Basically, the more energy inside any system, the hotter it is. Likewise, when we characterize the electronic noise in a system, the higher the level of noise energy in the system, the higher the system effective noise temperature.

Many individual components make up a modern system. These include amplifiers, attenuators, mixers, and phase shifters. Each component can add to the overall system noise temperature. The common figure of merit that defines the performance of a component is the noise figure, which is equal to the ratio of the output noise density of an input source and that added by the component divided by the source alone (a full mathematical definition is given later). This type of noise characterization only characterizes the component in its small signal input regime (i.e., when no carrier signal is present). If a significant carrier signal is present, then the input power may be considered large, and the noise temperature and figure become functions of carrier power and offset frequency from the carrier. This is because many devices are nonlinear, and nonlinear up-conversions of the signal can cause a frequency dependence known as flicker noise, which can also be power-dependent. If a minimum noise temperature is essential, then

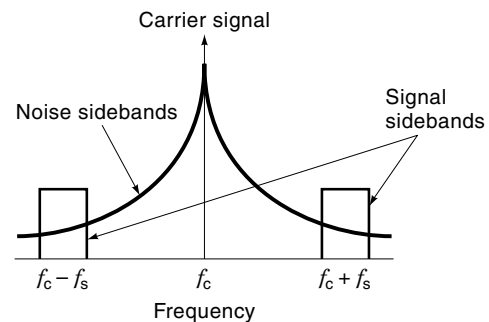


Figure 1. Spectral representation of a carrier signal with signal sidebands modulating the carrier in the presence of noise sidebands. f_C represents the carrier frequency and f_s represents the band of audio frequencies that modulate the carrier.

we must determine the restriction to the input power that is required to maintain small signal noise performance (8).

If a communication system is designed well and is not limited by technical noise, then it usually will be limited by the noise temperature of the components in the system. The noise power in this regime is independent of the carrier power. Thus, if the carrier power can be increased without creating any noise, the signal-to-noise ratio will increase as a result of the associated increase of signal power. Recently new interferometric techniques have allowed the detection of noise in the presence of a microwave carrier to be measured at high carrier powers close to a watt, with a noise performance still governed by the system noise temperature (8,9). This technique has allowed for a factor of 10,000 improvement in sensitivity of noise energy measurements in the presence of a carrier and is expected to have a significant impact on radar and metrology design for the future.

A carrier signal is a sinusoidal tone of single frequency f_c with phase and amplitude. Thus the signal frequencies f_s may be encoded with respect to the phase or amplitude of the carrier signal, and are known as phase modulation (PM) or amplitude modulation (AM). These two “quadratures” of modulation are ideally orthogonal and, thus, are independent and may exhibit vastly different levels of noise. When noise is present in the phase quadrature, it is known as *phase noise*; when noise is present in the amplitude quadrature, it is known as *amplitude noise*. Technical noise sources usually modulate the carrier with a constant fraction of phase or amplitude independent of the carrier power. Thus, if the carrier power is increased, both the signal sidebands and the noise sidebands increase, with the signal-to-noise ratio remaining constant. This is why it is useful to characterize the phase or amplitude noise in the presence of a carrier signal by a power ratio. However, it is not the purpose of this article to dwell on AM and PM signal encoding and noise measurement techniques. More on these topics can be found in the article on Measurement of Frequency, Phase Noise and Amplitude Noise. In this article, the focus is on the measurement and characterization of the small signal component noise, such as the more physically fundamental electric noise sources—thermal noise and shot noise—in the absence of a carrier signal.

Manufacturers of microwave components usually classify the noise performance in terms of the noise figure or temperature in the absence of a carrier (it is much harder to characterize the nonlinear flicker component in the presence of a carrier). Measurements of this kind are thus commonplace, and commercial noise figure test sets are available. High accuracy is available when characterizing modular components based on coaxial or waveguide input and outputs. However, the electronics industry has moved toward miniaturizing components, and many microwave circuits are produced not in modular form but rather are manufactured as part of one integrated circuit collectively called a *wafer*. Accurate characterization techniques of components “on-wafer” remains a principal concern of the metrology community and industry.

It is apparent that the accurate determination of the level of noise in a system is very important. Measurements of the noise temperature have become an art as many standards laboratories around the world strive to make their measurements of the absolute noise more accurate. The advent of the Dicke radiometer has made this possible (10). This article will focus only on the principles of noise types and measurements. To understand

the principles of the radiometer the reader is referred to the article on Radiometers. Recently, in the quest for more accurate measurements, an international comparison of noise temperature measurements has been undertaken between 2 GHz and 12 GHz. The institutes that took part in this comparison were: the National Institute of Standards and Technology, Boulder, Colorado, United States; the Laboratoire Central des Industries Electriques, Fontenay-aux Roses, France; the National Physical Laboratory, Teddington, United Kingdom; the Physikalisch-Technische Bundesanstalt, Braunschweig, Germany (11). The reader is also referred to the home page of these institutes for more details of state-of-the-art measurement techniques (12–15).

MATHEMATICAL REPRESENTATIONS OF NOISE

Noise is a fluctuating quantity that is described statistically by a random variable representing a fluctuating current, voltage, or power with time. When defining a random variable $X(t)$, it is important to have a measure of the extent of the fluctuations. A quantity that achieves this is known as the variance of X . To define the variance, the mean of the fluctuating quantity is first defined as $\langle X \rangle$ (it is common to use angle brackets to represent the mean of a quantity). Then a new variable representing the fluctuation about the mean can be defined, $\Delta X = X - \langle X \rangle$. Taking the mean square of this quantity defines the variance $\langle \Delta X^2 \rangle$. To calculate this quantity, we must know the probability density function $h(X)$ that describes the process. If $h(X)$ and thus the mean and variance are independent of time, the process is known as stationary.

For our purposes, the time domain representation given previously has limited use, and it is informative to introduce the Fourier method and represent $X(t)$ as a spectral density in the frequency domain. The relation between the frequency and time domains is given by the well-known Wiener–Khinchine theorem (16,17). Many undergraduate textbooks deal with this important theorem; for more details see Refs. 1, 18, and 19. The main result for our purposes is a particular case of the Wiener–Khinchine theorem that gives

$$\langle X^2 \rangle = \langle \Delta X^2 \rangle = \int_0^\infty S_x(f) df \quad (1)$$

for an independent stationary random process. Here, $S_x(f)$ is defined as the single-sided spectral density of the random variable, and we refer to f as the Fourier frequency. Equation (1) assumes that $\langle X \rangle = 0$, which is true for a sinusoidal variable. This term represents a direct current (dc) term and is not measured with a modern spectrum analyzer. In effect, the spectrum analyzer will set $\langle X \rangle$ to zero even if it is not.

Equation (1) gives an important result because it means that the spectral density of electronic noise may be measured very easily in principle. This can be done with a quadratic detector and a narrow band amplifier centered at the particular frequency f_0 . Assuming that the bandwidth of the amplifier B_r is small in comparison to the variation in frequency of the spectral density, a constant value of $S_x(f_0)$ can be assumed. From Eq. (1), the quadratic detector will measure a voltage at the output of the amplifier equivalent to

$$\langle v^2 \rangle = G S_x(f_0) B_r \quad (2)$$

and the spectral density is easily determined. Here, G is the power gain of the amplifier, and B_r is defined as the resolution bandwidth of the spectrum analyzer.

DEFINITIONS

Flicker and White Noise

White noise is a random noise with a flat frequency spectral density over the Fourier frequency range of interest. This type of noise is usually produced in the laboratory for testing purposes. For example, white noise generators such as hot or cold resistors are used as calibrated noise sources to calibrate noise measurement systems. In general, noise in system circuits is not white and has a frequency dependence. An example of a nonwhite noise is flicker noise, which has a spectrum proportional to $1/f$ and is prevalent in many circuit systems. Generally, at large values of Fourier frequencies, system noise is white; at small values of Fourier frequencies, system noise is flicker. The frequency at which the transition occurs is known as the flicker corner, which is typically but not necessarily about 10^5 Hz.

Signal-to-Noise Ratio

In general, the signal-to-noise ratio (SNR) in a system is defined as the ratio of available signal power to available noise power present in the system over the bandwidth of detection and is given simply by

$$\text{SNR} = \frac{\text{Signal power}}{\text{Noise power}} \quad (3)$$

In general, this value is dependent on the detection bandwidth. For example, if we detect a sinusoidal frequency of root mean square (rms) power S in the presence of white noise with a power spectral density of S_N (W/Hz), then the signal-to-noise ratio is given by

$$\text{SNR} = \frac{S}{BS_N} \quad (4)$$

where B is the detection bandwidth. Thus, the larger the bandwidth, the larger the detected noise power, which degrades SNR. In general, the signal might consist of more than one frequency and be more complicated than a single sinusoid frequency. To optimize the SNR, a standard result in signal detection theory states that the signal-to-noise ratio is optimized by a filter that has a transfer function proportional to the complex conjugate of the signal Fourier transform divided by the total noise spectral density (1). This technique accepts frequencies where the signal is large and the noise is low and filters out frequencies where the signal is low and the noise is large. Thus, it is prudent to know the signal bandwidth when designing a detection system.

Noise Temperature

From the equi-partition theorem, a gas molecule possesses a mean energy of $k_B T/2$ J/DOF and applies also to electronics, where T is the equilibrium temperature and k_B is Boltzmann's constant. Assuming that an electronic circuit has two degrees of freedom (i.e., amplitude and phase or resistance and

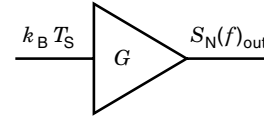


Figure 2. Schematic of a two-port device under test with the associated input and output.

reactance), then the mean noise energy in a passive circuit is $k_B T$ J. Because thermal noise is white, we may express this noise energy as a power spectral density given by $S_N(f) = k_B T$ W/Hz over all Fourier frequencies. Thus, the noise temperature at a selected frequency and reference point in the circuit may be defined as

$$T_N = S_N(f)/k_B \quad (5)$$

where $S_N(f)$ is the noise power spectral density measured or calculated at the circuit reference point and frequency. For a resistor, the noise temperature is approximately the actual temperature, whereas that observed by a diode or a transistor may be significantly larger than the absolute temperature.

Standard Reference Temperature

The standard reference temperature T_S for noise measurements is 290 K. This is an arbitrary choice; however, historically this has been chosen because it is close to room temperature.

Noise Figure

The spot noise figure (or narrow band noise figure) is defined as the ratio of the total noise power spectral density delivered to the output of the system divided by that caused by the source alone, when the source is at the standard reference temperature of 290 K. Assuming that the gain of the system is G (which may be fractional if the system has an overall attenuation), then the noise figure at a particular frequency is defined as (see Fig. 2)

$$F(f) = \frac{S_N(f)_{\text{out}}}{Gk_B T_S} \quad (6)$$

To calculate the overall noise figure $\overline{CIF}$, the spot noise figure is averaged over the frequency response of the system and is given by (18)

$$\overline{F} = \frac{\int_0^\infty F(f)G(f)df}{\int_0^\infty G(f)df} \quad (7)$$

Here $G(f)$ is now the system transfer function.

If a device is noiseless, then $S_N(f)_{\text{out}} = Gk_B T_S$, and both $\overline{CIF}$ and $F(f)$ are unity. However, for an active or dissipative device, there is some associated noise, and the noise figure is usually greater than unity. It is usual to compare noise figures in decibels ($10 \log_{10} [F]$); this is especially useful if the noise figures are close to unity.

NYQUIST'S THEOREM

When a resistor is at a finite temperature, a fluctuating current will occur as a result of random thermal motion of electrons in the resistor, which will in turn will give rise to a fluctuating electromagnetic force (emf). This is the electrical analogue to dissipative Brownian Motion giving rise to a fluctuating force. A detailed description of these principles are given in Ref. 19. This fluctuating emf was predicted by Einstein (20) and first discovered by Johnson [it is known as Johnson noise (21)]. Later, the physics was explained by Nyquist and is known as Nyquist's theorem (22).

To illustrate Nyquist's theorem, first we consider a resistance R in series with an arbitrary reactance $X(f)$, as shown in Fig. 3. It is known that $V(t)$ is a white noise source, and thus its spectral density $S_v(f)$ will be constant. From the Wiener-Khitchine theorem and linear circuit theory, the relation between the voltage and current spectral densities is given by

$$S_i(f) = \frac{S_v}{R^2 + X(f)^2} \quad (8)$$

Assuming $X(f)$ is inductive (we can also assume that it is capacitive and achieve the same result) and equal to $2\pi fL$, where L is the inductance, it can be shown from Eqs. (8) and (1) that

$$\langle i^2 \rangle = \int_0^\infty S_i(f) df = S_v \int_0^\infty \frac{df}{R^2 + 4\pi^2 f^2 L^2} = \frac{S_v}{4RL} \quad (9)$$

Now the equipartition theorem is invoked. It states the mean energy stored in an inductor (a similar equation governing the energy stored in a capacitor exists) is given by (19)

$$\frac{1}{2} L \langle i^2 \rangle = \frac{1}{2} k_B T \quad (10)$$

By combining Eqs. (10) and (9), Nyquist's theorem is derived:

$$S_v = 4k_B T R \quad (11)$$

This derivation gives only the classical solution and is not general for low temperatures and high frequencies. When $\omega \geq k_B T$, there is a quantum correction and takes the form (18,19,22)

$$S_v = \frac{4\hbar\omega R}{e^{\hbar\omega/k_B T} - 1} \quad (12)$$

where $\hbar$ is Planck's constant. When T is large Eq. (12) collapses back to the form given by Eq. (11).

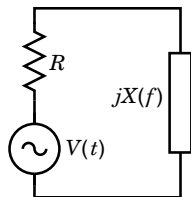


Figure 3. Model of a resistor connected in series with a reactance at temperature T . The voltage generator represents the internal Johnson noise supplied by the passive circuit.

EQUIVALENT CURRENT AND VOLTAGE NOISE GENERATORS

The noise at the output of a two-terminal (or one-port) network can be represented as either a noise current generator in parallel with its admittance or a noise emf in series with its impedance. The equivalent networks are shown in Fig. 4. From Nyquist's theorem, the thermal noise of a resistance R at temperature T , measured in a frequency bandwidth Δf , can be represented by the voltage generator $\sqrt{4k_B T R \Delta f}$ in series with a resistance R , as shown in Fig. 4(b). Likewise, the noise may be equally represented by the current generator $\sqrt{4k_B T g \Delta f}$ in parallel to the resistance $R = 1/g$, where g is the conductance. Thus, the *equivalent noise resistance* or *equivalent noise conductance* of a device at temperature T_0 may be written as

$$R_n = \frac{\langle v^2 \rangle}{\Delta f} \frac{1}{4k_B T_0} = \frac{S_v(f)}{4k_B T_0}, \quad g_n = \frac{\langle i^2 \rangle}{\Delta f} \frac{1}{4k_B T_0} = \frac{S_i(f)}{4k_B T_0} \quad (13)$$

Thus, from a measure of the open circuit voltage $S_v(f)$ [V^2/Hz] or short circuit current spectral density $S_i(f)$ [A^2/Hz], the equivalent noise resistance and conductance may be calculated. Figure 4 implies that the noise spectral densities are related by $S_v(f)/S_i(f) = |Z|^2$; thus the noise conductance and resistance are in general related by

$$R_n = g_n |Z|^2 \quad (14)$$

The noise resistance and conductance can be related to the noise temperature by equating the spectral densities as follows:

$$S_i(f) = 4k_B T_n g, \quad S_v(f) = 4k_B T_n R \quad (15)$$

Thus by equating Eqs. (15) and (13), the relation between the noise temperature, noise resistance, and noise conductance may be written as

$$T_n = \frac{g_n}{g} T_0 = \frac{R_n}{R} T_0 \quad (16)$$

The noise of the device has been expressed in terms of the temperature T_0 , which is not necessarily the device temperature. This is because the device temperature is not necessarily well defined because a temperature gradient might exist, especially for active devices. Thus, it is common to define the noise characteristics with respect to the standard temperature $T_0 = T_s$ (290 K). However, if a device is passive and gives the expected thermal noise, then $T_n = T_0$, $g_n = g$, and $R_n = R$, and the

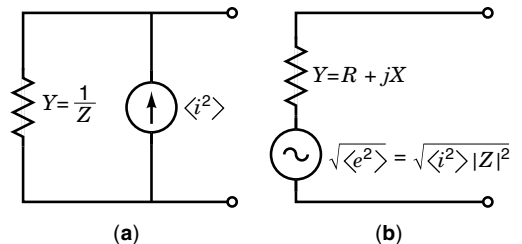


Figure 4. (a) Two-terminal network represented by a noise current generator in parallel with an admittance. (b) Two-terminal network represented by a noise voltage generator in series with an impedance.

device can be defined with respect to its own temperature as a reference.

It is necessary to mention here that these generators are one-port or two-terminal devices that, for example, may represent a diode or resistive noise source. If a device has more ports or terminals associated with its structure, then more than one noise generator must be considered when calculating the noise parameters. One such example is a transistor that is a three-terminal device. For example, a bipolar transistor has an emitter, base, and collector. In general, three noise generators between each terminal must be considered along with any correlated components. This article does not discuss this problem, and the reader is referred to Ref. 18 for more details. For our considerations, the noise added by a two-port system will be described by the noise figure and noise temperature concept and is discussed in more detail later.

TYPES OF NOISE

Thermal Noise

The basics of thermal noise was given previously by the Nyquist description and holds for a passive resistor as long as its temperature is in equilibrium. This description allows for the general quantification of any measurable noise power in terms of equivalent noise temperature, equivalent noise figure, equivalent noise conductance, and equivalent noise resistance. Thermal noise in a resistor is caused by the random motion of the current carriers, which produce a fluctuating voltage across its terminals. The problem can also be treated as a diffusion problem or velocity fluctuation problem of the current carriers (18). These descriptions are particularly useful when analyzing the thermal noise in a semiconductor. A semiconductor device has a nonlinear voltage current characteristic, and Nyquist's theorem holds for a $p-n$ junction at zero bias where the resistance is considered as dV/dI at the temperature of equilibrium.

Generation and Recombination Noise

Generation and recombination noise ($g-r$ noise) occurs in semiconductors that involve donors, traps, and recombination centers of the charge carriers. In general, the resistance of the material is dependent on the number of charge carriers; thus if the number of carriers fluctuates, the resistance will fluctuate also, causing a fluctuating voltage across the terminals. The appearance and disappearance of carriers by this process is described by

$$\frac{d}{dt}(\Delta n) = -\frac{\Delta n}{\tau} + H(t) \quad (17)$$

Here $H(t)$ is a random noise term, Δn is the fluctuation in number of carriers, and τ is the carrier life time. If we take the Fourier transform of Eq. (17) and apply the Wiener-Khinchine theorem, then the frequency domain representation can be written as

$$S_n(f) = \frac{S_H(f)\tau^2}{1 + 4\pi^2\tau^2 f^2} \quad (18)$$

Assuming that the spectral density of $H(t)$, S_H , is white,

the value of $\langle \Delta n^2 \rangle$ may be calculated from Wiener-Khinchine theorem as

$$\langle \Delta n^2 \rangle = \int_0^\infty S_n(f) df = S_H \tau \int_0^\infty \frac{\tau df}{1 + 4\pi^2\tau^2 f^2} = \frac{S_H \tau}{4} \quad (19)$$

Thus, by combining Eqs. (18) and (19), the spectral density of the number of fluctuating carriers can be shown to be

$$S_n(f) = \frac{4\langle \Delta n^2 \rangle \tau}{1 + 4\pi^2\tau^2 f^2} \quad (20)$$

The spectrum $S_n(f)$ can be calculated as soon as τ and $\langle \Delta n^2 \rangle$ are known.

Shot Noise

Shot noise occurs when there is current flow across a potential barrier. The current fluctuates around an average value as a result of a random emission of current carries across the potential barrier. This effect occurs in $p-n$ junctions in diodes and transistors, at the cathode surface of a vacuum tube, and so on. Shot noise can be driven by thermal fluctuations or other mechanisms such as noise due to recombination centers in the space charge region. To describe shot noise, we will consider a n^+-p junction as shown in Fig. 5.

The characteristic of the current crossing the depletion (or space charge) region can be written as

$$I(V) = I_0(V) \left(\exp \left[\frac{qV}{kT} \right] - 1 \right) \quad (21)$$

where q is the charge of the electron. Here, the first term in Eq. (21) is caused by electrons injected from the n^+ -region to the p -region, and the second term is the reverse current caused by electrons injected from the p -region into the n^+ -region. To proceed further, Schottky's theorem is invoked (23). It states that the spectral density of current fluctuations in an emission process is related to the current by

$$S_i = 2qI \quad (22)$$

for frequencies less than the inverse of the transit time. Because noise spectral densities are additive and do not depend on the direction of current flow, the spectral density of

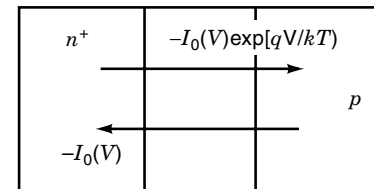


Figure 5. Current flow across the space charge region (or depletion region) in a n^+-p junction when a voltage is applied from the p to n^+ terminal. The separation of charge in the space charge region causes an internal electric field that has the potential to create shot noise. The arrows show the direction of the electron flow, which is in the opposite direction of the current flow.

shot noise current in a $p - n$ junction is given by

$$S_i(f) = 2qI_0(V) \left[\exp\left(\frac{qV}{kT}\right) + 1 \right] = 2q[I(V) + 2I_0(V)] \quad (23)$$

Applying some algebra, it is also common to write Eq. (23) as

$$S_i(f) = 2qI \frac{\left[\exp\left(\frac{qV}{kT}\right) + 1 \right]}{\left[\exp\left(\frac{qV}{kT}\right) - 1 \right]} = 2qI \coth\left(\frac{qV}{kT}\right) \quad (24)$$

or

$$S_i(f) = 2kTg_0 \left(\frac{I + 2I_0}{I + I_0} \right), \text{ where, } g_0 = \frac{dI}{dV} = \frac{q(I + I_0)}{kT} \quad (25)$$

At zero bias, the conductance g_0 supplies the equivalent of full thermal noise as Eq. (25) reduces to $S_i(f) = 4kTg_0$, and half this value when $I \gg I_0$. A similar equation can be written for the open circuit voltage noise spectrum. At high frequencies, the transit time across the depletion region must be considered. This is cumbersome and will not be discussed here; the reader is referred to some standard texts (18,24) to understand this effect. Also, at low temperatures and high frequencies quantum shot noise must be considered just as quantum Nyquist noise was considered previously. Consideration of this effect can be found in Ref. 25.

It should be pointed out here that not all currents exhibit shot noise. If the arrival time of charge carriers are correlated and not random, then suppression occurs. Common examples are space charge limited emissions from a tube and a chemical battery (26). Also, if we design a feedback control system to detect and cancel the noise below, shot noise may be attained. This type of system will then be limited by the noise in the detection system.

Flicker Noise

At large Fourier frequencies most devices can be explained in terms of thermal or shot noise processes. However at small Fourier frequencies excess noise exists. Usually, low-frequency noise has an $f^{-\alpha}$ dependence where $\alpha \sim 1$, and is known as *flicker noise*. The frequency at which the flicker noise is equal to the normal white noise is known as the flicker corner f_c and can vary considerably from device to device. Also, it is not always constant because it can depend on the operating conditions such as input power and bias voltage. Also, it does not in general decrease with temperature as thermal and shot noise do. Similar components made from the same manufacturer can have very different levels of flicker noise, which suggests that it is associated with the fine details and is not under the manufacturer's control.

Flicker noise can be enhanced in nonlinear and chaotic systems, thus it is always prudent to try to maintain the linear operation of a device to keep the flicker noise to a minimum. Recently, a interferometric method, which significantly reduces the effects of flicker noise, was developed at microwave frequencies and is discussed later.

Flicker noise in general cannot be precisely calculated; instead it is common to introduce the noise model in terms of the

white power spectrum S_w

$$S_i(f) = S_w \left(1 + \frac{f_c}{f} \right) \quad (26)$$

In general, no single model can describe the physical principles of flicker noise, unlike shot and thermal noise. However, specific descriptions of flicker noise have been made based on surface and generation and recombination effects (27), quantum effects (28–30), and number and mobility fluctuations (31–33). No attempt will be made to explain these specifics, and the reader is referred to the cited references.

A model worth mentioning is a result of a continuous distribution of time constants (34). A process with a single relaxation time τ_j will have a power spectral density of

$$S_j(f) = \frac{4A_j\tau_j}{1 + (2\pi f\tau_j)^2} \quad (27)$$

At low frequencies, Eq. (27) varies as $1/f^2$ so flicker noise cannot be explained in terms of a single relaxation time process. However, assuming a distribution of relaxation times in the frequency range $1/\tau_2 \ll 2\pi f \ll 1/\tau_1$, the power spectral density can be shown to be

$$S_j(f) = \left[\frac{A}{2\pi \ln\left(\frac{\tau_2}{\tau_1}\right)} \right] \left[\frac{\tan^{-1}(2\pi f\tau_2) - \tan^{-1}(2\pi f\tau_1)}{f} \right] \quad (28)$$

This spectrum is constant at very low frequencies— $1/f$ for an intermediate range and $1/f^2$ for high frequencies. The model given by Eq. (26) cannot be true for very low frequencies because if the spectrum is integrated between 0 and ∞ , it diverges at both limits. Therefore, there must be a lower frequency, in which Eq. (26) is no longer valid and the spectrum varies slower than $1/f$, and an upper frequency, in which Eq. (26) varies faster than $1/f$; Eq. (28) fits this requirement. However, flicker noise in some systems can still be measured in the submicrohertz regime, and thus many processes must have long relaxation times indeed. However, we know that this should at least be limited by the age of the universe!

Burst Noise

Burst noise (also known as popcorn noise) occurs in semiconductor devices and is the result of a defect in manufacturing. The waveform typically consists of random pulses of variable length and constant height (35) and can be described by a random switch model (36). The mechanism of the burst is the result of edge dislocations in the emitter space charge region (37) resulting from lattice distortions caused by large emitter doping densities and metallic impurity. The way to reduce this effect is to keep doping densities below the critical density and to improve manufacturing process. The spectral density of the noise is typically $1/f^\gamma$, where γ is typically 2.

NOISE FIGURE AND TEMPERATURE OF A DEVICE UNDER TEST

To measure the noise in a system, a standard noise source is introduced at the input, as shown in Fig. 6. Referring the device

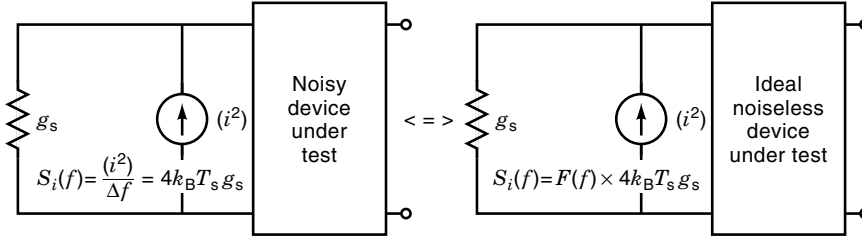


Figure 6. A standard noise source ($4k_B T_s g_s$) at the input of a noisy device under test may be represented as an enhanced noise generator at the input of an ideal noiseless device under test. The enhancement factor $F(f)$ is in general frequency dependent and known as the “spot noise figure.”

noise to the input of the device under test means that the equivalent enhanced current generator gives a noise output power of $F(f)$ times more than the reference temperature generator. This quantity is the spot noise figure defined in Eq. (6).

The spectral density of the equivalent enhanced current generator may be written as

$$S_i(f) = F(f)4k_B T_s g_s = 4k_B T_s g_s + [F(f) - 1]4k_B T_s g_s \quad (29)$$

The first term of the left-hand side in Eq. (29) is the thermal noise of the reference source and the second term is the noise of the device under test. The device is noiseless if the noise figure is unity; in this case, the only noise present in the system is the noise of the source. If the device under test adds the same level of noise as the reference, the noise figure is two.

To relate the noise figure to the noise temperature, T_{nDUT} , of the stage represented by the device under test, the second term on the left-hand side in Eq. (29) may be written as

$$[F(f) - 1]4k_B T_s g_s = 4k_B T_{nDUT} g_s \quad (30)$$

This means that the noise figure and temperature are related by

$$T_{nDUT} = [F(f) - 1]T_s \quad (31)$$

The advantage of the noise temperature notation is that if noise sources are uncorrelated, then noise temperatures are additive. The equivalent noise temperature of the system of Fig. 6 is thus given by $T_{eq} = T_s + T_{nDUT}$, and the spectral density of the equivalent current generator from Eq. (29) may also be represented by

$$S_i(f) = 4k_B T_{eq} g_s \quad (32)$$

When dealing with microwaves, it is more common to consider the noise as a power spectral density rather than a voltage or current spectral density. Microwave circuits are usually matched with the standard impedance of 50Ω . The available power at the output of the noise source is defined as the power fed into a matched load, and the equivalent circuit of a source connected to a load is shown in Fig. 7. If $g_1 = g_s$, then the current will be split into both resistances equally and the available noise power at the load will be

$$N_{av} = \frac{\langle i^2 \rangle}{4g_s} \quad (33)$$

Given that $S_i(f) = \langle i^2 \rangle / \Delta f = 4k_B T_s g_s$, then the power spectral

density of the available noise power will be

$$S_N = N_{av} / \Delta f = k_B T_s \quad (34)$$

This formula is similar to the initial definition of noise temperature in Eq. (5) (i.e., if we measure the noise power spectral density referred to the input of a device, then to calculate the noise temperature, we simply divide the spectral density by Boltzmann's constant).

In reality, devices are never perfectly matched; the effect of a mismatch in the context of a scattering matrix and reflection coefficient description will be discussed later. Also, the noise temperature and generator concept can be generalized to multiport devices, and in general correlations between noise sources must be considered. This will not be presented in this article, and the reader is referred to the literature (38).

NOISE FIGURE AND TEMPERATURE OF CASCADED STAGES

If two networks of noise temperature T_{N1} and T_{N2} and power gains G_1 and G_2 are cascaded and fed by a noise source of temperature T_s at the input of the first, then the output noise density is given by

$$S_N(f)_{out} = G_1(f)G_2(f)k_B T_s + G_1(f)G_2(f)k_B T_{N1} + G_2(f)k_B T_{N2} \quad (35)$$

To refer the noise to the input of the first stage, the noise power must be divided by the total gain $G_1(f)G_2(f)$, and the equivalent noise power density at the input is

$$S_N(f)_{in} = k_B \left(T_s + T_{N1} + \frac{T_{N2}}{G_1(f)} \right) \quad (36)$$

Thus, the equivalent noise temperature of the system is

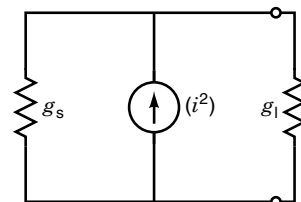


Figure 7. Equivalent current generator of a resistor of conductance g_s with a resistive load g_l at the output.

given by

$$T_{\text{eq}} = T_s + T_{N1} + \frac{T_{N2}}{G_1(f)} \quad (37)$$

For the single stage introduced in the last section, it was noted that the noise temperature is additive. In general for cascaded networks, the noise temperature must be normalized by the preceding gain before it is summed. In general for a cascade of m networks,

$$T_{\text{eq}} = T_s + T_{N1} + \sum_{j=2}^m \left(\frac{T_{Nj}}{\prod_{i=1}^{j-1} G_i(f)} \right) \quad (38)$$

The noise figure for the cascaded network when $m = 2$ is given by

$$F_{\text{eq}}(f) = \frac{S_N(f)_{\text{in}}}{k_B T_s} = 1 + \frac{T_{N1}}{T_s} + \frac{T_{N2}}{G_1(f) T_s} \quad (39)$$

where $T_s = 290$ K. Combining the relationship between noise figure and noise temperature for a single network derived previously in Eq. (31), the following is obtained:

$$F_{\text{eq}}(f) = F_1(f) + \frac{F_2 - 1}{G_1(f)} \quad (40)$$

This may be generalized to a system consisting of m cascaded networks and is given by

$$F_{\text{eq}}(f) = F_1(f) + \sum_{j=2}^m \left(\frac{F_j(f) - 1}{\prod_{i=1}^{j-1} G_i(f)} \right) \quad (41)$$

This equation is known as Friis's formula (39).

NOISE FIGURE AND TEMPERATURE OF LOSSY TRANSMISSION LINE

When a system has a significant length of transmission line, which may be at different temperatures, then the losses may significantly contribute to the equivalent noise temperature of the system. Examples of this type can occur when undertaking cryogenic measurements of a device under test. If the noise source is at room temperature, then the transmission line must connect to the input via a long cable with a temperature gradient. Another example is a link connecting a satellite ground station antenna and a low noise amplifier. In this section, the physical temperature will be denoted by T and the noise temperature by T_N .

First, a piece of transmission line at temperature T and loss L , where $L = 1/G$, is considered. There are two ways in which the transmission line can degrade the noise performance of a system: (1) The loss attenuates the signal and thus effectively enhances the noise temperature of the following stages [see Eq. (38)]; (2) The lossy system is itself a noise generator in a similar way to a resistor (i.e., it generates Nyquist noise). This noise is dependent only on the dissipated power.

The noise generated by the transmission line is dependent on the power lost in transmission (i.e., the power dissipated).

The fraction of power that is transmitted is equal to $1/L$ (or G) and thus by conservation of energy, the fraction of dissipated power must be equal to $(1 - 1/L)$ or $(1 - G)$. Thus the available noise power density at the output of a lossy transmission line with a standard noise source of T_s at its input will be

$$S_N(f)_{\text{out}} = G_1 k_B T_s + (1 - G_1) k_B T_1 \quad (42)$$

The second term in Eq. (42) is the contribution of the first piece of transmission line after the source input (denoted by subscript 1). To calculate the effective noise temperature refer to the input of the transmission line, T_{N1} ; this term must be equated with $G_1 k_B T_{N1}$ and can be calculated to be

$$T_{N1} = \frac{(1 - G_1)}{G_1} T_1 = L_1 \left(1 - \frac{1}{L_1} \right) T_1 \quad (43)$$

From Eqs. (43) and (31), the noise figure of the transmission line may be calculated to be

$$F_1(f) = 1 + (L_1 - 1) \frac{T_1}{T_s} \quad (44)$$

If a transmission line is without loss, it will not add any extra noise to the system. However, if it is not, then the second term in Eq. (44) represents the extra noise resulting from the power dissipation. Also, if there are any other networks cascaded after the lossy line, the noise added when referred to the input of the transmission line will be degraded by its loss L (or $1/G$).

If a lossy transmission line exists in a system under measurement, it may be considered as a separate network among a cascaded system and treated as in the previous section. This gives a method for correcting for any significant transmission line losses if they are known.

TWO-NOISE SOURCE METHOD

The two-noise source technique makes use of connecting two different known noise sources to the input of a device under test and measuring the change in output (40). It is assumed that two known noise sources of temperatures T_h and T_c are connected in turn to the input of the device under test. In this case, the ratio of output power from the two sources will be equal to

$$Y = \frac{T_e + T_h}{T_e + T_c} \quad (45)$$

where T_e is the effective noise temperature of the device under test. (This is the same as T_N introduced previously, however, to be consistent with the literature for this technique we will use T_e .) Thus

$$T_e = \frac{T_h - Y T_c}{Y - 1} \quad (46)$$

This is referred to as the operational definition of T_e (40).

This method assumes perfect matching between components. In reality, the measurement technique is made more

complicated if mismatches occur between the source and device under test. To characterize the mismatch requires the use of more sources and a more complicated measurement procedure and is discussed in the next section.

EFFECTS OF MISMATCH

When considering the effect of mismatch on noise measurements of a device under test, it has been useful to use a scattering matrix or reflection coefficient method to describe the measurement (41,42). Mismatch effects are pronounced when measuring amplifiers because the optimum input impedance is not the same as the matching condition for maximum power flow. For this reason, it is common for a low-noise commercial amplifier to come with an isolator at the input. Complications in the measurement procedure occur because not only does the mismatch change the level of available power, but it also means that reflections will occur at two planes, as shown in Fig. 8. Consequently, reverse flow of power through the device under test must be considered. In the previous analysis, it was always assumed that the flow of power was one way (i.e., from the source to the output of the device under test). In general, the value of noise temperature is dependent on the mismatch, and to characterize the system completely, four independent parameters must be measured (43–46). There are two common approaches with different parameter sets that warrant further discussion.

Noise Parameter Set $T_{e(\min)}$, R_n , and Γ_{opt} (Complex)

The dependence on the reflection coefficient at the input port between the source and the device under test must be corrected for, and it has been traditional to use the four standard IEEE parameters given to characterize this correction. To solve for the four parameters, at least four different noise sources are needed, but it is advantageous to have more to allow averaging and to help ensure a unique solution. The classical way of measuring these parameters relies on using tuners on each end of the device under test to simulate the input and output matching networks when calibrating the system (47–50). This method relies on scalar information only and is sometimes unreliable. To determine the noise parameters more accurately, a vector network analyzer is needed to calculate the device under test scattering parameters and the reflection coefficient of the receiver (50).

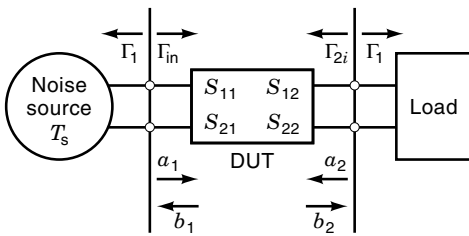


Figure 8. Schematic of the noise temperature measurement of a device under test such as an amplifier. In general, the model uses either the reflection coefficient Γ concept or the scattering parameter S_{ij} concept. Many modern undergraduate textbooks discuss these concepts so they are not explained here.

T_e may be written in terms of $T_{e(\min)}$, R_n , and Γ_{opt} (contains two parameters because it is complex) as

$$T_e = T_{e(\min)} + \frac{4T_s R_n G_{\text{opt}} |\Gamma_1 - \Gamma_{\text{opt}}|^2}{(1 - |\Gamma_{\text{opt}}|^2)(1 - |\Gamma_1|^2)} \quad (47)$$

where

$$G_{\text{opt}} = \frac{(1 - |\Gamma_{\text{opt}}|^2)}{Z_0 |1 + \Gamma_{\text{opt}}|^2}$$

Here, T_s is a reference temperature (typically 290 K), Z_0 is the characteristic line impedance (normally 50 Ω for a microwave coaxial system). $T_{e(\min)}$ is the minimum effective input noise temperature, which is realized when $\Gamma_1 = \Gamma_{\text{opt}}$, and R_n characterizes how quickly T_e increases above $T_{e(\min)}$.

Noise Parameter Set T_a , T_{rev} , and β (Complex)

This set of parameters, developed at NIST (43), is useful because they are terminal invariant (i.e., their values do not change if a lossless two port is added or subtracted from the input). The method was developed to enable a direct measurement of one of the parameters, namely T_{rev} , the noise temperature of the radiation emerging from the input. Recently this method was shown to give an accuracy of ± 0.04 dB when measuring commercial low-noise microwave amplifiers (51). However, a disadvantage is that it requires skill in low-temperature measurements because T_{rev} is typically at cryogenic temperatures for a low-noise amplifier.

T_e can be expressed in relation to this model as

$$T_e = \frac{T_a + T_{\text{rev}} |\Gamma'_i - \beta|^2}{1 - |\Gamma'_i|^2} \quad (48)$$

where

$$\Gamma'_i = \frac{\Gamma_1 - S_{11}^*}{1 - S_{11} \Gamma_1}$$

Here, S_{11} is the input scattering parameter to the device under test (see Fig. 8), T_{rev} is the available noise power from the internal noise sources when the output of the amplifier is terminated in a noiseless matched load, β is a measure of the correlation of the available noise power emanating from the two amplifier ports, and T_a is the amplifier noise temperature if no mismatch exists at the amplifiers input.

As before, this method still requires accurate determination of scattering parameters and reflection coefficients using a vector network analyzer, as well as at least four different sources to determine the four parameters. However, it is also common to use more than four noise sources to add some redundancy to improve the accuracy of the experiments (46,51,52).

For a properly made measurement when mismatch is corrected, the major source of inaccuracy is the accuracy to which the noise source is calibrated. Other noise parameters are obtained from relative measurements and are not affected by noise source calibration errors. This fact has led some of the world's national metrological institutes to do a comparison of noise source calibration and is discussed later.

NOISE SOURCES

Calibrated noise sources are essential if accurate noise measurements as discussed previously are to be made. Noise sources may be categorized as either primary or nonprimary.

Primary Noise Standards

Primary noise standards are thermal noise sources that include a resistive device held at a known temperature. Examples include cooled coaxial resistive terminations immersed in liquid nitrogen at its boiling temperature (cold standard), high-temperature oven standards that operate at typically 500 K to 800 K (hot standards), and room temperature standards that operate at about 290 K. More information on these types of standards can be found in Refs. 53–59. The measurement and characterization relies on knowing the calculable output noise power of a black body radiator such as SiC at cryogenic temperatures or the known temperature and resistance of a passive termination.

Other noise sources can be classified as nonprimary and are usually calibrated against a primary standard if accurate measurements are to be made. They include gas discharge tubes, solid-state diode sources, and commercial “cold load” thermal noise sources operating at liquid nitrogen temperature as well as “hot load” noise sources operating at elevated temperatures. Because we have already discussed in detail thermal noise, only the former two will be discussed further.

Diode Noise Sources

In forward bias, a diode produces shot noise and may be used as a noise source. However, at high frequencies such as microwave, the noise is reduced as a result of transit time effects. Also, flicker noise is present. This means that a white noise spectrum is not generated so the effective noise temperature is a function of Fourier frequency.

Another way a diode may be used to produce noise is to reverse bias the diode near the breakdown (or avalanche) region. Such noise sources give white noise up to the gigahertz region and can be used as wide-band noise sources. They are relatively small devices that require only small voltages and operate with noise temperatures of order 10^4 K. The diode itself has a $T_N = 10^5$ to 10^6 K, which is reduced by a sizeable attenuator used to provide matching. These types of noise sources need calibration, remain calibrated for long times, and are available commercially.

Gas-Discharge Noise Sources

Gas discharges become electrical conductors at high temperatures. Typically a gas discharge tube consists of an ionized noble gas at low pressure. High voltages across the tube are necessary to create the discharge. Typically, these devices produce a discharge with excellent long-term stability that is practically independent of operating conditions, which varies little from tube to tube. The available noise power exhibits a flat power spectrum and an effective noise temperature of order 10^4 K. To make use of the noise generation at microwave frequencies, the tube can be mounted in a microwave waveguide.

INTERNATIONAL COMPARISON OF THERMAL NOISE TEMPERATURE MEASUREMENTS

The world's major state-of-the-art metrological institutes are the places where the most accurate measurements of absolute noise temperature can be made. Recently, an international comparison was undertaken to measure the noise temperature of two commercial microwave noise sources to try to obtain a measure of the absolute error (11). The institutes that took part were the National Institute of Standards and Technology in Boulder, Colorado, United States (13); the Laboratoire Central des Industries Electriques in Fontenay-aux Roses, France (14); the National Physical Laboratory in Teddington, Worcestershire, United Kingdom (12); and the Physikalisch-Technische Bundesanstalt in Braunschweig, Germany (15).

The measurements required each institute to provide its own noise standard and radiometer to perform measurements at three different frequencies. Two entirely different primary standards (cryogenic and oven) were implemented, along with two different types of radiometer (switching and total power). Uncertainties (2σ) between the institutes ranged from 0.5% to 2.9%. Expense and effort was required to fix the primary standards and sources under measurement at a fixed operating temperature. This achievement represents the best determination of absolute measurement uncertainties.

ON-WAFER NOISE MEASUREMENTS

The advent of integrated circuits designs on a single chip has enabled the industry to miniaturize circuits and reduce the expense of microwave and millimeter wave technologies. Collectively, a circuit of this type is often referred to as a *wafer*. The characterization of such devices is more challenging than the waveguide or coaxial counterpart because it is harder to isolate a single device in a wafer. Moreover, impedance matching the device under test to the measurement apparatus is difficult, and these characterizations are always implemented in a large mismatch environment. For accurate measurement, it is important to have a vector network analyzer, which is calibrated for *S*-parameter measurements. The measurement process is quite intensive because it requires the measurement of the *S*-parameters of the noise source, noise receiver, adaptors, probes, and the like.

Much work has been done to develop on-wafer techniques (50,60–63). Commercial systems exist; however, work still is being pursued to assess the accuracy and reliability of such measurements, and ways of improving on-wafer measurements are still under investigation. The chief problem arises as a result of the mismatch and loss of the adaptor to coaxial and waveguide noise standards. To reduce these errors, on-wafer calibrated noise standards are needed. Measurements have been achieved with off-wafer noise standards, and some initial steps using on-wafer uncalibrated noise sources have been achieved (64,65). Other problems include radiation entering the open stripline format of a wafer.

INTERFEROMETRIC NOISE MEASUREMENTS

Previous discussions have mainly been about noise temperature and figure measurements of a device under test (DUT)

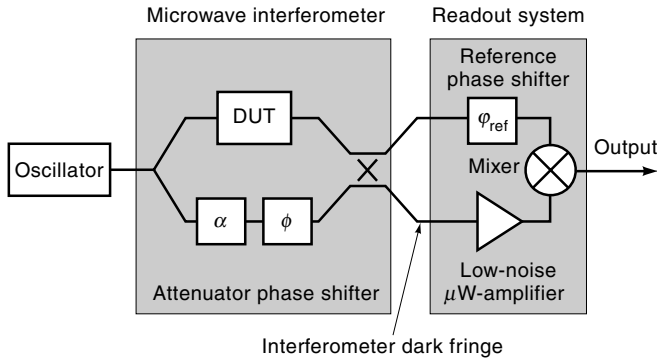


Figure 9. Schematic of the microwave interferometer measurement system of a DUT in the presence of a carrier frequency.

when it is operating in the small signal regime (i.e., with no carrier signal present). A large carrier signal can cause non-linear effects such as up-conversions, which enhance noise at small values of Fourier frequency (offset from the carrier). Recently, interferometric measurement and signal generation schemes at microwave frequency have been developed (8,9) and allow component characterization and operation at the intrinsic small signal noise temperature independent of the carrier power. The basic noise measurement setup is shown in Fig. 9. It includes the microwave interferometer consisting of the DUT and a compensating branch that enables the cancellation of the carrier at the dark fringe of the interferometer. The signal without the carrier is then amplified by the low-noise microwave amplifier operating in the small signal regime. The low-noise microwave amplifier and a mixer form the microwave readout system, which can be either phase- or amplitude-sensitive depending on the setting of the reference phase shifter ϕ_{ref} .

One of the features of the interferometric measurement system is a greatly enhanced sensitivity to phase and amplitude fluctuations of the DUT. The sensitivity enhancement results from the ability of interferometric system to satisfy, on the first glance, two contradictory requirements: (1) having a high power at the input of the interferometer, and (2) enabling low-noise operation of the readout system. These requirements are met by interfering the two signals destructively from the output of the DUT and compensating branch (suppressing the carrier) before the noise is detected by the readout system. This enables the microwave amplifier in the readout system to operate in the small signal regime devoid of flicker noise and reduces the effect of the mixer flicker noise by the amplifier gain.

The effective noise temperature of the readout system T_{RS} limits the sensitivity of the interferometric noise measurement system, and the noise temperature is given by

$$T_{\text{RS}} = T_0 + T_a + \frac{T_m}{G_a} \quad (49)$$

where $T_0 \approx 290$ K is the ambient temperature; T_a and T_m are the effective noise temperature of the microwave amplifier and mixer, respectively; and G_a is the gain of the microwave amplifier. In general, T_{RS} is a function of Fourier frequency f and the power at the input of the low-noise microwave amplifier. This dependence is caused by the amplifier flicker noise, which generally scales proportionally with input power (66,67).

Suppressing the carrier at the interferometer output minimizes the degradation of T_{RS} as a result of the amplifier flicker noise, and choosing an amplifier with a low level of Johnson noise allows the measurement to be close to ambient temperature. There is some conflicting data published (68) that suggests flicker noise in amplifiers can be reduced as input power is increased. However, when the input to the amplifier is totally nulled, the resulting noise spectrum is white and devoid of flicker noise, and this represents the optimum operating condition.

These types of measurements are relatively new; consequently, the accuracy has not been determined, and mismatch correction techniques have not yet been applied.

ACKNOWLEDGMENTS

The author thanks Dr. Eugene Ivanov and Richard Woode for reading the manuscript and helping with references. He also thanks the Librarian Will Hamilton for his patient help in finding reference material.

BIBLIOGRAPHY

1. L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise*, Englewood Cliffs, NJ: Prentice-Hall, 1962.
2. C. H. Cook and M. Bernfeld, *Radar Signals, An Introduction to Theory and Application*, Boston: Artech House, 1993.
3. A. D. Whalen, *Detection of Signals in Noise*, New York: Academic Press, 1971.
4. D. O. North, An analysis of the factors which determine signal/noise discrimination in pulsed carrier systems. *Proc. IEEE*, **51**: 1016, 1963.
5. J. C. Mather et al., A preliminary measurement of the cosmic microwave background spectrum by the cosmic background explorer (COBE) satellite, *Astrophys. J. Lett.*, **37**: 354–356, 1990.
6. D. G. Blair et al., High sensitivity gravitational wave antenna with parametric transducer readout, *Phys. Rev. Lett.*, **74**: 1908–1911, 1995.
7. P. R. Saulson, *Fundamentals of Interferometric Gravitational Wave Detectors*, Singapore, World Scientific, 1994.
8. E. N. Ivanov, M. E. Tobar, and R. A. Woode, Microwave interferometry: Application to precision measurements and noise reduction techniques, *IEEE Trans. Microw. Theory Tech.*, 1998.
9. E. N. Ivanov, M. E. Tobar, and R. A. Woode, A study of noise phenomena in microwave components using an advanced noise measurement system, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **44**: 161–163, 1997.
10. R. H. Dicke, The measurement of thermal radiation at microwave frequencies, *Rev. Sci. Instrum.*, **17**: 268–275, 1946.
11. J. Randa et al., International Comparison of Thermal Noise-Temperature Measurements at 2, 4, and 12 GHz. Communicated at CPEM, Gaithersburg, MD, 1998.
12. National Physics Laboratory, Teddington, UK. <http://www.npl.co.uk/npl/cetm/rfm/rfcontacts.html>
13. National Institute for Science and Technology, Boulder, Colorado, USA. <http://www.boulder.nist.gov/div813/81306/noise/noise.htm>
14. Laboratoire Central des Industries Electriques, Fontenay-aux-Roses, France. <http://www.ccip.fr/club/92/lcie-gb'ps.html>
15. Physikalisch-Technische Bundesanstalt, Braunschweig, Germany. <http://www.ptb.de/english/org/2/hp.htm>
16. N. Wiener, Generalized harmonic analysis, *Acta Math.*, **55**: 117, 1930.

17. A. Khintchine, Korrelations theorie der stationaren stoichastischen Prozesse, *Math. Ann.*, **109**: 604, 1934.
18. A. v. d. Ziel, *Noise in Solid State Devices and Circuits*, New York: Wiley, 1986.
19. F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw Hill, 1981.
20. A. Einstein, Theory of Brown Motion, *Ann. Phys.*, **19**: 289, 371, 1906.
21. J. B. Johnson, Thermal agitation of electricity in conductors, *Phys. Rev.*, **32**: 97, 1928.
22. H. Nyquist, Thermal agitation of electric charge in conductors, *Phys. Rev.*, **32**: 110, 1928.
23. W. Schottky, Electric oscillations, *Ann. Phys.*, **57**: 541, 1918.
24. F. N. H. Robinson, *Noise and Fluctuations in Electronic Devices and Circuits*, Oxford: Clarendon Press, 1974.
25. J. Tucker, Quantum limited detection in tunnel junction mixers, *IEEE J. Quant. Electron.*, **15**: 1234, 1979.
26. C. K. Boggs et al., Measurement of voltage noise in chemical batteries, *Proc. IEEE Int. Freq. Control Symp.*, 367–374, 1995.
27. K. Amberiadis, G-R and $1/f$ noise in semiconductor devices, Ph.D. Thesis, University of Minnesota, 1982.
28. P. H. Handel, $1/f$ —an “inferred” phenomenon, *Phys. Rev. Lett.*, **34**: 1492, 1975.
29. P. H. Handel, Quantum approach to $1/f$ noise, *Phys. Rev.*, **22**: 745, 1980.
30. K. M. v. Vliet, P. H. Handel, and A. v. d. Ziel, Superstatistical emission noise, *Physica A*, **108**: 511, 1981.
31. F. N. Hooge, $1/f$ noise is no surface effect, *Phys. Lett. A*, **29**: 139, 1969.
32. Hooge, $1/f$ noise, *Physica B*, **83**: 19, 1976.
33. R. P. Jindal and A. v. d. Ziel, Phonon fluctuation model for flicker noise in elemental semiconductors, *J. Appl. Phys.*, **52**: 2884, 1981.
34. A. L. McWhorter, $1/f$ Noise and Related Surface Effects in Germanium, Lincoln Lab., Boston 80, May 1955.
35. G. Giralt, J. C. Martin, and F. X. Matea-Perez, Sur un phénomène de bruit dans les transistors, caractérisé par des créniaux de courant d'amplitude constante, *Compt. Rend. Acad. Sci. Paris*, **261**: 5350, 1965.
36. S. Machlup, Noise in semiconductors: spectrum of two-parameter random signal, *J. Appl. Phys.*, **25**: 341, 1954.
37. M. Mihaila and K. Amberiadis, Noise phenomena associated with dislocations in bipolar transistors, *Solid State Electron.*, **26**: 109, 1983.
38. D. F. Wait, Thermal noise from a passive linear multiport, *IEEE Trans. Microw. Theory Tech.*, **16**: 687–691, 1968.
39. H. T. Friess, Noise figures of ratio receivers, *Proc. IRE*, **32**: 419, 1944.
40. D. F. Wait, Comments Concerning on Wafer Noise Parameter Measurements, presented at 36th ARFTG Conf. Digest, Monterrey, CA, 1990.
41. T. Nemoto and D. F. Wait, Microwave circuit analysis using the equivalent generator concept, *IEEE Trans. Microw. Theory Tech.*, **16**: 866–873, 1968.
42. G. F. Engen, A new method of characterizing amplifier noise performance, *IEEE Trans. Instrum. Meas.*, **19**: 344–349, 1970.
43. D. F. Wait and G. F. Engen, Application of radiometry to the accurate measurement of amplifier noise, *IEEE Trans. Instrum. Meas.*, **40**: 433–437, 1991.
44. H. Rothe and W. Dahlke, Theory of noisy fourpoles, *Proc. IRE*, **44**: 811–818, 1956.
45. H. A. Haus and R. Q. Lane, Representation of noise in linear two-ports, *Proc. IRE*, 69–74, 1959.
46. R. Q. Lane, The determination of device noise parameters, *Proc. IEEE*, **57**: 1461–1462, 1969.
47. M. Pospieszalski et al., Comments on ‘Design of microwave GaAs MESFET’s for broadband, low noise amplifier,’ *IEEE Trans. Microw. Theory Tech.*, **34**: 194, 1986.
48. A. Cappy, Noise modelling and measurement techniques, *IEEE Trans. Microw. Theory Tech.*, **36**: 1–10, 1988.
49. E. Strid, Measurements of losses in noise matching networks, *IEEE Trans. Microw. Theory Tech.*, **29**: 247–252, 1981.
50. A. C. Davidson, B. W. Leake, and E. Strid, Accuracy improvements in microwave noise parameter measurements, *IEEE Trans. Microw. Theory Tech.*, **37**: 1973–1978, 1989.
51. D. F. Wait and J. Randa, Amplifier noise measurements at NIST, *IEEE Trans. Instrum. Meas.*, **46**: 482–485, 1997.
52. V. Adamian and A. Uhler, A novel procedure for receiver noise characterization, *IEEE Trans. Instrum. Meas.*, **22**: 181–182, 1973.
53. W. C. Daywitt, Design and error analysis for the WR10 thermal noise standard, *Natl. Bur. Stand. (US) Technol. note 1071*, 1983.
54. W. C. Daywitt, The noise temperature of an arbitrarily shaped microwave cavity with application to a set of millimeter wave primary standards, *Metrologia*, **30**: 471–478, 1994.
55. J. Randa, Noise-temperature measurement system for the WR-28 band, *Natl. Inst. Stand. Technol. Tech. note 1395*, 1997.
56. J. S. Wells, W. C. Daywitt, and C. K. S. Miller, Measurement of effective temperatures of microwave noise sources, *IEEE Trans. Instrum. Meas.*, **13**: 17–28, 1964.
57. F.-I. Buchholz and W. Kessel, A primary broad-banded coaxial thermal noise standard for the range 100 MHz to 10 GHz, *IEEE Trans. Instrum. Meas.*, **36**: 474–479, 1987.
58. M. W. Sinclair and A. M. Wallace, A new national electrical noise standard in the X-band, *IEEE Proc. A*, **133**: 272–274, 1986.
59. W. C. Daywitt, A reference noise standard for millimeter waves, *IEEE Trans. Microw. Theory Tech.*, **21**: 845–847, 1973.
60. M. S. Gupta et al., Microwave noise characterization of GaAs MESFET’s: Evaluation by on-wafer low-frequency output of noise current measurement, *IEEE Trans. Microw. Theory Tech.*, **35**: 1208–1218, 1987.
61. L. Dunleavy, A Ka-band on-wafer s-parameter and noise figure measurement system, *Proc. 34th ATFTG Conf. Digest*, Ft. Lauderdale, FL, 1989.
62. G. Dambrine et al., A new method for on-wafer noise measurement, *IEEE Trans. Microw. Theory Tech.*, **41**: 375–381, 1993.
63. A. Boudiaf, C. Dubon-Chevallier, and D. Pasquet, Verification of on-wafer noise parameter accuracy, *IEEE Trans. Instrum. Meas.*, **44**: 332–335, 1995.
64. J. Randa, R. L. Billinger, and J. L. Rice, On-wafer measurements of noise temperature, submitted to *IEEE Trans. Instrum. Meas.*, 1998.
65. J. Randa, Noise temperature measurements on-wafer, *Natl. Inst. Stand. Technol. Tech. note 1390*, 1997.
66. F. G. Ascarrunz, E. S. Ferre, and F. L. Walls, Investigations of AM and PM noise in X-band devices, *Proc. IEEE Frequency Control Symp.*, 1993.
67. T. E. Parker, Characteristics and sources of phase noise in stable oscillators, *Proc. 41st IEEE Frequency Control Symp.*, 1987.
68. M. C. Delgado Aramburo et al., Comparison of $1/f$ noise in commercial amplifiers, *Proc. IEEE Freq. Control Symp.*, 470–477, 1997.

MICHAEL E. TOBAR
The University of Western Australia

ELECTRIC DISTORTION MEASUREMENT

In the context of the electrical domain, the term *distortion* may be broadly defined as any deviation of a signal in any parameter (time, amplitude, or wave shape) from that of an ideal signal. The term is usually associated with lack of fidelity. When referring to analog signals, distortion means shape alteration. For binary data transmission the term has a particular meaning commonly defined as a displacement in time of a signal from the time that the receiver expects to be correct. When applied to a system, distortion is a manifestation of its nonideal behavior in which the manipulation of at least some type of signal is involved.

DISTORTIONLESS SYSTEM

A periodic signal $s(t)$ may be represented by the following Fourier series:

$$s(t) = S_0 + \sum_{n=1}^{\infty} S_n \cos(n\omega_1 t + \varphi_n) \quad (1)$$

where S_0 represents the dc component of the signal, S_n is the amplitude and φ_n the phase of the n th harmonic, and ω_1 is the angular frequency of its fundamental component. If the signal is not periodic and it satisfies some conditions, the representation does not involve a series but an integral, the Fourier integral given by

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \overline{S}(\omega) e^{j\omega t} d\omega \quad (2)$$

where

$$\overline{S}(\omega) = \int_{-\infty}^{+\infty} s(t) e^{-j\omega t} dt \quad (3)$$

is the Fourier transform of $s(t)$.

According to the Fourier representation, each signal $s(t)$ may be viewed as the sum of an infinite number of sinusoidal signals having a well-defined amplitude, frequency, and phase. The representation is unique. No two signals have the same Fourier representation and vice versa.

An electric analog signal $s(t)$ is said to suffer from distortion when, for some reason, it is transformed into a signal $s'(t)$ that does not satisfy the condition

$$s'(t) = ks(t - t_0) \quad (4)$$

where k is a constant that accounts for a change in amplitude and t_0 is a time delay. Any signal that does not satisfy Eq. (4) is not an exact replica of $s(t)$.

There are several causes for signal distortion: conducted or radiated interference (harmonics in the power network voltage and current, crosstalk between conductors, for instance); signal manipulation (modulation, mixing, etc.); and nonideal behavior of the media or of the systems used to transmit and manipulate the signal.

Even though the first two causes are important, the last deserves further consideration because real transmission media

and systems are always nonideal and thus may contribute significantly to signal distortion. To understand the mechanism of distortion produced in a signal by a medium or system, let us first consider that the media or systems are linear and time-invariant so that they are characterized by a transfer function in the frequency domain:

$$\overline{T}(\omega) = T(\omega) e^{j\theta(\omega)} \quad (5)$$

In such cases, both media and systems may be analyzed in terms of systems theory, and thus we will refer to both as a "system." Because the transfer function relates the output of a system to its input, it accurately describes the system's frequency response and thus the effect of such a system on its signals.

Considering that the input of the system is a periodic signal $x(t)$ represented by its Fourier series

$$x(t) = X_0 + \sum_{n=1}^{\infty} X_n \cos(n\omega_1 t + \varphi_n) \quad (6)$$

then the steady state output signal $y(t)$ is given by

$$y(t) = Y_0 + \sum_{n=1}^{\infty} Y_n \cos(n\omega_1 t + \Psi_n) \quad (7)$$

where

$$\begin{cases} Y_n = T(n\omega_1) X_n \\ \Psi_n = \varphi_n + \theta(n\omega_1) \end{cases} \quad n = 0, 1, 2, \dots \quad (8)$$

If $\overline{CTT}(\omega)$ is such that

$$T(\omega) = k \quad (9a)$$

and

$$\theta(\omega) = -t_0\omega \quad (9b)$$

then replacing Eqs. (9a) and (9b) in Eq. (7) yields

$$y(t) = k \left\{ X_0 + \sum_{n=1}^{\infty} X_n \cos[n\omega_1(t - t_0) + \varphi_n] \right\} = kx(t - t_0) \quad (10)$$

The same result is obtained for a nonperiodic signal provided that it may be represented by a Fourier transform $\overline{CTX}(\omega)$ and that the transfer function of the system satisfies Eqs. (9a) and (9b). In fact, extending the concept of the transfer function to the domain of negative frequencies and taking into consideration that because $x(t)$ and $y(t)$ are real, $T(\omega)$ must be an even function of ω , and $\theta(\omega)$ an odd function of ω , one may write:

$$\overline{Y}(\omega) = \overline{T}(\omega) \overline{X}(\omega) = k e^{-j\omega t_0} \overline{X}(\omega) \quad (11)$$

$y(t)$ is obtained by replacing Eq. (11) in Eq. (2):

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} k \overline{X}(\omega) e^{j\omega(t-t_0)} d\omega = kx(t - t_0) \quad (12)$$

The analysis just presented leads to the important conclusion that a distortionless system must be linear and time-invariant

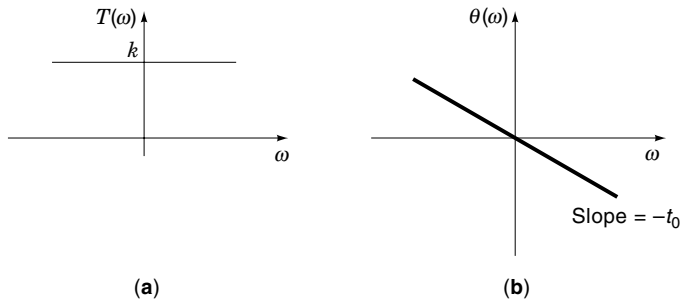


Figure 1. Transfer function of an ideal system. (a) Amplitude response; (b) phase response. A distortionless system is a linear and time-invariant system that has a constant amplitude frequency response k and a phase response that changes in frequency according to a straight line passing through the origin and whose slope is a time delay t_0 .

and have an amplitude response constant for all frequencies and a phase response linear with frequency. The transfer function of an ideal system is represented in Fig. 1.

In a nonlinear system, the output must be related to the input by means of a n -degree polynomial commonly called a transfer characteristic. In those systems distortion always occurs, because a sinusoidal input produces an output with an enriched spectral content.

DISTORTION TAXONOMY AND TERMINOLOGY

As already discussed, electric distortion produces a change in a signal parameter and, in the context of the Fourier representation of a signal, it corresponds to one or several of the following alterations: appearance of energy in new frequencies (nonlinear amplitude distortion); nonproportional change in the amplitudes of the fundamental and harmonic components of the signal (frequency distortion); or change in the phase of the frequencies of the signal (phase or delay distortion). These are the three principal types of distortion. The first is characteristic of nonlinear systems and thus is called nonlinear distortion. The other two are characteristic of dispersive systems and are called linear distortions.

The distortion produced in electrical signals becomes a problem when the information they convey is altered or even lost or when those distorted signals interfere with other signals. Because of the different origins of distortion and the need to characterize and evaluate the performance of systems, the terminology related to distortion includes several expressions that are worth examining because some of them define parameters used in specifying systems distortion performance.

Telecommunications is probably the electrical subdomain where distortion causes the most problems. Two main factors contribute to that: (a) the use of components and devices that are intrinsically nonlinear (mixers, loudspeakers, etc.) or potentially nonlinear (e.g., amplifiers) affect the spectra of the signals they process; (b) the congestion of the frequency spectrum. Thus it is only natural to expect the terminology of distortion to be directly related to sound and video signals.

The following terms are strongly supported by Ref. 1. The list is not exhaustive but aims to include those terms most relevant in electrical distortion measurements.

Amplitude nonlinearity (amplitude distortion) is the phenomenon through which frequency components appear at the

output of a system. They are dependent on the characteristics of the input signal but not present in it. Amplitude nonlinearity produces two types of amplitude distortion: harmonic distortion and intermodulation distortion. Crossover distortion caused by the nonlinear characteristics of a device that changes operating modes, such as a push-pull amplifier, is an example of amplitude distortion.

Harmonic distortion is the amplitude nonlinearity expressed in terms of the ratio of the harmonics in the output signal to the total output signal when a sinusoidal input signal is applied. It is caused by nonlinearity in the system's transfer characteristic. In harmonic distortion measurements, a single sinusoidal signal is applied to the system, and wave analysis at harmonic frequencies determines the percentage distortion.

Harmonic distortion of the n th order is the harmonic distortion expressed in terms of the ratio of the rms output signal due to the component of harmonic order n to the total rms output signal. Usually, the difference between the total rms and the rms of the fundamental is very small, and thus the latter is also used in the ratio. *Total harmonic distortion* is the harmonic distortion expressed in terms of the ratio of the rms output signal due to distortion to the total rms output signal. *Weighted total harmonic distortion*, used in sound system equipment, is the total harmonic distortion measured with a frequency weighting. *Noise harmonic distortion* (or noise distortion) is the harmonic distortion when one-third octave-band filtered noise is used as the input signal.

Intermodulation distortion is the amplitude nonlinearity expressed in terms of the ratio of the input signal of frequencies $pf_1 + qf_2 + \dots$ (where $p, q, \dots$, are positive or negative integers) to the total output signal, when (at least two) sinusoidal input signals having the fundamental frequencies $f_1, f_2, \dots$, are applied at the input. In radio frequency (RF) power amplifiers, for instance, the major causes of intermodulation distortion are crossover effects, gain reduction at high current, and device saturation.

In intermodulation distortion measurements, two sinusoidal signals of different frequencies f_1 and f_2 are applied to the system. *Modulation distortion* is the intermodulation distortion where the input signal is composed of a large-amplitude, low-frequency signal f_1 and a small-amplitude, high-frequency signal f_2 . In some systems two kinds of modulation distortion are present, both having the same spectral components and differing only in phase: (a) amplitude modulation distortion caused by the amplitude modulation due to nonlinearity; (b) frequency modulation caused by frequency modulation having no relationship to nonlinearity. In such cases, it is necessary to distinguish between these two types of distortion. The reference output at which the distortion occurs is taken as the arithmetic sum of the output signals at frequencies f_1 and f_2 .

Modulation distortion of the n th order is the modulation distortion in terms of the ratio of the arithmetic sum of the rms output signal components at frequencies $f_2 \pm (n-1)f_1$ to the rms output signal component at frequency f_2 . *Total modulation distortion* is the modulation distortion in terms of the ratio of the arithmetic sum of the rms output signal components to the rms output signal at frequency f_2 . *Difference-frequency distortion* is the intermodulation distortion where the input signal is composed of two sinusoidal signals f_1 and f_2 of similar or equal amplitude. The difference in the frequency of the two signals is less than the lower of the frequencies. The reference output at which the distortion occurs is taken as the arithmetic

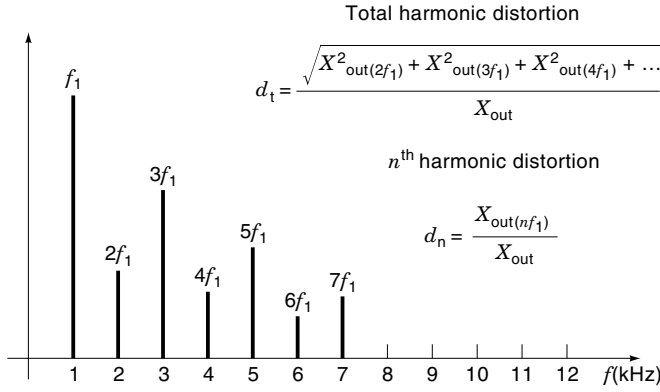


Figure 2. Harmonic distortion. Total harmonic distortion, d_t , and n^{th} harmonic distortion, d_n , are two parameters for characterizing the harmonic distortion of a system. In the expressions, X_{out} represents the rms of the output signal (an electric voltage, in general), $X_{\text{out}}(f_1)$ represents the rms of the fundamental in the output signal, and $X_{\text{out}}(nf_1)$ represents the rms of the harmonic component of nf_1 in the output signal. The input of the system is a sine wave signal of frequency f_1 .

sum of the output signals at frequencies f_1 and f_2 . *Noise intermodulation distortion* is the intermodulation distortion where one-third octave-band filtered noise is used as the input signal. *Transient intermodulation distortion*, important in characterizing loudspeakers, results from nonlinear response to steep wavefronts. It is measured by adding square wave (3.18 kHz) and sine wave (15 kHz) inputs with a 4:1 amplitude ratio and observing the multiple sum and difference-frequency components added to the output spectrum.

In Figs. 2–5, excerpted from Ref. 1, some of the concepts previously discussed are presented in graphical and mathematical form.

Frequency distortion is the effect on a signal that results from variation in the amplitude response of a system as a function of frequency. Some authors also use attenuation distortion

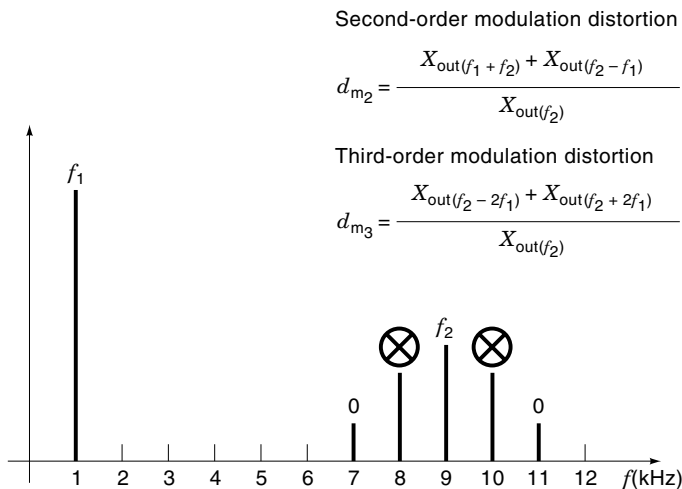


Figure 3. Modulation distortion. Modulation distortion of the n^{th} order d_{m_n} is a parameter for characterizing the intermodulation distortion of a system. The input of the system is a signal with frequencies f_1 and f_2 . Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

$$X_{\text{out}}(f_1) = X_{\text{out}}(f_2)$$

$$X_{\text{out,ref}} = X_{\text{out}}(f_1) + X_{\text{out}}(f_2) = 2X_{\text{out}}(f_2)$$

$$f_2 - f_1 = 80 \text{ Hz, for instance}$$

$$f_m = \frac{f_2 + f_1}{2} \text{ is a preferred one-third octave band center frequency (for instance, 10 kHz)}$$

Second-order difference-frequency distortion

$$d_{d2} = \frac{X_{\text{out}}(f_2 - f_1)}{2X_{\text{out}}(f_2)} = \frac{X_{\text{out}}(f_2 - f_1)}{X_{\text{out,ref}}}$$

Third-order difference-frequency distortion

$$d_{d3} = \frac{X_{\text{out}}(2f_2 - f_1) + X_{\text{out}}(2f_1 - f_2)}{X_{\text{out,ref}}}$$

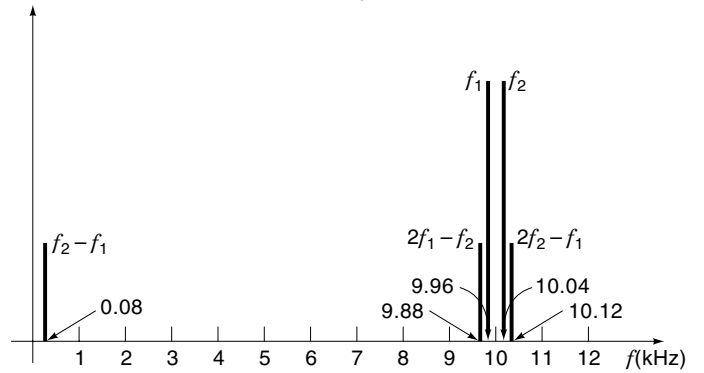


Figure 4. Difference-frequency distortion. Difference-frequency distortion of the n^{th} order d_{d_n} is a parameter for characterizing the intermodulation distortion of a system when a signal having two closely spaced frequency components of similar amplitudes is supplied to the input. In general, this parameter indicates the in-band distortion introduced by the system. Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

or amplitude distortion to designate this effect. If the amplitude response assumes values between k_{max} and k_{min} , the parameter $(k_{\text{max}} - k_{\text{min}})/[(k_{\text{max}} + k_{\text{min}})/2]$ may be used to express frequency distortion.

Phase or delay distortion results from the deviation from a constant slope of the output phase versus frequency response of a system. This produces echo responses in the output that precede and follow the main response and a distortion of the output signal when an input signal having a large number of frequency components is applied. When the phase characteristic of a linear system assumes the value θ at frequency f_1 , the system introduces at that frequency a time delay $t_{d1} = \theta_1/2\pi f_1$ between the input and the output. If the system is not ideal, the time delay $t_{d2} = \theta_2/2\pi f_2$ introduced at frequency f_2 differs from t_{d1} . In that case, the derivative of the phase with respect to frequency is not constant. In our opinion, the maximum value of that derivative over any frequency interval expressed in time units better characterizes phase distortion. Some authors (2) designate this parameter as envelope delay distortion. The experimental evaluation of this parameter may be cumbersome or even impossible, which leads to the implementation of alternative methods. Details are presented in a forthcoming section.

$$X_{\text{out}(f_1)} = X_{\text{out}(f_2)} \quad \text{with, for instance,}$$

$$f_1 = 8 \text{ kHz} \quad f_2 = 11.95 \text{ kHz}$$

$$X_{\text{out}_{\text{ref}}} = X_{\text{out}(f_1)} + X_{\text{out}(f_2)} = 2X_{\text{out}(f_2)}$$

$$f' = f_2 - f_1 = 3.95 \text{ kHz} \quad f'' = 2f_1 - f_2 = 4.05 \text{ kHz}$$

Total difference-frequency distortion:

$$d_{\text{d}_{\text{tot}}} = \frac{\sqrt{(X'_{\text{out}})^2 + (X''_{\text{out}})^2}}{X_{\text{out}}}$$

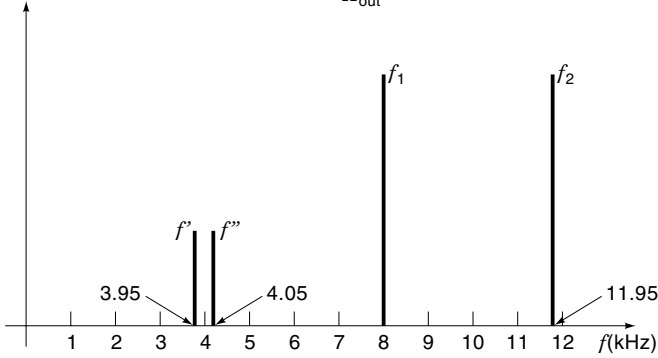


Figure 5. Total difference-frequency distortion. Total difference-frequency distortion $d_{\text{d}_{\text{tot}}}$ is a parameter particularly relevant in assessing the out-of-band distortion introduced by a system. Note that the frequencies of the two-tone signal supplied to the input are not closely spaced as in Fig. 4. Refer to the legend of Fig. 2 for the meaning of the variables in the expressions.

SIGNAL DISTORTION MEASUREMENT

Electric distortion measurements are usually carried out by examining signals (electric voltage, as a rule) in the frequency domain (harmonic analysis). However, both frequency-domain and time-domain instrumentation are used for that purpose. Frequency-domain instrumentation analyzes signals by using analog filtering techniques. Wave analyzers, such as the frequency selective voltmeter, the heterodyne tuned voltmeter, the heterodyne harmonic analyzer (wavemeter), and the heterodyne spectrum analyzer, are examples of this type of instrumentation designed to measure the relative amplitudes of single-frequency components in a complex signal. Time-domain instrumentation analyzes by time sampling the signals and subsequent numerical handling of the sampled data commonly using the fast Fourier transform (FFT) algorithm. The FFT spectrum analyzer is an example of time-domain instrumentation.

Special-purpose instruments, such as the one whose block diagram is presented in Fig. 6 (total harmonic distortion meter), directly display many distortion measurements. The spectrum analyzer is, however, the general-purpose instrument most often used to measure distortion. With it, the entire spectrum within its frequency band is analyzed even though second and third harmonic measurements are enough for many applications. The most common spectrum analyzers are the superheterodyne spectrum analyzer and the FFT spectrum analyzer.

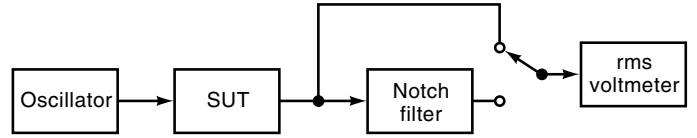


Figure 6. Block diagram of a typical fundamental-suppression total harmonic distortion meter. Total harmonic distortion by the system under test (SUT) is evaluated by internally computing the ratio between the rms values of the output voltage and its value upon suppression of its fundamental frequency. The instrument includes the oscillator, and the rms values are measured by an rms responding voltmeter.

In the FFT spectrum analyzer the input signal is converted, the samples converted from analog to digital, and a FFT is performed. As a result, magnitude and phase spectra of the input signal are obtained. The main advantages of FFT spectrum analyzers compared to superheterodyne spectrum analyzers are their ability to measure phase and the possibility of characterizing single-shot phenomena. Their limitations are related to the frequency range (limited by the ADC maximum conversion rate) and sensitivity (related to quantization noise). FFT spectrum analyzers are very easily implemented by using PC-based automatic measuring systems with plug-in data acquisition boards or by using digitizers and digital oscilloscopes having computer interfaces like RS232 or IEEE488. Now manufacturers are including FFT spectrum analyses as an additional feature of digital oscilloscopes.

Figure 7 shows a simplified block diagram of a superheterodyne spectrum analyzer. After the input attenuator, the signal is applied to a low-pass filter, whose function is analyzed later. The output of the filter is applied to a mixer. Here the signal is mixed with the output of a voltage-controlled oscillator (VCO). The ramp generator sweeps the VCO linearly from f_{min} to f_{max} . Because the mixer is a nonlinear device, its output contains the two original signals and also their harmonics, the sums and differences of the original frequencies, and their harmonics. When any of the frequency components of the mixer output falls within the passband of the filter, a nonzero voltage is applied to the envelope detector and after amplification, to the vertical plates of a cathode-ray tube (CRT), producing a vertical deflection of the electron beam. As the ramp that commands the VCO is also applied to the horizontal plates of the CRT, the horizontal axis can be calibrated in frequency.

The central frequency f_{if} and the bandwidth of the intermediate-frequency filter are chosen so that at any time the only frequency component at the output of the mixer that is within the band of the filter is the difference between the frequency f_{vco} of the VCO and that of the input signal f_{signal} . This implies that f_{if} must be out of the input band. Otherwise, apart from the difference of frequencies we could have a component of the input signal within the passband of the filter. In this case, because the output of the mixer includes the original input signal, this would produce a constant vertical deflection of the CRT during all the frequency scanning of the VCO.

To display frequencies near 0 Hz, the lower frequency f_{min} of the VCO must be equal to f_{if} , because $f_{\text{if}} = f_{\text{vco}} - f_{\text{signal}}$. To display all the frequencies within the band of the spectrum analyzer, the maximum output frequency f_c of the VCO must be $f_{\text{max}} = f_{\text{if}} + f_c$.

Now let us suppose that we have a spectrum analyzer with an input range of frequencies from 0 to 3 GHz. In this case, f_{if}

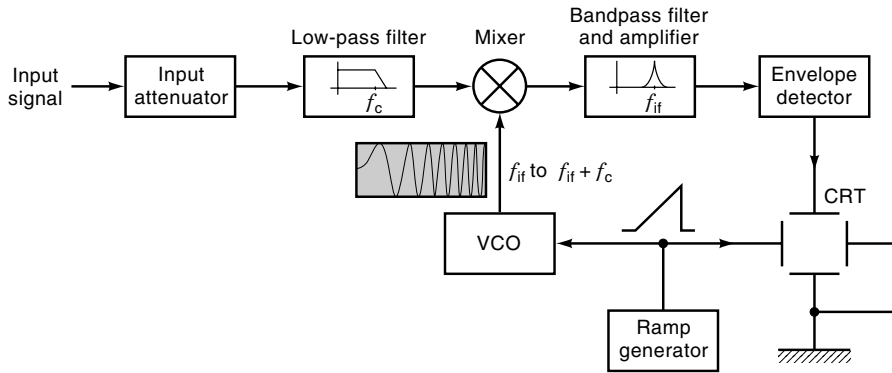


Figure 7. Simplified block diagram of the superheterodyne spectrum analyzer. The superheterodyne spectrum analyzer allows measuring the amplitude or rms values of the frequency components of an electric voltage in a defined frequency band. Those values are obtained by measuring the output voltage of a narrow-passband, fixed-frequency filter when, upon being heterodyned, the spectrum of the input voltage passes the frequency window of the filter.

could be, for instance, 3.5 GHz, and then the output frequency of the VCO should vary from 3.5 GHz to 6.5 GHz. Suppose that we have an input signal with two frequency components, one at 1 GHz (f_{s1}) and the other at 2 GHz (f_{s2}). When the ramp begins, the beam is deflected to the left of the CRT screen, and the VCO oscillates at 3.5 GHz. As the ramp amplitude grows, the beam is moving to the right of the screen, and f_{vco} increases. Suppose that at a given moment $f_{vco} = 3.6$ GHz at the output of the mixer. Then we have the following components: 1 GHz (f_{s1}), 2 GHz (f_{s2}), 3.6 GHz (f_{vco}), 2.6 GHz ($f_{vco} - f_{s1}$), 4.6 GHz ($f_{vco} + f_{s1}$), 1.6 GHz ($f_{vco} - f_{s2}$), and 5.6 GHz ($f_{vco} + f_{s2}$). Because the bandwidth of the bandpass filter is much less than 0.1 GHz, none of the components appear after the filter, and so no vertical deflection occurs at the screen. This is the case until f_{vco} reaches 4.5 GHz. Then the components at the output of the mixer are 1 GHz (f_{s1}), 2 GHz (f_{s2}), 4.5 GHz (f_{vco}), 3.5 GHz ($f_{vco} - f_{s1}$), 5.5 GHz ($f_{vco} + f_{s1}$), 2.5 GHz ($f_{vco} - f_{s2}$) and 6.5 GHz ($f_{vco} + f_{s2}$). The output of the bandpass filter is no longer zero. There is a component at 3.5 GHz ($f_{vco} - f_{s1}$) whose amplitude is proportional to the amplitude of the input signal component of 1 GHz. This produces a vertical deflection when the horizontal deflection produced by the ramp corresponds to 1 GHz. During the rest of the sweeping the vertical deflection would be zero except when the frequency of the VCO reaches 5.5 GHz. At that time the difference between f_{vco} and f_{s2} is within the band of the filter, and a vertical deflection proportional to the amplitude of f_{s2} appears on the screen.

Now we are ready to understand the function of the low-pass filter at the input. Suppose that a component at 8.5 GHz (f_{s3}) is present in the input signal and that the spectrum analyzer does not have the low-pass filter. When $f_{vco} = 5$ GHz the difference between f_{s3} and f_{vco} is 3.5 GHz and falls within the passband of the filter. So it produces a vertical deflection on the screen proportional to the amplitude of f_{s3} . The problem is that this component would be displayed when the ramp reaches the voltage level corresponding to 2 GHz. This would give the user the erroneous indication that a 2 GHz component is present in the input signal, instead of the 8.5 GHz component really present. Then the function of the low-pass filter is to prevent these high frequencies from getting to the mixer. The bandwidth of this filter should be equal to the range of the spectrum analyzer.

Superheterodyne spectrum analyzers are not real-time instruments. They need the input signal to remain unchangeable during the sweep time, and storage CRTs are necessary to display the spectrum of the input signal.

To resolve signals with closely spaced frequency components, spectrum analyzers have bandpass filters with band-

widths as narrow as 10 Hz. Such narrow filters are difficult (or impossible) to achieve, especially at high center frequencies as in our example at 3.5 GHz. Adding mixing stages solves this problem. Figure 8 shows a simplified block diagram of a spectrum analyzer with two mixing steps. The output of the bandpass filter centered at 1 MHz differs from zero only when $f_{vco} - f_{\text{signal}}$ is equal to the central frequency of the first bandpass filter (3.5 GHz in the example), plus or minus the bandwidth of the last bandpass filter (the bandwidth of the 1 MHz filter in the example).

In some spectrum analyzers the signal is converted after the last bandpass filter from analog to digital (at a much lower rate than if it were converted at the instrument input), and then digital filtering is performed, allowing implementation of very narrow filters.

Sensitivity is the measure of the smallest amplitude that the spectrum analyzer can display. The ultimate limitation in measuring a low-level signal is the random noise generated by the instrument itself. The thermal noise generated in the circuit elements is amplified by the different gain stages, added to the noise they generate, and displayed on the screen as a noise signal below which one cannot make measurements. The instrument sensitivity is determined by measuring the noise level on the display without any applied input signal. Signals at lower levels cannot be measured because they are masked by the noise. Even though the input attenuator and mixers have little effect on the actual system noise before the first gain stage, they do have a marked effect on the ability of the instrument to measure low-level signals because they attenuate the input and so they reduce the signal-to-noise ratio. Choosing the minimum input attenuation maximizes the instrument's sensitivity.

To minimize the internal spectrum analyzer's harmonic distortion, signal levels should be kept as low as possible at the input of the spectrum analyzer mixer. This means that to increase accuracy, the input attenuator of the spectrum analyzer must be used to decrease the level of the signal applied to the mixer when high-level signals are applied. However, this reduces the signal-to-noise ratio and so the instrument's sensitivity.

The bandwidth of the bandpass filter (resolution bandwidth) affects sensitivity. The spectrum analyzer generates random noise of constant amplitude over a wide range of frequencies. Because part of the internally generated noise is present at the input of the bandpass filter, the noise present at the output also decreases and sensitivity increases when the filter bandwidth decreases.

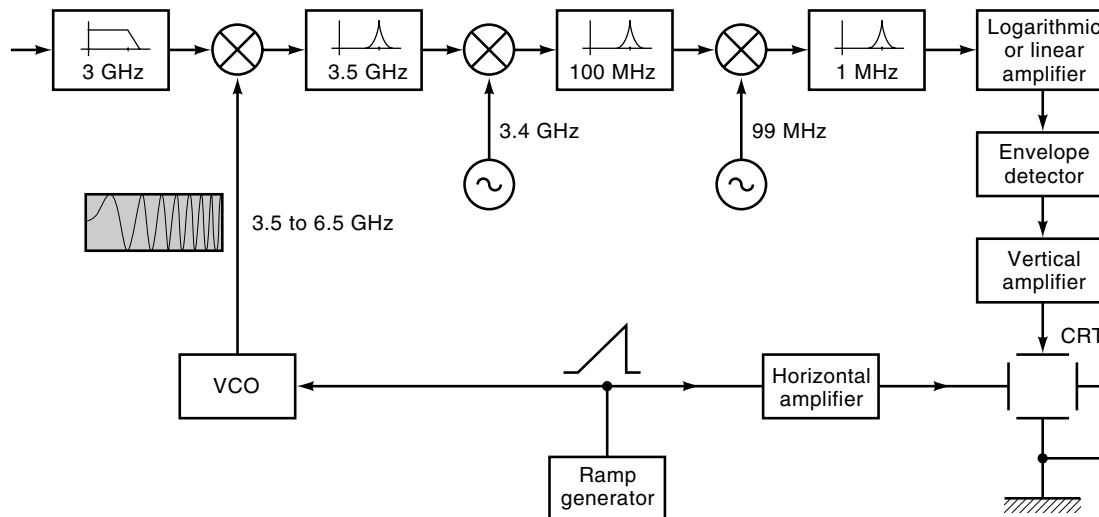


Figure 8. Block diagram of the superheterodyne spectrum analyzer with two mixing steps. The inclusion of multiple mixing stages in a superheterodyne spectrum analyzer allows high-resolution spectral analysis of high frequency voltages and also analog-to-digital conversion of the voltage representing the components of the input voltages.

The dynamic range of a spectrum analyzer is defined as the difference between its maximum input voltage and its noise level.

Frequency resolution is the ability of the spectrum analyzer to separate closely spaced input signal frequency components. It depends on the bandwidth of the narrowest filter in the chain (see Fig. 8). As the VCO is swept in frequency, the input of the bandpass filter is also swept. Unless two distinct input signal frequency components are far enough apart when compared with the filter bandwidth, the traces they produce on the screen fall on top of each other and look like only one response.

Band-pass band filters, as with band-limited circuits, require finite time to respond to an input stimulus. Because the rise time of a filter is inversely proportional to its bandwidth, the narrower the resolution of the filter, the greater the time it needs to respond to the input. If the VCO is swept too quickly, there is a loss of displayed amplitude in the screen of the spectrum analyzer. This means that when a narrow filter is selected, sweep time must increase, otherwise the instrument's accuracy is affected. It can be shown (3) that the sweep time must decrease with the square of the bandwidth to assure that the time when the mixer output is within the passband is of the order of magnitude of the rise time of the filter. This means that each time the resolution bandwidth is reduced by a factor of 10, the sweep time goes up by a factor of 100. If we select a very narrow filter, the sweep time becomes prohibitive. For instance, a bandwidth of 30 Hz in a 10 division display with 50 MHz/div selected, leads to a sweep time of 34 days!!! Some spectrum analyzers automatically set sweep time to the span and bandwidth resolutions selected to maintain the instrument's calibration. Others allow the user to select sweep time also, but when this is too small it indicates that the display is uncalibrated.

The amplitude accuracy of the spectrum analyzer depends on several factors. The input attenuator and the first mixer must present a flat frequency response over the entire band of the instrument. In a low-frequency instrument, ± 0.5 dB of deviation from a flat response is a typical value, but for a spectrum

analyzer with a frequency range of tens of GHz, ± 4 dB is an acceptable value. The fidelity of the logarithmic characteristic of the log amplifiers and the linearity of the envelope detector characteristic also affect amplitude accuracy. Impedance mismatch is also a source of error at high frequencies. Spectrum analyzers do not have perfect input impedances. In most cases, however, uncertainty is relatively small. When this is not the case, the use of a well-matched attenuator at the instrument input solves the problem.

The measurement of signal distortion must consider the characteristics of the signal. The change with time of the harmonic components of a signal may create four types of signals: (a) signals with quasi-stationary components; (b) signals with fluctuating components; (c) signals with rapidly changing components; and (d) signals with interharmonics and other spurious components. Continuous real-time analysis is required for (b). For (c) continuous real-time measurement is absolutely necessary because the value of each component is meaningful only when obtained through statistical analysis of a set of values measured over time. Very precise requirements are necessary to get reproducible results.

The use of a superheterodyne spectrum analyzer is not incompatible with real-time analysis, but it requires (1) the possibility of external control of the analyzer's local oscillator; (2) that the analyzer have an analog-to-digital converter at the output of the bandpass filter, a digital memory, and a CPU so that successive values of each frequency component may be stored and processed. These features are common in many spectrum analyzers now commercially available.

SYSTEMS DISTORTION MEASUREMENT

The evaluation of distortion introduced into electrical signals by systems, such as electronic devices and equipment, consists of one or several of the following basic measurements involving electric voltages:

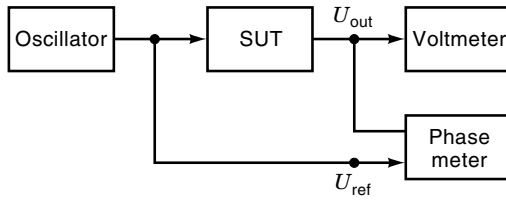


Figure 9. Setup for linear electric distortion measurement. The system under test (SUT) is subjected to a sine wave voltage. The voltmeter may be rms or peak responding. The function of the voltmeter and of the phase meter may be performed by a vector network analyzer.

1. Linear distortion (frequency and phase distortion): measurement of the amplitude and phase of a system's output voltage as a function of the frequency. The system is driven by a sine wave voltage whose amplitude is kept constant and whose frequency is swept in the range of interest. The output voltage amplitude and the phase shift between the input and output voltages are measured. Frequency distortion is evaluated by dividing the difference between the maximum and minimum indications of the voltmeter by half their sum. The result is expressed either as a percentage or in logarithmic units (dB). Delay distortion, expressed in time units, commonly μs or ms , is determined by dividing each phase shift (in rad) by the corresponding angular frequency (in rad/s) and selecting the maximum difference of the obtained ratios. The test setup for linear distortion measurements is shown in Fig. 9.
2. Harmonic distortion (nonlinear): measurement of the absolute values or values relative to the fundamental of the harmonics in the output voltage of the system when a sine wave voltage is applied to the system. In some cases, when the nonlinearity of the system depends heavily on frequency, the system is subjected to a band-limited noise signal. Harmonic distortion is usually expressed either as a percentage or in logarithmic units (dB or, sometimes, dBc, that is, dB relative to the fundamental, when the harmonics have small amplitudes compared to the fundamental). Figure 10 represents a possible test setup for harmonic distortion measurement.
3. Intermodulation distortion (nonlinear): the system is supplied with at least two sine waves of different frequencies, and the frequency components in the output voltage are measured. As in the case of harmonic distortion measurement, when the nonlinearity of the system depends heavily on frequency, the system is subjected to a band-limited noise signal. Intermodulation distortion is usually expressed either as a percentage or in logarithmic units (dB). The test equipment for intermodulation distortion measurement is shown in Fig. 11.

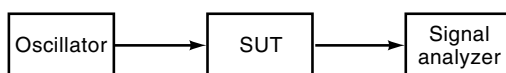


Figure 10. Setup for electric harmonic distortion measurement. The signal analyzer is generally a spectrum analyzer with an input voltage attenuator.

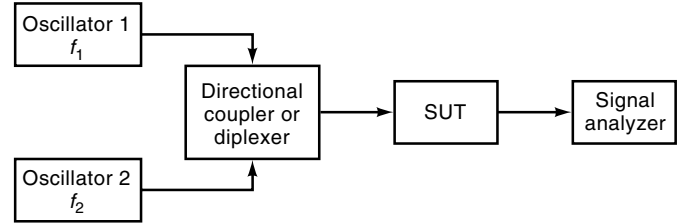


Figure 11. Setup for electric intermodulation distortion measurement. The two oscillators are connected to the system under test (SUT) through a directional coupler or a diplexer. The oscillators may be replaced by a multitone generator.

Testing to assess the performance of a system depends on the application. In sound systems, where harmonics and intermodulation products that fall in the audible spectrum produce distortion, harmonic and intermodulation distortion measurements are mandatory. Because the human ear is relatively insensitive to delay distortion, however, this type of electrical distortion needs no attention. In video and data signals, delay distortion constitutes one of the most limiting impairments, which means that it must be measured.

Following the ideas just presented, the measurement of electric distortion has been a subject of standardization. Thus the test of a loudspeaker, an audio amplifier, or a TV receiver involves many different distortion measurements according to specific methods and procedures. Standards IEC 268 Parts 3, 4, 5, and 6 for sound system equipment, IEC 244 Parts 4 and 4A for radio transmitters and IEC 244 Part 5 for television transmitters are texts where the reader may find useful information concerning the measurement of electric distortion in those systems.

The following are some considerations on equipment and measuring methods for distortion measurement:

1. The level of total harmonic distortion of the source of signals shall be at least 10 dB below the lowest level of distortion to be measured.
2. To correctly measure the distortion of a system, one must consider the distortion introduced by the test setup. For that purpose, it is good practice to calibrate the setup before testing the system under test (SUT). A measurement on the setup alone provides values that are used as correction factors.
3. The signal analyzer is often a spectrum analyzer. Network analyzers are also very much in use, particularly for testing RF and microwave devices (4). Available spectrum and network analyzers have three useful features: (a) an internal oscillator that may be used to excite the SUT; (b) digital interfaces for remote control that allow automated measurement procedures; (c) an internal CPU useful for reducing data and presenting distortion parameters.
4. Delay distortion is expressed by a parameter that is a function of the derivative of the phase with respect to frequency $d\theta/d\omega$. Several methods leading to different parameters are in use (4,5). One method very commonly implemented in network analyzers consists of measuring $d\theta/d\omega$ at two frequencies, one of which is a reference. Then delay distortion is expressed as the difference of the two derivatives.

Measuring the phase shift introduced by a system requires a reference signal. This may be a problem when testing tuners, for instance. The AM/FM-delay method for group delay measurement (4) is one solution to overcoming that difficulty.

Another delay distortion measurement problem arises when the input and output of the system are far apart as is the case with some communications systems. Solutions for this problem are discussed in (5).

Apart from the distortion due to the nonlinearity or the non-ideal frequency response of a system considered in this article, a signal may be distorted because of the interference of signals of the same system, for example, cross talk and cross-modulation. All of these types of distortions are included in what we designate an intrasystem distortion. A signal in a system, however, may be distorted by a signal from another system. This type of distortion that is designated as intersystem distortion is produced when a coupling between the two systems exists either by conduction or radiation. Intersystem distortion measurement and evaluation is an electromagnetic interference (EMI) or electromagnetic compatibility (EMC) problem and is thus beyond the scope of the present article. EMI/EMC is presently an extremely important domain of electrical engineering and it will be even more important in the future. The proper operation of electrical and electronic equipment requires increased attention to interference and susceptibility issues. Interested readers may refer to Refs. 6 and 7.

Legislation and standards for measurements on EMI/EMC have been produced all over the world. References 8–14, are examples of standards that may assist the reader in evaluating intersystem distortion. For a more detailed list, readers should consult Ref. 7.

BIBLIOGRAPHY

1. Standard IEC 268-2, Sound system equipment—explanation of general terms, International Electrotechnical Commission, Geneva, Switzerland, 1987.
2. R. L. Freeman, *Telecommunications Transmission Handbook*, New York: Wiley, 1991.
3. Spectrum Analysis Basics, Application Note 150, Hewlett-Packard Company, Palo Alto, CA, 1989.
4. RF and microwave device test for the 90s, *Seminar Papers*, Hewlett-Packard, Palo Alto, CA, 1995.
5. B. M. Oliver and J. M. Cage, *Electronic Measurements and Instrumentation*, New York: McGraw-Hill, 1975.
6. B. E. Keyser, *Principles of Electromagnetic Compatibility*, Norwood, MA: Artech House, 1985.
7. V. P. Kodali, *Engineering Electromagnetic Compatibility*, Principles, Measurements and Technology, Piscataway, NJ: IEEE Press, 1996.
8. Standard EN 50082-1, *Electromagnetic compatibility general immunity standard*. Part 1: Residential, commercial and light industry, Electrotechnical Standardization European Committee, Brussels, Belgium.
9. Standard EN 50082-2, *Electromagnetic compatibility general immunity standard*. Part 2: Industrial environment, Electrotechnical Standardization European Committee, Brussels, Belgium.
10. Standard CISPR 20, *Limits and methods of measurement of the immunity characteristics of radio broadcast and television receivers and associated equipment*, International Electrotechnical Commission, Geneva, Switzerland, 1990.
11. Standard CISPR 22, *Limits and methods of measurement of radio disturbance characteristics of information technology equipment*, International Electrotechnical Commission, Geneva, Switzerland, 1993.
12. Standard IEC 555-2, *Disturbances in supply systems caused by household appliances and similar electrical equipment*. Part 2: Harmonics, International Electrotechnical Commission, Geneva, Switzerland, 1982.
13. Standard IEC 555-3, *Disturbances in supply systems caused by household appliances and similar electrical equipment*. Part 3: Voltage fluctuations, International Electrotechnical Commission, Geneva, Switzerland, 1982.
14. Standards IEC 1000-4-x, *EMC: Test and measurement techniques*, International Electrotechnical Commission, Geneva, Switzerland.
15. S. Haykin, *Communication Systems*, New York: Wiley, 1994.
16. Joseph J. Carr, *Elements of Electronic Instrumentation and Measurement*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.
17. A. D. Helfrick and W. D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

PEDRO M. B. SILVA GIRÃO
ANTÔNIO M. CRUZ SERRA
HELENA M. GEIRINHAS RAMOS
Instituto Superior Técnico
Department of Electrical and
Computer Engineering

INTERMODULATION MEASUREMENT

Virtually all electronic circuits and systems exhibit nonlinear input-output transfer characteristic. Mixers, frequency multipliers, modulators, and square-law detectors represent examples of intentional class members, while linear power amplifiers, active filters, and microwave transmitters, in which nonlinearity represents an undesirable deviation of the system from ideal, or linear, operation, are examples of unintentional members.

Whenever a number of signals of differing frequencies pass through a nonlinear device, energy is transferred to frequencies that are sums and differences of the original frequencies. These are the intermodulation products (IMPs). In such cases, the instantaneous level of one signal may effectively modulate the level of another signal, whence the name *intermodulation*. It is interesting to note that the ear's cochlea has a similar nonlinear response and produces sums and differences of the input frequencies in the same way, particularly with loud sounds (1).

It has also been found that passive components, normally considered to be linear, can also generate IMPs. A variety of situations can arise in which nonlinear-resistance junctions can be formed at metallic mating surfaces. Such junctions may result from salt or chemical depositions or from corrosion. The result is sometimes known as the "rusty bolt effect" because rusted bolts in structures have been known to exhibit such nonlinearities. This phenomenon is referred to as passive intermodulation (PIM). Sources of PIM include waveguides, directional couplers, duplexers, and antennas (2–6).

Intermodulation may also occur at the amplifier-loudspeaker interface (7), or in general due to the nonlinear interaction between the input signal of a two-port network and a signal injected to the output port and propagating into the

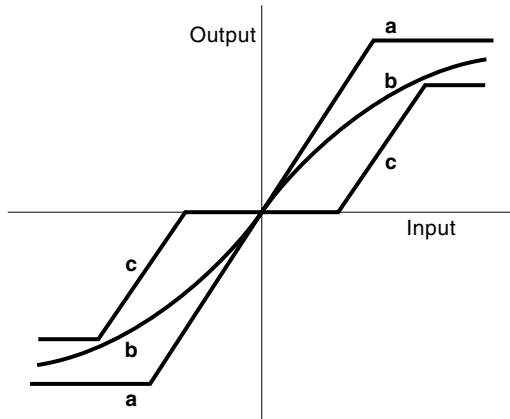


Figure 1. Different types of static nonlinearities: (a) clipping. (b) soft. (c) crossover.

input via a feedback network (8). Externally induced transmitter intermodulation, also known as reverse intermodulation, back intermodulation, and antenna-induced intermodulation, is the mixing of a carrier frequency with one or more interfering signals in a transmitter's final stage (9). Moreover, lack of screening of open-wire transmission lines can result in significant coupling to adjacent lines, frequently giving rise to intermodulation products (10). Furthermore, intermodulation may arise when an array of receiving antennas is illuminated with a transient impulsive electromagnetic plane wave (11).

In discussing the sources of IMPs it is convenient to divide nonlinear mechanisms yielding IMPs into two principal forms. The first is due to a nonlinear amplitude input/output characteristic (AM/AM), which causes amplitude compression with increasing input amplitude. The second mechanism occurs because of the variation of phase shift through the device, or the system, as the input amplitude is changed (AM/PM).

Depending on the signal characteristics, sources of IMPs can be divided into two categories: (1) static nonlinearity, depending solely on the amplitude of the signal, and (2) dynamic nonlinearity, depending not only on the amplitude but also on the time properties or frequency composition of the signal.

Static nonlinearities usually encountered in electronic circuits and systems can be classified into clipping, crossover, and soft nonlinearities (12), as shown in Fig. 1. In contrast with the *hard* nonlinearities of clipping (which is significant near maximum input amplitudes) and crossover (mostly significant at small input amplitudes), soft nonlinearity is usually the most important in the transfer characteristic of an electronic circuit. If the frequency content or the time properties of the input signal affect the transfer characteristic of the circuit or the system, the resulting nonlinearities may be called *dynamic*. Intermodulation products resulting from dynamic nonlinearities are referred to as transient intermodulation (TIM), slew-induced distortion (SID), or dynamic intermodulation distortion (DIM) (13–16).

SIMPLE INTERMODULATION-PRODUCT THEORY

IMPs occur when two or more signals exist simultaneously in a nonlinear environment. In general, if N signals, with frequencies f_1 to f_N , are combined in a static nonlinearity, the output

will contain spectral components at frequencies given by

$$f = \sum_{n=1}^N k_n f_n$$

where k_n is a positive integer, a negative integer, or zero and $\sum_{n=1}^N |k_n|$ is the order of the IMP. Even with a small number of input signals, N , a very large number of IMPs are generated. Fortunately, not all products are equally troublesome. Depending on the system involved, some of these IMPs can be neglected, since they will be filtered out at some point. For example, most communication systems operate over a limited frequency band. Thus, IMPs falling out of the band will be attenuated. Moreover, the amplitudes of the IMPs generally decrease with increasing order of the products, and high-order products can often be neglected. Low-order intermodulation components such as the second-order components $f_m - f_n$ and $f_m + f_n$ and the third-order components occurring at frequencies $2f_m - f_n$ and $f_m + f_n - f_q$ are usually the most troublesome, having the largest magnitudes and/or lying close to the originating frequencies, which makes their removal by filtering practically difficult. However, a salient characteristic of PIM, as distinguished from the conventional IM counterpart discussed in the preceding, is that the PIMs causing trouble are of a high order, say the eleventh through the twenty-first.

Analysis of nonlinear systems differs from that of linear systems in several respects: There is no single analytical approach which is generally applicable (such as Fourier or Laplace transforms in linear systems); closed-form analytical solutions of nonlinear equations are not ordinarily available; and there is rarely sufficient information available to enable a set of equations which accurately model the system to be derived. These factors preclude the exact analytical determination of nonlinear effects, such as IMPs, in the general case. In order to get anything done at all, it is usually necessary to make various simplifying assumptions and then to use an approximate model which will provide results of acceptable accuracy for the problem in hand.

A simple approach, therefore, is to use frequency-domain techniques, which provide a separate solution for each frequency present in the output. In general, such methods are (1) centered around a description of the nonlinear mechanism by a continuous-function characteristic (for example, a polynomial or a Fourier-series representation of the output in terms of the input) and (2) based on the simplifying assumption that this characteristic does not vary with frequency—that is, it is a memoryless characteristic.

Memoryless nonlinear circuits are often modeled with a power series of the form

$$V_{\text{out}} = \sum_{n=0}^N k_n V_i^n \quad (1)$$

The first coefficient, k_0 , represents the dc offset in the circuit. The second coefficient, k_1 , is the gain of the circuit associated with linear circuit theory. The remaining coefficients, k_2 and above, represent the nonlinear behavior of the circuit. If the circuit were completely linear, all of the coefficients except k_1 would be zero.

The model can be simplified by ignoring the terms that come after the k_3 term. For soft nonlinearities, the size of k_n decreases

rapidly as n gets larger. For many applications the reduced model of Eq. (2) is sufficient, since the second-order and third-order effects dominate. However, there are many devices, circuits, and systems presenting difficulties for the polynomial approximation.

We have, then,

$$V_{\text{out}} = k_0 + k_1 V_i + k_2 V_i^2 + k_3 V_i^3 \quad (2)$$

Assuming that the input is a two-tone signal of the form

$$V_i = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t \quad (3)$$

then combining Eqs. (2) and (3) yields

$$\begin{aligned} V_{\text{out}} = & a_0 + b_1 \cos \omega_1 t + c_1 \cos \omega_2 t + b_2 \cos 2\omega_1 t + c_2 \cos 2\omega_2 t \\ & + b_3 \cos(\omega_1 + \omega_2)t + c_3 \cos(\omega_1 - \omega_2)t + b_4 \cos 3\omega_1 t \\ & + c_4 \cos 3\omega_2 t + b_5 \cos[(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t] \\ & + c_5 [\cos(2\omega_2 + \omega_1)t + \cos(2\omega_2 - \omega_1)t] \end{aligned} \quad (4)$$

where

$$a_0 = k_0 + \frac{k_2}{2}(V_1^2 + V_2^2)$$

$$b_1 = k_1 V_1 + \frac{3}{4}k_3 V_1^3 + \frac{3}{2}k_3 V_1 V_2^2$$

$$c_1 = k_1 V_2 + \frac{3}{4}k_3 V_2^3 + \frac{3}{2}k_3 V_1^2 V_2$$

$$b_2 = \frac{1}{2}k_2 V_1^2$$

$$c_2 = \frac{1}{2}k_2 V_2^2$$

$$b_3 = c_3 = k_2 V_1 V_2$$

$$b_4 = \frac{1}{4}k_3 V_1^3$$

$$c_4 = \frac{1}{4}k_3 V_2^3$$

$$b_5 = \frac{3}{4}k_3 V_1^2 V_2$$

$$c_5 = \frac{3}{4}k_3 V_1 V_2^2$$

For equal-amplitude input tones, Eq. (4) shows that the second-order terms, of amplitudes b_2 , c_2 , b_3 , and c_3 , will be increased 2 dB in amplitude when input tones are increased by 1 dB. Third-order terms, of amplitudes b_4 , c_4 , b_5 , and c_5 , are increased 3 dB in amplitude when input tones are increased by 1 dB.

SIMPLE TRANSIENT-INTERMODULATION THEORY

To illustrate how TIM distortion arises, consider a differential amplifier with negative feedback applied between the output and the inverting input and with a voltage step applied to the noninverting input. If the open-loop gain of the amplifier were flat and the time delay through it were zero, the voltage step would instantaneously propagate undistorted through the

amplifier, back through the feedback loop, and into the inverting input. There it would be subtracted from the input signal, and the difference signal, which is a voltage step occurring at the same time that the input voltage does, would be amplified by the amplifier. However, this is not the case when the open-loop gain of the amplifier is not flat and the time delay through it is not zero. When the voltage step occurs, the limited high-frequency response of the amplifier prevents the appearance of a signal at the amplifier output terminal until the internal capacitors of the amplifier can charge or discharge. This causes the momentary absence of a feedback signal at the inverting input to the amplifier, possibly causing the amplifier to severely overload until the feedback signal arrives.

If the input signal to the differential amplifier is formed from a sine wave superimposed on a square wave, the amplifier will exhibit the same response to the abrupt level changes in the square wave as it did to the voltage step discussed in the preceding. During the momentary absence of the feedback when the square wave changes level, the amplifier can either saturate or cut off. If this occurs, the sine wave momentarily disappears from the signal at the output terminal of the amplifier, or it momentarily decreases in amplitude. This happens because the saturated or cut-off amplifier appears as a short circuit or open circuit, respectively, to the sine wave, and this component of the input signal is interrupted from the output signal, thereby resulting in TIM (16).

A point to be noted is that if the term TIM were understood literally, it would imply transients of both high and low frequencies and/or high or low operating levels—in other words, all transients. In actual practice, however, TIM occurs only for signals with simultaneous high level and high frequencies, not lower levels or lower frequencies. The key characteristic of such signals is high signal slopes, not just high frequencies or high levels. Neither high frequencies nor high levels in themselves necessarily result in distortion, unless their combination is such that a high effective signal slope is produced. TIM is actually generated when the signal slope approaches or exceeds the amplifier slew rate. This can happen for either transient or steady-state signals. Thus, a more easily understood term to describe what actually happens would be one which relates slew rate and signal slope, such as *slew-induced distortion* (SID). Other descriptive variations of the terminology are “slew-rate distortion” and “slewing distortion” (17).

Due to the complexity of the mechanism resulting in TIM, especially the frequency dependence of the amplifier and the incorporation of the feedback, the use of Volterra series to characterize the output as a function of the input (18) is needed to provide closed-form expressions for all the distortion products. However, techniques using Volterra series suffer from the disadvantage that a complex mathematical procedure is required to obtain a closed-form expression for the output amplitude associated with a single component of the output spectrum. Moreover, the problem of obtaining output products of orders higher than the third becomes prohibitively difficult unless it may be assumed that higher-order contributions vanish rapidly (19). The Volterra series approach is therefore most applicable to mild nonlinearities where low-order Volterra kernels can adequately model the circuit behavior.

With appropriate assumptions and simplifications, many useful features of the Volterra series technique can be used to find approximate expressions for TIM (SID). These are quite accurate for small distortion conditions (20,21).

INTERMODULATION CHARACTERIZATION

Although it is important to understand the origin of intermodulation and the engineering techniques for avoiding it, it is equally important to be able to characterize it objectively, preferably in a way that correlates well with the subjective perception of the intermodulation. Being able to characterize an imperfection in this way is an important step toward eliminating it as a system performance degradation.

Several techniques for characterizing intermodulation distortion have been proposed. While some of these techniques measure the total intermodulation distortion, others distinguish between the various intermodulation products. The latter are much to be preferred, because subjective perception of intermodulation shows that equal amounts of total intermodulation distortion differ widely in their effect according to how the total is made up.

Depending on the signal characteristics, techniques for characterization of intermodulation distortion can be classified into two categories: (1) steady-state techniques, where characterization is performed on the assumption that the input to the system under consideration is a multisinusoidal signal, and (2) dynamic techniques, where characterization is performed on the assumption that the input to the system under consideration is formed of a sinusoidal signal superimposed on another signal characterized by rapid changes of state—for example, a square wave.

Steady-State Techniques

The Intercept Point. Increasing the signal level at the input to a weakly nonlinear device will cause the IMPs to increase at the output (22). In fact, the increase in the amplitudes of the IMPs is faster than the increase in the output version of the input signal. For increasing fundamental input power, the fundamental output power increases in a linear manner, according to the gain or loss of the device. At some point, gain compression occurs and the fundamental output power no longer increases with input power. The output power of the second-order intermodulation products also increases with fundamental input power, but at a faster rate. Recall that the second-order intermodulation changes 2 dB per 1 dB of change in the fundamental. Similarly, the third-order intermodulation changes 3 dB per 1 dB of change in the fundamental. Thus, on a logarithmic scale, as shown in Fig. 2, the lines representing the second- and third-order intermodulation products have twice and three times, respectively, the slope of the fundamental line.

If there were no gain compression, the fundamental input power could be increased until the second-order intermodulation eventually caught up with it and the two output power levels were equal. This point is referred to as the second-order intercept point. The third-order intermodulation product also increases faster than the fundamental, and those two lines will intersect at the third-order intercept point. Rarely can either of these two points be measured directly, due to the gain compression of the fundamental. Instead, the intercept points are extrapolated from measurements of the fundamental and intermodulation products at power levels below where gain compression occurs. The intercept points are usually specified in dBm and may refer either to the output or to the input; the two points will differ by the gain of the system under consideration. The second-order and third-order intercept points are figures of

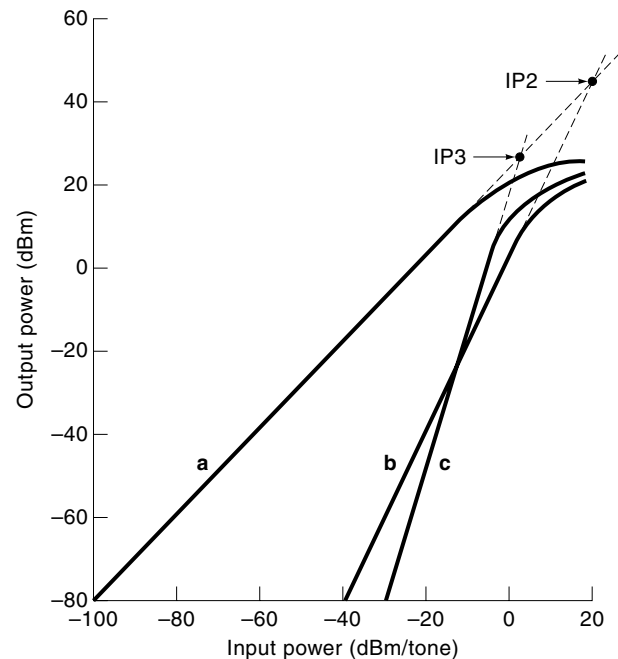


Figure 2. Third-order and second-order intercept points are determined by extending the fundamental, the second-order, and the third-order intermodulation transfer function lines. (a) Fundamental transfer function, slope = 1; (b) second-order intermodulation, slope = 2; (c) third-order intermodulation, slope = 3. IP3, third-order intercept point; IP2, second-order intercept point.

merit which are independent of the signal level. Therefore, the intermodulation performance of two different systems can be compared quite easily if their intercept points are known (22).

Using the intercept point, it is easy to calculate the relative intermodulation level corresponding to a given input signal level. In fact, the difference between the level of the second-order intermodulation and the fundamental signal level is the same as the difference between the fundamental signal level and the intercept point. Thus, if the second-order intercept point is +15 dBm and the fundamental signal level is -10 dBm (both referred to the output of the device), the difference between these two values is 25 dB. Therefore, the second-order intermodulation products will be 25 dB below the fundamental, or -35 dBm. So the intercept point allows easy conversion between fundamental signal level and the intermodulation level.

The difference between the level of the third-order intermodulation products and the fundamental signal level is twice the difference between the fundamental signal level and the third-order intercept point. (Note that the second-order intercept point is not the same as the third-order intercept point.) Suppose that the third-order intercept point is +5 dBm and the fundamental signal is -25 dBm, both referred to the output of the device. The difference between the intercept point and the fundamental is 30 dB, so the third-order intermodulation products will be two times 30 dB down from the fundamental. The relative distortion level is -60 dB, and the absolute power of the intermodulation products is -85 dBm.

It is important, however, to note that the preceding analyses assume that the second-order and third-order intermodulation curves have slopes of 2 dB/dB and 3 dB/dB respectively.

Thus, theoretically, the intercept points are not functions of the input power level. If a power sweep is performed, it is expected that the intercept points will remain constant. The intercept points can therefore be calculated from measurements at only one power level. However, if the input signal exceeds a certain limit, the amplitudes of the output fundamentals and the resulting intermodulation products will start to saturate, and the intercept points will usually drop off, indicating an invalid measurement. It is essential to know this limit. Expanding the model of Eq. (2) to include fourth- and fifth-order terms (23) can do this.

Moreover, at low power levels, the intercept points will start to change as the noise floor of the measuring instrument (usually a spectrum analyzer) is approached, thus indicating an invalid measurement. It is important, therefore, to look at the variation of the intercept points as functions of power, as this provides a good way of checking the valid measurement range. This is particularly useful for large-dynamic-range circuits and systems with low output power where the intermodulation is low, but only for signals that are low enough.

Two-Tone Test. The two-tone test is extensively used in characterizing a wide range of devices. A few examples are (a) magnetic tapes (24), (b) microwave- and millimeter-wave diode detectors (25), (c) analog-to-digital converters (26, 27), (d) gamma correctors (28), and (e) electrical components such as resistors, capacitors, and inductors, as well as contacts of switches, connectors, and relays (29). The two-tone test is also used to characterize the performance of the basilar membrane of the cochlea (30).

The two-tone test can also be used to determine the transfer characteristic of a nonlinear device modeled by the polynomial approximation of Eq. (2). With the input formed of two properly selected frequencies ω_1 and ω_2 , if the second-order and third-order intermodulation products are measured separately, then it is possible to find, from the measured data, the coefficients of the quadratic and cubic terms, k_2 and k_3 respectively, in the polynomial approximation of Eq. (2). If, in addition, the IMPs are measured at two sets of values of ω_1 and ω_2 , then it is possible to identify the dominant physical nonlinear process from the variation of IMPs with test frequencies (13).

The two-tone test can also be used to determine the complex transfer characteristic of a nonlinear device exhibiting AM/AM nonlinearity only, with fixed phase shift between the output and the input. In this case a complete set of measurements, for all the two-tone intermodulation products produced by the nonlinearity, at two different power levels is necessary (31). If the device under consideration exhibits both AM/AM and AM/PM nonlinearities, then determination of a unique set of polynomial coefficients requires a complete set of intermodulation measurements at three different power levels (31). The set obtained at the highest power level will decide the amplitude range within which the characterization will be valid.

Due to the basic assumption that the nonlinearities are represented by polynomials, high-accuracy representation of the device characteristics will require difficult accurate measurements of higher-order intermodulation products, in addition to the increased complications and considerable efforts involved in the analysis (31). Another difficulty from which this method suffers arises from the necessity of measuring complete sets of two-tone intermodulation products which are spread over a wide frequency range and consequently may put stringent

specifications on the measuring instruments and techniques if accurate measurements are to be achieved.

In the two-tone test, use is made of one of the in-band IMPs to describe a device, a circuit, or a system nonlinearity. Measurements are made in or near the frequency range of interest. In this test, the input signal consists of two frequencies, ω_1 and ω_2 , of equal amplitude and a fixed amount of frequency spacing. At the output of the circuit or the system under test, the amplitudes of the third-order intermodulation products, $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, are measured. The intermodulation distortion is defined as the ratio between the root sum square of the intermodulation products and the root sum square of the twin-tone amplitudes.

Unless a wave analyzer or a spectrum analyzer is available, the implementation of the two-tone test invariably requires amplification of the whole output spectrum to get the components ω_1 and ω_2 at a normalized value (100%). Then, ω_1 and ω_2 are suppressed, and the remaining components $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are measured with an ac voltmeter or oscilloscope. Especially at audio frequencies, this approach requires steep filters, one set of filters for each set of ω_1 and ω_2 . For the same reason, $\omega_2 - \omega_1$ cannot be too low. So it will never be a real narrow-band system. This narrowband aspect is particularly important for higher frequencies, where equalizers, in the reproduction audio channel, may give unequal amplification of the components in the spectrum (32).

In the audio-frequency range, a new version of the two-tone test overcomes the above-mentioned disadvantages (32). This is based upon the multiplication of the spectrum by itself. Thus, if the two-tone input signal is given by Eq. (3), with $V_1 = V_2 = V$, then multiplying the input spectrum by itself—that is, squaring—yields

$$\begin{aligned} V_i^2 &= V^2(\cos \omega_1 t + \cos \omega_2 t)^2 \\ &= V^2 \left(1 + \frac{1}{2} \cos 2\omega_1 + \frac{1}{2} \cos 2\omega_2 \right. \\ &\quad \left. + \cos(\omega_1 + \omega_2) + \cos(\omega_1 - \omega_2) \right) \end{aligned} \quad (5)$$

Assuming that the system under test is narrowband, Eq. (4) reduces to

$$\begin{aligned} V_{\text{out}} &= b_1 \cos \omega_1 t + c_1 \cos \omega_2 t + b_5 \cos(2\omega_1 - \omega_2)t \\ &\quad + c_5 \cos(2\omega_2 - \omega_1)t \end{aligned} \quad (6)$$

Multiplying the output spectrum by itself—that is, squaring—yields

$$\begin{aligned} V_{\text{out}}^2 &= [b_1 \cos \omega_1 t + c_1 \cos \omega_2 t + b_5 \cos(2\omega_1 - \omega_2)t \\ &\quad + c_5 \cos(2\omega_2 - \omega_1)t]^2 \end{aligned} \quad (7)$$

$$\begin{aligned} &= \frac{1}{2}(b_1^2 + c_1^2 + b_5^2 + c_5^2) + \left(\frac{1}{2}b_1^2 + c_1b_5 \right) \cos 2\omega_1 t \\ &\quad + \left(\frac{1}{2}c_1^2 + b_1c_5 \right) \cos 2\omega_2 t + \frac{1}{2}b_5^2 \cos(4\omega_1 - 2\omega_2)t \\ &\quad + \frac{1}{2}c_5^2 \cos(4\omega_2 - 2\omega_1)t + (b_1c_1 + b_5c_5) \cos(\omega_1 + \omega_2)t \\ &\quad + (b_1c_1 + b_1b_5 + c_1c_5) \cos(\omega_1 - \omega_2)t + b_1b_5 \cos(3\omega_1 - \omega_2)t \\ &\quad + c_1c_5 \cos(3\omega_2 - \omega_1)t + (b_1c_5 - c_1b_5) \cos(2\omega_1 - 2\omega_2)t \\ &\quad + b_5c_5 \cos(3\omega_1 - 3\omega_2)t \end{aligned} \quad (8)$$

where

$$b_1 = c_1 = k_1 V + \frac{9}{4} k_3 V^3$$

and

$$b_5 = c_5 = \frac{3}{4} k_3 V^3$$

Inspection of the spectra of Eqs. (5) and (8) shows that:

1. Both spectra are split into two parts, a lower region and a higher region. All components in the lower region are unaffected by the choice of ω_1 and ω_2 . They are affected only by the difference in frequencies $\omega_2 - \omega_1$. This means that as long as $\omega_2 - \omega_1$ is kept constant, the lower region of the spectrum will not be affected if measurement is swept through the whole band.
2. If the distortion is not too high—that is, for $k_3 \ll k_1$ —then the dc component is essentially constant. Also, Eqs. (5) and (8) show that the amplitudes of the components at frequencies ω_1 and ω_2 are equal to the dc component. In fact this dc component represents the 100% level.
3. The amplitude of the components at frequencies $3\omega_2 - \omega_1$ and $3\omega_1 - \omega_2$ is given by $b_1 b_5 = c_1 c_5$. Thus if $b_1 = c_1$ are normalized to unity—that is, the dc component is unity—then the amplitude of these frequency components is directly proportional to the amplitude of the third-order IMPs at frequencies $2\omega_2 - \omega_1$ and $2\omega_1 - \omega_2$.

Thus the IMP measurement reduces to Ref. 32:

1. Squaring
2. Keeping the dc level to a normalized value
3. Filtering out the component at frequency $3\omega_2 - \omega_1$ or $3\omega_1 - \omega_2$
4. Measuring the result with an ac voltmeter

While this approach offers simpler and more flexible instrumentation, its application is limited to low-distortion circuits and systems.

Three standard two-tone test methods are in common use for testing audio-frequency circuits and systems. These are the SMPTE, the CCIF, and the IEC intermodulation tests.

SMPTE Intermodulation Test. In the SMPTE (Society of Motion Picture and Television Engineers) test of intermodulation distortion, the system input is a combination of a large-amplitude low-frequency sine wave with a small-amplitude high-frequency sine wave (33). Often the large-amplitude component is 80% of the rated maximum input amplitude, and its frequency ω_1 is either 50 Hz or 60 Hz. The small-amplitude component is often 20% of the rated maximum input amplitude and therefore falls within the small-signal regime of the system operation; its frequency ω_2 is often 7 kHz. The large sine wave excites nonlinearities in the system under test, and hence it modulates the small-signal transfer function. Because the two input components are independent, the response of the system under test (in the presence of the large component) is

effectively the response of a linear time-varying network. The SMPTE intermodulation test quantifies nonlinearity by reference to the amplitude modulation of the small-amplitude high-frequency component as it appears at the output of the system under test.

CCIF Intermodulation Test. In the CCIF (International Telephonic Consultative Committee) test of intermodulation distortion, the input to the system under consideration is a mixture of two sinusoids, each 50% of full rated amplitude, one at $\omega_1 = 14$ kHz and the other at $\omega_2 = 15$ kHz (34). Intermodulation distortion is quantified by reference to the amplitude of the resulting $\omega_2 - \omega_1 = 1$ kHz difference-frequency tone at the output.

The CCIF intermodulation test has been successfully used for measurement of high-frequency distortion in audio systems. But it is sensitive only to asymmetrical distortion mechanisms which produce even-order products. If a spectrum analyzer, or other sharp-cutoff filtering technique, is used to look at the odd-order products, $2\omega_2 - \omega_1$ and $2\omega_1 - \omega_2$, as well, which in this case lie at 16 kHz and 13 kHz, then the test is also sensitive to symmetrical distortion mechanisms.

IEC Total Difference-Frequency Test. In the IEC (International Electrotechnical Commission) total difference-frequency distortion test, the input is a mixture of two sinusoids, each 50% of full rated amplitude, and angular frequencies ω_1 and ω_2 chosen such that $\omega_1 = 2\omega_0$ and $\omega_2 = 3\omega_0$. Intermodulation distortion is quantified by reference to the amplitude of the difference-frequency tone at ω_0 . In the original proposal for the test, ω_0 was chosen to correspond to 5 kHz; this has been changed to 4 kHz (35).

The IEC intermodulation test is fully in-band and detects both even-order and odd-order nonlinearities. However, it does not distinguish between them. A modified version of this test using $\omega_1 = 7\omega_0$ and $\omega_2 = 5\omega_0$ results in second-order IMPs at $2\omega_0$ and $12\omega_0$, and it results in third-order IMPs at $3\omega_0$, $9\omega_0$, $17\omega_0$, and $19\omega_0$. Thus, all the IMPs are well separated in frequency from each other and from harmonics of the test signals. Proper selection of ω_0 results in ω_1 and ω_2 , and at least one second-order and one third-order product fall within the bandwidth of the system under consideration (13).

Three-Tone Test. In this test, again, specific in-band IMPs are selected to characterize the overall system nonlinearities (36). The more even spectral distribution and flexibility, while still allowing discrete frequency evaluation, make this an attractive test for multifrequency systems such as communication and cable television systems.

In this test, three equal-amplitude tones are applied to the input of the nonlinear system under consideration. Thus,

$$V_i = V(\cos \omega_1 t + \cos \omega_2 t + \cos \omega_3 t) \quad (9)$$

Combining Eqs. (2) and (9) and using simple trigonometric identities, it is easy to show that the third-order term, $k_3 V_i^3$, will contribute to the output spectrum the following:

1. Three components at frequencies ω_1 , ω_2 , and ω_3 , each with amplitude given by

$$A_1 = \frac{15}{4} k_3 V^3 \quad (10)$$

2. Three components at frequencies $3\omega_1$, $3\omega_2$, and $3\omega_3$, each with amplitude given by

$$A_3 = \frac{1}{4}k_3 V^3 \quad (11)$$

3. Twelve components at frequencies $2\omega_m \pm \omega_n$, $m, n = 1-3$, each with amplitude given by

$$A_{21} = \frac{3}{4}k_3 V^3 \quad (12)$$

4. Four components at frequencies $\omega_m + \omega_n \pm \omega_p$, $m, n, p = 1-3$, each with amplitude given by

$$A_{111} = \frac{3}{2}k_3 V^3 \quad (13)$$

Equations (12) and (13) show that an intermodulation product of frequency $\omega_m + \omega_n \pm \omega_p$ is 6 dB higher than an intermodulation product of frequency $2\omega_m \pm \omega_n$.

The intermodulation distortion is defined as the ratio of the amplitude of one of the intermodulation products of $\omega_m + \omega_n \pm \omega_p$ to the amplitude of one of the three output tones.

In this test the choice of frequencies ω_1 , ω_2 , and ω_3 used to make the measurement is important. This is because a system's intermodulation performance may not be constant over its operating frequency range.

The three-tone test is widely used in characterizing the performance of RF amplifiers used in television broadcast transposers, where the vision carrier, color subcarrier, and sound carrier frequency components interact in the presence of amplifier nonlinearities. If the three frequency components are represented as single frequencies— ω_v the vision carrier, ω_{sc} the color subcarrier, and ω_s the sound carrier with amplitudes V_v , V_{sc} , and V_s , respectively—then the input signal can be expressed as

$$V_i = V_v \cos \omega_v t + V_{sc} \cos \omega_{sc} t + V_s \cos \omega_s t \quad (14)$$

Combining Eqs. (2) and (14) and using simple trigonometric identities, it is easy to show that the third-order term of Eq. (2) produces, among others, two in-band intermodulation components given by

$$V_{ip} = \frac{3}{2}k_3 V_v V_{sc} V_s \cos(\omega_v + \omega_s - \omega_{sc})t + \frac{3}{4}k_3 V_s V_{sc}^2 \cos(2\omega_{sc} - \omega_s)t \quad (15)$$

The intermodulation performance of the transposer is measured by taking it out of service and using the three-tone simulation of a composite video and sound signal, given by Eq. (14), as its input. The three levels and frequencies vary from system to system. Typical levels, below the peak synchronous pulse level, are $V_v = -6$ dB, $V_{sc} = -17$ dB, and $V_s = -10$ dB. Under these conditions, the first term of Eq. (15) is the most visible, and the second term will be much lower in amplitude, typically 17 dB less. Using a spectrum analyzer, the relative amplitude of the major in-band intermodulation is measured and referenced to the level of peak synchronous pulse. Usually, the permissible level of the major in-band intermodulation component is -53 dB below the reference level.

This three-tone test method is slow and requires spectrum analyzers with large dynamic ranges. Moreover, it measures

the system performance at one luminance level and one chrominance level only. Thus, it does not test the system over its full operating range (37).

The inadequacy of the internationally accepted three-tone test method can be overcome by using a modified color-bar test signal (37). The color bars are applied to the transposer via a test transmitter. The color bars and sound carrier therefore apply the three tones to the transposer, changing levels in rapid succession. With suitable processing, based on sampling the demodulated color-bar signal for short intervals corresponding to a selected color, intermodulation can be measured simultaneously at seven different luminance levels (37).

Four-Tone Test. This test is commonly used in voice-frequency circuits. In it, the input consists of two pairs of tones, thus approximating an input with Gaussian amplitude distribution, which is a more realistic representation of real-life signals. The frequencies of the input tones are selected to generate second- and third-order intermodulation products within the bandwidth of the device under test. These intermodulation products must be easily separated from the input tones. A variety of frequencies can be selected for the input tones; a typical widely used set is described here. The first pair of tones is separated by $6 \text{ Hz} \pm 1 \text{ Hz}$ and centered at $860 \text{ Hz} \pm 1 \text{ Hz}$, and the second pair of tones is separated by $16 \text{ Hz} \pm 1 \text{ Hz}$ and centered at $1380 \text{ Hz} \pm 1 \text{ Hz}$. The four tones are of equal level within ± 0.25 dB. Using Eq. (2), it is easy to show that the output of the device under test will contain six third-order intermodulation products in the range 1877 Hz to 1923 Hz, four second-order intermodulation products in the range 503 Hz to 537 Hz, and four second-order intermodulation products in the range 2223 Hz to 2257 Hz. Thus, the second-order products measured are the combination of two band-pass filters, one having a passband from 503 Hz to 537 Hz, and the other having a passband from 2223 Hz to 2257 Hz. The third-order products measured are the output of a band-pass filter having a passband from 1877 Hz to 1923 Hz.

Noise-Power Ratio (NPR) Test. In the NPR test, the input to the device under test is obtained from a white-noise source which is band-limited to the instantaneous frequency range of interest. This emulates a situation with many simultaneous input signals. Provided that none of the signals dominate, according to the central-limit theorem, the resulting voltage obtained when many uncorrelated signals are added will approach a Gaussian distribution. True white noise covers a frequency range of interest continuously, unlike discrete signals.

The NPR test measures the amount of intermodulation product power between two frequency ranges of white Gaussian noise. A white-noise generator is used with its output frequency range limited by a band-pass filter according to the bandwidth of the device under test. A quiet channel, of bandwidth $\delta\omega$, is formed by a switchable bandreject filter, as shown in Fig. 3. Then, the resulting white-noise signal of Fig. 3 is applied to the input of the device under test. At the output of the system under test is a receiver which is switch-tuned to the frequency of the bandreject filter used to produce the quiet channel.

The NPR test is widely used for evaluating the intermodulation performance of systems whose input-signal spectrum can be approximated by that of white noise. However, the NPR may

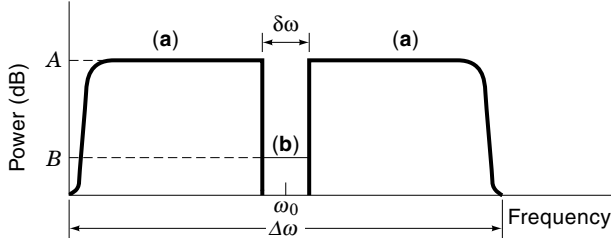


Figure 3. The output spectrum of a noise-power ratio measurement. (a) Injected noise. (b) Noise and intermodulation generated in the measurement bandwidth, $\delta\omega$, by the DUT. $\text{NPR} = A - B$.

be degraded by the noise floor of the system under test, especially under very low loading. It may also be degraded by the distortion products which are produced under high loading (38).

Cross-Modulation. Cross-modulation occurs when modulation from a single unwanted modulated signal transfers itself across and modulates the wanted signal. Cross-modulation is troublesome primarily if the desired signal is weak and is adjacent to a strong unwanted signal. Even when the carrier of the strong unwanted signal is not passed through the system, the modulation on the undesired carrier will be transferred to the desired carrier. Cross-modulation is therefore a special case of intermodulation. Recall that when the input to a nonlinear system is a two-tone signal of the form of Eq. (3), then the amplitudes of the output components at frequencies ω_1 and ω_2 will be given by

$$b_1 = k_1 V_1 + \frac{3}{4} k_3 V_1^3 + \frac{3}{2} k_3 V_1 V_2^2 \quad (16)$$

and

$$c_1 = k_1 V_2 + \frac{3}{4} k_3 V_2^3 + \frac{3}{2} k_3 V_1^2 V_2 \quad (17)$$

respectively. Thus, the output obtained at each frequency, ω_1 and ω_2 , is dependent upon the amplitude of the signal component of the other frequency. If the amplitude of the wanted unmodulated carrier is V_1 and the instantaneous amplitude of the unwanted amplitude-modulated carrier is

$$V_2(t) = V_2(1 + m \cos \omega_m t) \quad (18)$$

then, using Eq. (16), the amplitude of the wanted carrier will be

$$b_1 = k_1 V_1 + \frac{3}{4} k_3 V_1^3 + \frac{3}{2} k_3 V_1 V_2^2 (1 + m \cos \omega_m t)^2 \quad (19)$$

For small values of m and with $k_3 \ll k_1$, Eq. (19) can be approximated by

$$b_1 \approx k_1 V_1 + 3k_3 V_1 V_2^2 m \cos \omega_m t \quad (20)$$

Thus the wanted carrier will be modulated by a modulation index

$$p = 3 \frac{k_3}{k_1} V_2^2 m \quad (21)$$

The cross-modulation factor is then defined as

$$K = \frac{p}{m} \quad (22)$$

Thus, one frequency will be modulated by the modulation of the other frequency. Similar results can be obtained if the unwanted carrier is FM-modulated.

Cross-modulation can be measured as the change in the amplitude of the wanted unmodulated carrier as a function of the amplitude of the unwanted unmodulated carrier. This is the procedure used by the NCTA (National Cable Television Association) standard cross-modulation measurement (39).

Alternatively, the cross-modulation can be measured using the definition of Eq. (22)—that is, the percentage modulation that appears on an unmodulated desired carrier due to the presence of an undesired modulated carrier, divided by the percentage modulation on the undesired carrier (40).

Cross-modulation can also be measured using two equal-amplitude carriers. The wanted carrier, ω_2 , is unmodulated while the unwanted carrier, ω_1 , is FM-modulated. The output spectrum clearly shows the frequency deviation of the wanted carrier. Moreover, it can be shown that the frequency deviation of the intermodulation components of the output spectrum is larger than that of the original FM-modulated unwanted carrier. For the intermodulation product of frequency $\alpha\omega_1 \pm \beta\omega_2$, the deviation will be multiplied by α . Thus, it may be easier to measure the cross-modulation by measuring the deviation of an intermodulation product rather than the deviation of the wanted unmodulated carrier (41).

Differential Gain. The differential gain (DG), a parameter of special interest in color-TV engineering, is conventionally defined as the difference in gain encountered by a low-level high-frequency sinusoid at two stated instantaneous amplitudes of a superimposed slowly varying sweep signal. In video signal transmission, the high-frequency sinusoid represents the chromatic signal and the low-frequency sinusoid represents the luminance signal. Corresponding to the theoretical conditions of the differential measurement, DG measurement is performed with a signal of the form of Eq. (3) with $\omega_2 \gg \omega_1$ and $V_2 \rightarrow 0$ at $V_1 = 0$, X (42). Therefore—recalling that if the input to a nonlinear system is a two-tone signal of the form of Eq. (3) then the amplitude of the output component at frequency ω_2 will be given by

$$c_1 = k_1 V_2 + \frac{3}{4} k_3 V_2^3 + \frac{3}{2} k_3 V_1^2 V_2 \quad (23)$$

Thus, the DG can be expressed as

$$\text{DG} = 1 - \frac{k_1 + \frac{3}{4} k_3 V_2^2}{k_1 + \frac{3}{4} k_3 V_2^2 + \frac{3}{2} k_3 X^2} \quad (24)$$

DG can therefore be considered, to some extent, as a measure of the intermodulation performance of a system under test.

Dynamic Range. The dynamic range can be defined as the amplitude range over which a circuit or a system can operate without performance degradation. The minimum amplitude is dictated by the input thermal noise and the noise contributed by the system. The maximum amplitude is dictated by the

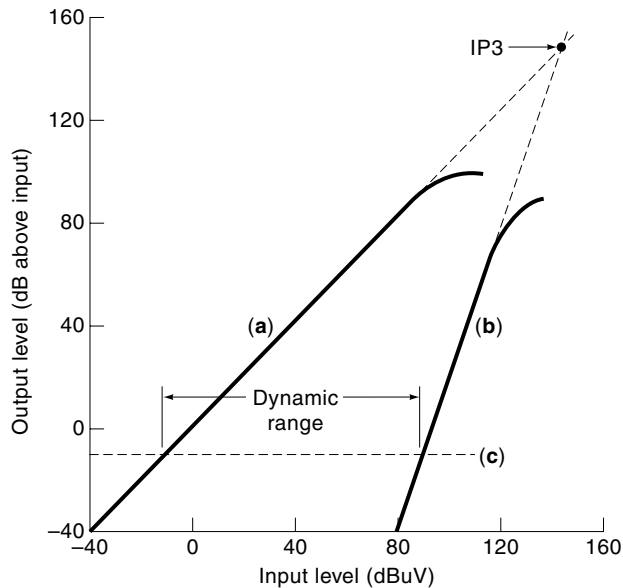


Figure 4. The dynamic range is the difference between the fundamental response input level and the third-order response input as measured along the noise floor. (a) Fundamental response. (b) Third-order intermodulation response. (c) Noise floor.

distortion mechanisms of the system under consideration. In general, the amount of tolerable distortion will depend on the type of the signals and the system under test. However, for the purpose of an objective definition the maximum amplitude will be considered the input signal level at which the intermodulation distortion is equal to the minimum amplitude (43). The dynamic range can therefore be considered, to some extent, as a measure of the intermodulation performance of a system under test.

A useful working definition of the dynamic range is that it is two-thirds the difference in level between the noise floor and the intercept point in a 3 kHz bandwidth (44). Or it is the difference between the fundamental response input level and the third-order response input as measured along the noise floor (sometimes defined as the 3 dB bandwidth above the noise floor) in a 3 kHz bandwidth, as shown in Fig. 4. Reducing the bandwidth improves the dynamic range because of the effect on noise.

Alternatively, in receiver circuits the intermodulation-free dynamic range (IFDR) is widely used to quantify the ability of the receiver to listen to a weak station without disturbance from an intermodulation product generated by strong stations on other frequencies. The IFDR is in fact a measure of how strong two signals can be before the level of their intermodulation products reaches the noise floor of the receiver. The IFDR is defined as the difference between the signal strengths of the strong signals and the noise floor of the receiver. The larger the IFDR, the better.

Adjacent-Channel and Cochannel Power Ratio Tests. In modern telecommunication circuits signals comprising one or more modulated carriers are handled. The characterization of the intermodulation performance of such circuits cannot, therefore, be performed using two-tone and three-tone input signals. A combination of equally spaced tones—in practice ten or more

sinusoids (45), with constant power and correlated or uncorrelated phases—is more appropriate (46).

Because of the nonlinearity of the device under test, intermodulation products will be generated. These intermodulation products can be classified as adjacent-channel distortion when their frequencies are located to the right or to the left of the original spectrum, or cochannel distortion when their frequencies are located exactly over the original spectrum. The adjacent-channel power ratio (ACPR) is defined as the ratio of the total linear output power to the total output power collected in the upper and lower adjacent channels (47). The cochannel power ratio (CCPR) is defined as the ratio of the total linear output power to the total distortion power collected in the input bandwidth (47). The intermodulation distortion ratio (IMR) is the ratio of the linear output power per tone to the output power of adjacent-channel tones (47).

In fact the ACPR, CCPR, and IMR distortion measurements are simple extensions of the two-tone intermodulation measurement (48). However, it is important to first generate a very clean multitone signal. This can be easily achieved using the technique described in Ref. (49).

Dynamic Techniques

Sine-Square Test. This test employs a square wave as a type of signal characterized by rapid change of state (50). In fact, this is a two-tone test where the signal consists of a 3.18 kHz square wave, which has been filtered with a simple one-pole, low-pass RC filter, at either 30 kHz or 100 kHz, and combined with a 15 kHz sine wave. The peak-to-peak amplitude ratio of the sine wave to the square wave is 1 : 4. The resulting square-wave signal component has a very high slope, which is in theory limited only by the low-pass filter. This test has the capability of stressing the amplifier to a high degree of nonlinearities related to signal slope and/or slew rate.

The output spectrum of the system under test is analyzed for the intermodulation products generated by nonlinear mixing of the sine and square waves. The rms sum of the intermodulation products relative to the amplitude of the 15 kHz sine wave is defined as the percentage distortion.

Because the test signal's rate of change depends heavily on out-of-band (>20 kHz) harmonics of the square wave, this test can give somewhat optimistic results for audio systems incorporating front-end low-pass filters (51). Moreover, each component in the output spectrum has two contributory parts: (1) the dynamic intermodulation component caused by the rise-time portion of the square wave driving the amplifier to frequency-dependent nonlinearity—that is, TIM—and (2) the static intermodulation component caused by the amplitude-dependent nonlinearity of the amplifier.

In order to separate the static intermodulation component, the sine-triangle test was proposed (50). The sine-triangle test is similar to the sine-square test with the square wave replaced by a triangular wave of equal peak-to-peak amplitude. This reduces drastically the rise time, leaving only the intermodulation components caused by the static nonlinearities. However, neither the sine-square nor the sine-triangle test uniquely separates the static and dynamic nonlinearities.

The high-pass-square-wave/sine (HP-SQ/S) test is a distinct modification of the sine-square and sine-triangle tests (52). The HP-SQ/S test is based on the sine-square test with the test signal further shaped by a single-pole RC high-pass filter

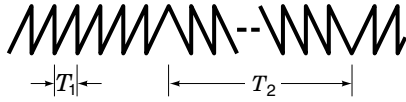


Figure 5. The input signal of a sawtooth wave test is derived from inverting the phase of a 30 kHz sawtooth waveform with a 30 kHz/256 period. $T_1 = 1/30$ kHz, $T_2 = 256T_1$.

with a 3 dB rolloff frequency of 5 kHz. Thus sine-square test signal generators can be applied, requiring only an additional *RC* high-pass filter. The HP-SQ/S test signal better resembles real-life signals and acquires both static and dynamic nonlinear distortions simultaneously. As in the sine-square test, the distortion factor is defined as the ratio of the rms sum of the amplitudes of the intermodulation products to the amplitude of the sinusoid.

Sawtooth-Wave Test. In this method a sawtooth wave is used as the signal that changes its state rapidly (53). The signal is derived from inverting the phase of a 30 kHz sawtooth waveform with a 30 kHz/256 period as shown in Fig. 5. The signal therefore consists of two alternating series of signals: one series of instantaneously rising, the other of instantaneously falling signals.

Without reversal, application of the high-frequency asymmetrical sawtooth signal to a system under test causes the system's symmetrical and asymmetrical nonlinearities to generate a dc offset whose magnitude depends on the severity of the nonlinearity. The periodic polarity reversal merely "chops" this dc offset into an easily measured low-frequency ac signal. Thus, if the system under test is prone to TIM, then at the output of the low-pass filter a signal appears whose shape is rectangular. Each time the input signal reverses polarity, a rectangular output waveform appears that is due to the shift in average voltage or dc level. This output signal represents the amount of TIM in the system under test. TIM is calculated as

$$\text{TIM} = \left(\frac{\text{amplitude (peak-to-peak) of sawtooth}}{\text{amplitude (peak-to-peak) of square wave}} \right) 100\% \quad (25)$$

Because this test depends so heavily on extreme signal slope (and hence on out-of-band sawtooth harmonics), it may easily be misled in its assessment of TIM by systems incorporating input low-pass filters (51).

Multitone Intermodulation (MIM) Test. The MIM test is a variation of the CCIF intermodulation test in which two high-frequency tones spaced apart by a small frequency difference are applied to the system under test (51). In order to retain the advantages of the CCIF test while incorporating sensitivity to symmetrical distortion mechanisms, a third tone, at frequency ω_3 , is added. The three frequencies are chosen so that the resulting triple-beat product at $\omega_3 - \omega_2 - \omega_1$ is slightly below 1 kHz, while two of the tones produce a CCIF-like difference-frequency product $\omega_2 - \omega_1$ at slightly above 1 kHz. Specifically, the three equal-amplitude tones are at frequencies 20 kHz, 10.05 kHz, and 9 kHz, resulting in a triple-beat product at 950 Hz and a difference-frequency product at 1050 Hz.

The difference-frequency and triple-beat products are selected to lie so close to each other that both products can be passed through a narrow band-pass filter centered about 1 kHz. The distortion percentage is defined as the value of the 950 Hz and 1050 Hz distortion products, measured together on an rms-calibrated average-responding ac voltmeter, referred to the rms value of the sine wave of the same peak-to-peak amplitude as the three-tone MIM test signal.

The MIM test enjoys the following attractive features:

1. Inexpensive instrumentation; no spectrum analyzers are required.
2. Simple measurement procedure.
3. Fully in-band stimulus and response.

The MIM test is not as stringent as the sine-square and sawtooth wave tests in terms of peak rate of change, and as a result it yields smaller TIM distortion percentages. However, because it does not resort to unrealistically high rates of change to stress the audio system under test, good subjective correlation can be expected.

INTERMODULATION MEASUREMENT

Measurement Equipment

Multitone Tests. A block diagram of the system used for multitone intermodulation measurement is shown in Fig. 6. The multiple frequency source can be implemented from two or three synthesized sine/square/triangular wave generators. Amplifier/attenuator pairs can be added at the output of each generator. Bandpass filters can also be added to suppress the harmonic contents at the output of each generator. For RF

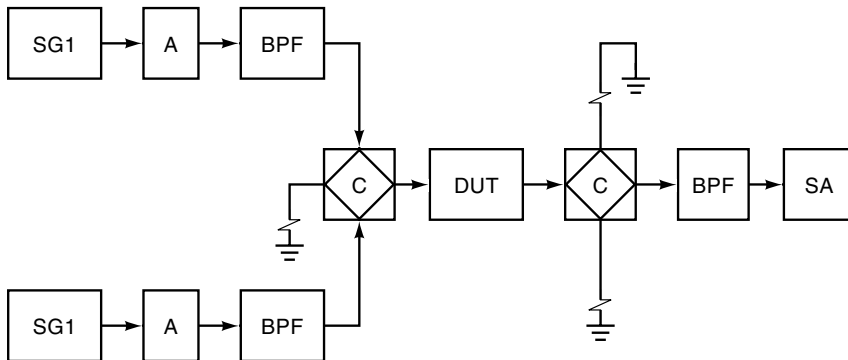


Figure 6. Block diagram of the two-tone test setup. Multitone tests require additional signal generators, combiners, amplifiers and bandpass filters. SG, signal generator; A, amplifier; BPF, bandpass filter; C, combiner; DUT, device under test; SA, spectrum analyzer.

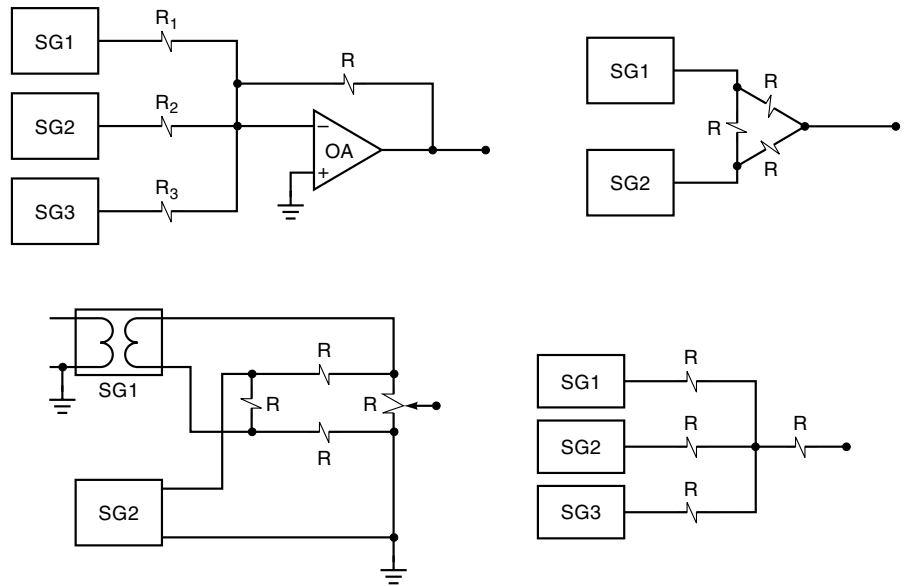


Figure 7. Different types of resistive combiners used in audio-frequency tests. SG, signal generator; OA, operational amplifier.

measurements, harmonic suppression and isolation between different generators is achieved by using amplifier/circulator combinations and cavity resonators (49). The synthesized sources are combined using hybrids or combiners of adequate isolation. Spectral purity at this point is crucial to the accuracy of the measurement. The multitone output is fed to the device under test (DUT). The output of the DUT is fed to the spectrum analyzer. For audio-frequency measurements, the intermodulation components of interest can be filtered out, using band-pass filters, and fed to ac voltmeters. Alternatively, for RF measurements, the output of the DUT can be fed to directional couplers. The outputs of the directional couplers are fed to a television oscilloscope and/or a spectrum analyzer.

For audio-frequency measurements, resistive combiners are widely used for combining the outputs of two, or more, signal generators. Figure 7 shows a number of the widely used resistive combining networks.

Measurement Using a Microcomputer. Intermodulation can also be measured using a microcomputer (50). The block diagram of this technique is shown in Fig. 8. This technique is based on measuring the single tone input-output characteristic of the DUT using a vector voltmeter. The output of the vector voltmeter is fed to a microcomputer, which converts it into three digital data lines representing the input amplitude, the output amplitude, and the phase lag between the input and output signals. After storing the data, the microcomputer increments the amplitude of the input signal. After storing all the necessary data, the microcomputer, using a stochastic method, calculates the amplitudes of the intermodulation components of the DUT. Although the procedure reported in Ref. 55 uses a stochastic method for calculating the amplitudes of the intermodulation components resulting from a two-tone input signal, the same procedure can be applied to any number of input tones, using different analytical techniques for modeling the nonlinear characteristics of the DUT.

Alternatively, microcomputers can be added to the measurement setup of Fig. 6 to:

1. Control the frequencies of the signal sources, especially in the millimeter-wavelength range, where the difference in

frequencies between the signal sources may be less than 0.001 times the base signal frequency (56).

2. Scan the base signal frequency over the measurement range of interest in predefined steps (57).
3. Correct the power from each source so that power delivery to the DUT will be the same across the whole frequency range scanned.
4. Read and calculate the parameters of interest during the measurements (58,59).

Noise-Power Ratio Test. Figure 9 shows a block diagram of a noise-power ratio test setup (38). The setup consists of a white-noise generator which applies an accurate level of white Gaussian noise power with known bandwidth (equal to $\Delta\omega$ and centered around ω_0) to the DUT. The output of the DUT is measured with the bandreject filter out. When the bandreject filter, with bandwidth $\delta\omega$ and centered around ω_0 , is switched in, a narrow band of frequencies is attenuated by about 70 dB, and

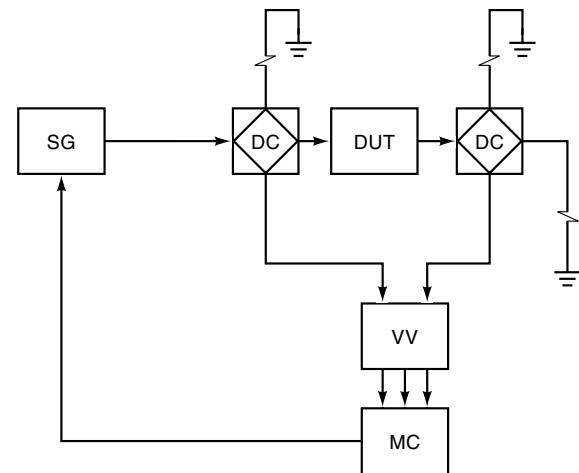


Figure 8. Block diagram of a microcomputer-based intermodulation measurement setup. SG, signal generator; DC, directional coupler; DUT, device under test; VV, vector voltmeter; MC, microcomputer.

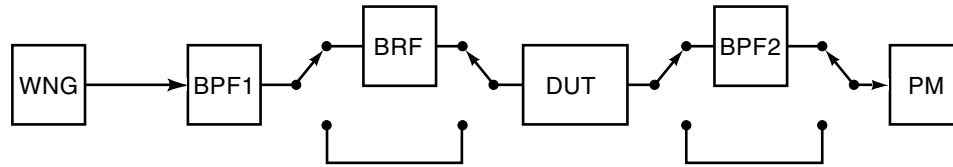


Figure 9. Block diagram of the noise-power ratio test setup. WNG, white-noise generator; BPF1, band-pass filter with bandwidth $\Delta\omega$ centered around ω_0 ; BRF, bandreject filter with bandwidth $\delta\omega$ centered around ω_0 ; DUT, device under test; BPF2, band-pass filter with bandwidth $\delta\omega$ centered around ω_0 ; PM, power meter.

a quiet channel, of width $\delta\omega$ and centered around ω_0 , is formed as shown in Fig. 3. At the output of the DUT, the noise power is measured in the quiet channel, using a band-pass filter with bandwidth $\delta\omega$ and centered around ω_0 . This noise power is due to the thermal noise and the intermodulation introduced by the device under test. The NPR is the ratio of the noise power measured without the bandreject filter inserted before the DUT to that measured with the bandreject filter inserted. The white-noise generator corrects the loading power level for the insertion loss of the bandreject filter.

Externally Induced Intermodulation Test. This is a two-tone test with one signal applied to the input and the other signal applied to the output (9). A test setup is shown in Fig. 10. Two directional couplers are used to gauge both the forward-carrier power and the intermodulation product levels. Two more directional couplers are added to inject the interfering signal and to measure the actual injected value using the spectrum analyzer.

Measurement Accuracy

Multitone Tests. For accurate measurements of the intermodulation products, using multitone tests, it is essential to reduce or remove the nonlinear distortion originating in the signal sources and/or the measurement equipment. Measurement accuracy may therefore be affected by the purity of the signal sources, the linearity of the combiners, and the performance of the spectrum analyzer.

Signal Sources. Measurement of the amplitudes of the intermodulation components requires the use of two or more signals. The frequencies of these signals must be noncommensurate. Otherwise, harmonics in one source might beat with the fundamental(s) of other signal(s) and interfere with the desired intermodulation components.

Ideally the signal generators would produce perfect sinusoids, but in reality all signals have imperfections. Of particular interest here is the spectral purity, which is a measure of the inherent frequency stability of the signal. Perhaps the most common method to quantify the spectral purity of a signal generator is its phase noise (60). In the time domain, the phase noise manifests itself as a jitter in the zero crossings of a sine wave. In the frequency domain, the phase noise appears as sidebands surrounding the original frequency. Thus, mixing with other frequencies due to the nonlinearities of the DUT would result in additional intermodulation products. It is therefore important to consider the intermodulation due to phase noise when calculating the intermodulation performance of the DUT (61).

Signal generators with automatic level control (ALC) may produce signals with unwanted modulation. The ALC is implemented by rectifying the output signal of the generator and feeding back the resulting dc voltage to drive an amplitude modulator. If a second signal is applied to the output of the signal generator, the detector will produce a signal at the difference in frequency between the two frequencies. This signal will modulate the generator's output. The frequency of the modulation

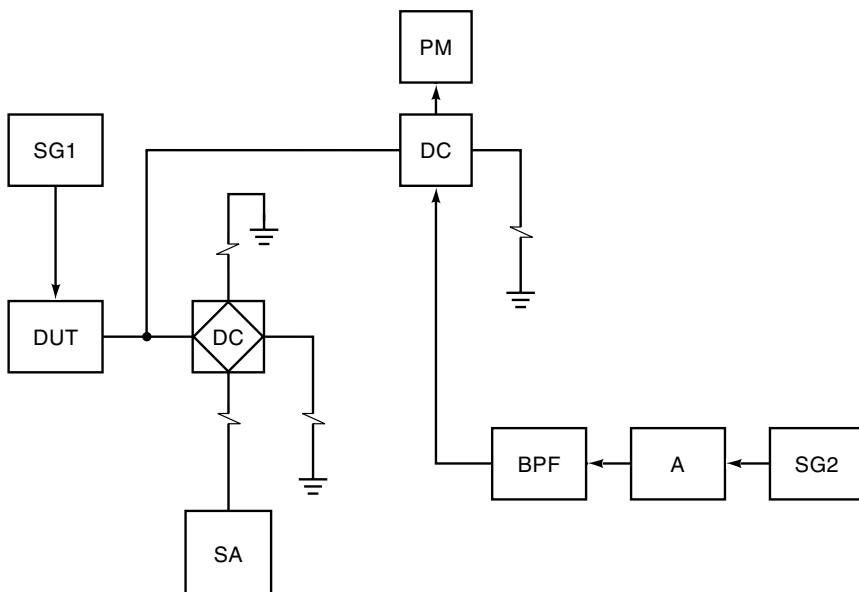


Figure 10. Measurement of externally induced intermodulation can be performed by using two tones: one injected at the input and one injected at the output of the DUT. SG, signal generator; DC, directional coupler; PM, power meter; SA, spectrum analyzer; BPF, band-pass filter; A, amplifier.

sidebands will share the same spectral lines as the intermodulation products of interest. Isolating the signal generators and the combiners can minimize this effect. This can be achieved by ensuring that there is as much attenuation as possible between them.

Combiners. Measurement of intermodulation products is performed by applying to the input of the circuit, or the system, under test a signal consisting of two or more different frequencies. The outputs of the signal generators are therefore combined by a combiner. The combiner must provide sufficient isolation between the signal sources to reduce the possibility of producing intermodulation products before the combined input signal is applied to the circuit or the system under test. While resistive combiners are adequate for input signal levels up to a few millivolts, for larger voltage levels the use of power combiners may be necessary (62). Insertion of an attenuator in each arm of the combiner helps in minimizing the distortion components resulting from the interaction between the two signal sources. Such components, if generated, should be at least 80 dB below the fundamental components.

A simple test to determine whether adequate isolation has been achieved can be effected by introducing a variable attenuator between the signal source combiner and the DUT in Fig. 6. This is set to a low value during measurements, but at setup, when IMPs have been located on the spectrum analyzer, increasing the attenuation by 3 dB will result in a reduction in the observed IMP level. If this reduction is only 3 dB, then it has to be assumed that the IMP observed has originated in the signal sources, not in the DUT. If, however, the reduction is 6 dB for a second-order IMP or 9 dB for a third-order [see Eq. (4)], then it is safe to assume that the IMP has originated in the DUT or the spectrum analyzer.

Alternatively, a technique which attenuates the parasitic intermodulation products arising from the interaction between the generators of the fundamental components, before the input of the spectrum analyzer, was described in Ref. 63.

A block diagram of the technique is shown in Fig. 11. The input to the system under test is formed by combining the outputs of two signal generators at frequencies ω_1 and ω_2 in the combiner. The hybrid splitter-combiner (HSC1) splits the combined signal into two branches with voltage-transfer ratio $a = \alpha$ and $b = \sqrt{1 - \alpha^2}$ at the first and second outputs. Using Eq. (1) and assuming that the system under test and the compensator have identical nonlinear characteristics, the inputs of the second hybrid splitter-combiner (HSC2) can be expressed as

$$V_a = \sum_{n=0}^N k_n (\alpha V_i)^n \quad (26)$$

and

$$V_b = \sum_{n=0}^N k_n (\sqrt{1 - \alpha^2} V_i^n) \quad (27)$$

Using Eqs. (26) and (27), the output of HSC2, with voltage transfer ratio opposite in sign and equal to the reciprocal of that of HSC1, can be expressed as

$$V_{out} = \sum_{n=0}^K -k_n [\sqrt{1 - \alpha^2} (\alpha V_i)^n - \alpha (\sqrt{1 - \alpha^2} V_i^n)] \quad (28)$$

According to Eq. (28), broadband compensation occurs for the linear components of the combined signal, with $n = 1$. Thus, all the linearly transformed spectral components are eliminated. This is also true for the intermodulation components which may result from the nonlinear interaction between the two signal generators. The output of HSC2 can therefore be applied directly to the spectrum analyzer.

This technique does not require complicated high-order selective filters and can attenuate the parasitic intermodulation components and the fundamental frequency components by about 50 dB over a wide range of frequencies differing by 7 to 10 octaves. However, it requires a compensator with a nonlinear characteristic similar to that of the system under test.

Spectrum Analyzers. Spectrum analyzers are widely used in measuring the intermodulation performance of electronic circuits and systems. Internal circuits of the spectrum analyzers are themselves imperfect and will also produce distortion products (22). The distortion performance of the analyzers is usually specified by the manufacturers, either directly or lumped into a dynamic-range specification. The performance of the analyzer can be stretched, however, if the nature of these distortion products is understood.

Amplitudes of the distortion products, resulting from the internal circuits of the analyzer, can be reduced by reducing the signal levels at the analyzer's input. Thus, using internal and/or external attenuators can reduce the input signal levels to the analyzer and hence reduce its distortion products and improve the intermodulation measurement range of the spectrum analyzer. However, reduced input levels to the analyzer mean reduced signal-to-noise ratio, and the distortion component to be measured may be buried in the noise. White noise can be reduced by reducing the resolution bandwidth of the analyzer, this may lead to a lower sweep rate. Thus, achieving an optimum dynamic range involves tradeoffs between input signal levels and analyzer distortion. Usually, data sheets of analyzers will contain information about the noise level in each

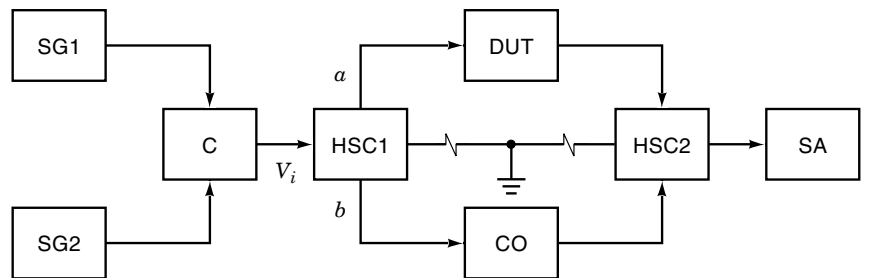


Figure 11. A technique for attenuating the intermodulation products arising due to the interaction between the signal generators of the fundamental components. SG, signal generator; C, combiner; HSC, hybrid splitter-combiner; DUT, device under test; CO, compensator; SA, spectrum analyzer.

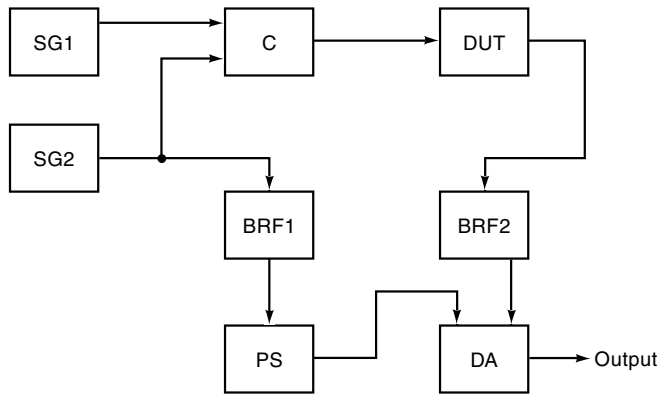


Figure 12. Compensation method for the measurement of small-amplitude intermodulation products in the presence of a large fundamental. SG, signal generator; C, combiner; DUT, device under test; BRF, bandreject filter; PS, phase shifter; DA, differential amplifier.

resolution bandwidth and about distortion products generated by the analyzer for each input level. Using this information, the dynamic range of the analyzer for various input levels can be determined (64).

Whenever good selectivity, as well as sensitivity and dynamic range, is of prime importance, test receivers may be used in preference to spectrum analyzers (6).

Alternatively, if the frequencies of the intermodulation components of interest are sufficiently lower (or higher) than the fundamental frequencies, then low-pass (or high-pass) filters can be used to remove the fundamental components which would give rise to other nonlinear distortion components in the spectrum analyzer. Attenuation factors of 80 dB or more, at frequencies outside the band of interest, are recommended. The insertion loss of the low-pass (or the high-pass) filter should be as small as possible; 0.4 dB or less is recommended.

If the frequency of the intermodulation component of interest is not sufficiently higher (or lower) than the fundamental frequencies, then it will be necessary to have complicated multiple-section high-order filters with amplitude-frequency characteristics that are nearly rectangular. Such filters will change, to some extent, the amplitude of the intermodulation components, and this will complicate the calculation of the intermodulation performance of the system under test. A method for compensating for a large fundamental component, thus allowing the measurement of small intermodulation components in its presence, was described in Ref. 65.

A block diagram of the compensation method is shown in Fig. 12. The input to the system under test is formed from one large-amplitude signal at frequency ω_1 and one small-amplitude signal at frequency ω_2 with $\omega_1 \ll \omega_2$. The output of the system under test contains fundamental components at frequencies ω_1 and ω_2 , and intermodulation components at frequencies $\omega_2 \pm n\omega_1$, $n = 1, 2, \dots, N$. In order to measure the small-amplitude intermodulation components, it is necessary to avoid applying to the analyzer the fundamental component at frequency ω_2 . This can be achieved as follows. The output of the system under test is fed to the bandreject filter BRF2 to suppress the fundamental component at ω_1 . The output of the signal generator of frequency ω_2 is fed to the bandreject filter BRF1 to suppress the possible entry through the combiner of any component at frequency ω_1

before reaching the phase shifter. The phase shifter compensates, at the frequency ω_2 , the phase shift through the system under test.

Ideally, the voltages, of frequency ω_2 , at the inputs of the differential amplifier are in phase with the voltage at the non-inverting input, equal to half the voltage at the inverting input. Thus, the output of the differential amplifier at frequency ω_2 is ideally zero. Practically the output voltage at ω_2 will be attenuated by 50 dB to 60 dB (6). The output of the differential amplifier, with the suppressed fundamental component at frequency ω_2 , can be applied to the spectrum analyzer.

This compensation technique, which entails additional filters and matching units, can be used only for broadband measurements with $\omega_1 \ll \omega_2$.

While spectrum analyzers using digital IF sections may not suffer from the internally generated distortion discussed in the preceding, they may suffer from the relatively low-level distortion products resulting from the analog-to-digital conversion. The amplitudes of these products are usually less sensitive to the amplitude of the signal components.

Noise-Power Ratio Test. The accuracy of the noise-power ratio (NPR) test is affected mainly by two factors: (1) the noise floor of the amplifier, which will dominate under very low loading, and (2) the distortion products, produced under very high loading. It is therefore recommended to sweep the loading between two prespecified start and stop levels. The NPR is measured at different levels, and the largest measured value of NPR is considered as the worst case.

Microcomputer-Based Tests. Quantization errors associated with the analog-to-digital conversion of the data in microcomputer-based intermodulation tests must be taken into account. Measurement errors due to quantization are affected by the length of the binary digits and determine the dynamic range of operation (53).

BIBLIOGRAPHY

1. L. E. Kinsler et al., *Fundamentals of Acoustics*, New York: Wiley, 1982, pp. 267–268.
2. K. Y. Eng and O.-C. Yue, High-order intermodulation effects in digital satellite channels, *IEEE Trans. Aerosp. Electron. Syst.*, **AES-17**: 438–445, 1981.
3. C. D. Bod, C. S. Guenzer, and C. A. Carosella, Intermodulation generation by electron tunneling through aluminum-oxide films, *Proc. IEEE*, **67**: 1643–1652, 1979.
4. W. H. Higa, Spurious signals generated by electron tunneling on large reflector antennas, *Proc. IEEE*, **63**: 306–313, 1975.
5. P. L. Aspden and A. P. Anderson, Identification of passive intermodulation product generation in microwave reflecting surfaces, *IEEE Proc.-H*, **139**: 337–342, 1992.
6. P. L. Liu, A. D. Rawlins, and D. W. Watts, Measurement of intermodulation products generated by structural components, *Electron. Lett.*, **24**: 1005–1007, 1988.
7. M. Ojala and J. Lammasniemi, Intermodulation at the amplifier-loudspeaker interface, *Wireless World*, **86**: 45–47, November 1980; 42–44, 55, December 1980.
8. E. M. Cherry and G. K. Cambrell, Output resistance and intermodulation distortion in feedback amplifiers, *J. Audio Eng. Soc.*, **30**: 178–191, 1982.

9. E. Franke, Test setup gauges externally-induced transmitter IM, *Microw. RF*, **32**: 95–98, 1993.
10. W. Wharton, S. Metcalfe, and G. C. Platts, *Broadcast Transmission Engineering Practice*, Oxford, UK: Butterworth-Heinemann, 1991, Chapter 5.
11. J. M. Lindsey, L. S. Riggs, and T. H. Shumpert, Intermodulation effects induced on parallel wires by transient excitation, *IEEE Trans. Electromagn. Compat.*, **31**: 218–222, 1989.
12. M. Ojala, Non-linear distortion in audio amplifiers, *Wireless World*, **83**: 41–43, January 1977.
13. E. M. Cherry, Intermodulation distortion in audio amplifiers, *IEEE Conf. Int.*, Australia, 1983, pp. 639–641.
14. W. G. Jung, M. L. Stephens, and C. C. Todd, An overview of SID and TIM—Part I, *Audio*, **63**: 59–72, 1979.
15. R. R. Cordell, Another view of TIM, *Audio*, **64**: 38–49, 1980.
16. W. M. Leach, Transient IM distortion in power amplifiers, *Audio*, **59**: 34–41, 1975.
17. W. M. Leach, Suppression of slew rate and transient IM distortions in audio power amplifiers, *J. Audio Eng. Soc.*, **25**: 466–473, 1977.
18. S. Naryanan, Transistor distortion analysis using the Volterra series representation, *Bell Syst. Tech. J.*, **46**: 999–1024, 1967.
19. P. Harrop and T. A. C. M. Claassen, Modelling of an FET mixer, *Electron. Lett.*, **14**: 369–370, 1978.
20. W. G. Jung, M. L. Stephens, and C. C. Todd, An overview of SID and TIM—Part III, *Audio*, **63**: 42–59, 1979.
21. M. T. Abuelma'atti, Prediction of the transient intermodulation performance of operational amplifiers, *Int. J. Electron.*, **55**: 591–602, 1983.
22. R. A. Witte, *Spectrum and Network Measurements*, Englewood Cliffs, NJ: Prentice-Hall, 1991, Chapter 7.
23. S. Hunziker and W. Baechtold, Simple model for fundamental intermodulation analysis of RF amplifiers and links, *Electron. Lett.*, **32**: 1826–1827, 1996.
24. G. A. A. Hueber, B. Nijholt, and H. Tendeloo, Twin-tone tape testing, *J. Audio Eng. Soc.*, **24**: 542–553, 1976.
25. J. Li, R. G. Bosisio, and K. Wu, A simple dual-tone calibration of diode detectors, *IEEE Instrum. Meas. Technol. Conf.*, Hamamatsu, Japan, 1994, 1994, pp. 276–279.
26. J. D. Giacomini, Most ADC systems require intermodulation testing, *Electron. Des.*, **40** (17): 57–65, 1992.
27. M. Benkais, S. L. Masson, and P. Marchegay, A/D converter characterization by spectral analysis in “dual-tone” mode, *IEEE Trans. Instrum. Meas.*, **44**: 940–944, 1995.
28. B. D. Loughlin, Nonlinear amplitude relations and gamma correction, in K. McIlwain and C. Dean (eds.), *Principles of Color Television*, New York: Wiley, 1956, pp. 200–256.
29. M. Kanno and I. Minowa, Application of nonlinearity measuring method using two frequencies to electrical components, *IEEE Trans. Instrum. Meas.*, **IM-34**: 590–593, 1985.
30. L. Robles, M. A. Ruggero, and N. C. Rich, Two-tone distortion in the basilar membrane of the cochlea, *Nature*, **349**: 413–414, 1991.
31. T. Maseng, On the characterization of a bandpass nonlinearity by two-tone measurements, *IEEE Trans. Commun.*, **COM-26**: 746–754, 1978.
32. H. Roering, The twin-tone distortion meter: A new approach, *J. Audio Eng. Soc.*, **31**: 332–339, 1983.
33. E. M. Cherry, Amplitude and phase intermodulation distortion, *J. Audio Eng. Soc.*, **31**: 298–303, 1983.
34. H. H. Scott, Audible audio distortion, *Electronics*, **18**: 126–131, 1945.
35. A. N. Thiele, Measurement of nonlinear distortion in a band-limited system, *J. Audio Eng. Soc.*, **31**: 443–445, 1983.
36. G. L. Heiter, Characterization of nonlinearities in microwave devices and systems, *IEEE Trans. Microw. Theory Tech.*, **MTT-21**: 797–805, 1973.
37. A. D. Broadhurst, P. F. Bouwer, and A. L. Curle, Measuring television transposer intermodulation distortion, *IEEE Trans. Broadcast.*, **34**: 344–355, 1988.
38. B. Hessen-Schmidt, Test set speeds NPR measurements, *Microw. RF*, **33**: 126–128, 1994.
39. B. Arnold, Third order intermodulation products in a CATV system, *IEEE Trans. Cable Telev.*, **CATV-2**: 67–79, 1977.
40. O. A. Dogha and M. B. Das, Cross-modulation and intermodulation performance of MOS-FET's in tuned high-frequency amplifiers, *Int. J. Electron.*, **45**: 307–320, 1978.
41. J. H. Foster and W. E. Kunz, Intermodulation and crossmodulation in travelling-wave tubes, *Conf. Int. Tubes Hyperfrequencies*, Paris, 1964, pp. 75–79.
42. Differential Phase and Gain at Work, Hewlett-Packard Appl. Note 175-1, 1975.
43. J. Smith, *Modern Communication Circuits*, New York: McGraw-Hill, 1987, Chapter 3.
44. J. Dyer, The facts and figures of HF receiver performance, *Electron. World Wireless World*, **99**: 1026–1030, 1993.
45. R. Hajji, F. Beauregard, and F. Ghannouchi, Multitone power and intermodulation load-pull characterization of microwave transistors suitable for linear SSPA's design, *IEEE Trans. Microw. Theory Tech.*, **45**: 1093–1099, 1997.
46. N. B. Carvalho and J. C. Pedro, Multitone intermodulation distortion performance of 3rd order microwave circuits, *IEEE Int. Microwave Theory and Techniques Symp. Digest*, 1999, pp. 763–766.
47. J. C. Pedro and N. B. Carvalho, On the use of multitone techniques for assessing RF components' intermodulation distortion, *IEEE Trans. Microw. Theory Tech.*, **47**: 2393–2402, 1999.
48. N. B. Carvalho and J. C. Pedro, Compact formulas to relate ACPR and NPR to two-tone IMR and IP3, *Microw. J.*, **42**: 70–84, December 1999.
49. R. Hajji, F. Beauregard, and F. Ghannouchi, Multi-tone transistor characterization for intermodulation and distortion analysis, *IEEE Int. Microwave Theory and Techniques Symp. Digest*, 1996, pp. 1691–1694.
50. E. Leinonen, M. Ojala, and J. Curl, Method for measuring transient intermodulation distortion (TIM), *J. Audio Eng. Soc.*, **25**: 170–177, 1977.
51. R. R. Cordell, A fully in-band multitone test for transient intermodulation distortion, *J. Audio Eng. Soc.*, **29**: 578–586, 1981.
52. P. Skritek, A combined measurement method for both dynamic intermodulation and static nonlinear distortions, *J. Audio Eng. Soc.*, **35**: 31–37, 1987.
53. S. Takahashi and S. Tanaka, A new method of measuring transient intermodulation distortion: Comparison with the conventional method, *J. Audio Eng. Soc.*, **30**: 10–16, 1982.
54. G. Hamer, S. Kazeminejad, and D. P. Howson, Test set for the measurement of IMDs at 900 MHz, *IEE Colloq. Passive Intermodulation Prod. Antennas Relat. Struct.*, London, IEE Dig. No. 1989/94, 1989.
55. T. Sasaki and H. Hataoka, Intermodulation measurement using a microcomputer, *IEEE Trans. Instrum. Meas.*, **IM-30**: 262–264, 1981.
56. P. A. Morton et al., Large-signal harmonic and intermodulation distortions in wide-bandwidth GaInAsP semiconductor lasers, *IEEE J. Quantum Electron.*, **25**: 1559–1567, 1989.

57. S. Mukherjee, Vector measurement of nonlinear transfer function, *IEEE Trans. Instrum. Meas.*, **44**: 892–897, 1994.
58. C. Tsironis, Two tone intermodulation measurements using a computer-controlled microwave tuner, *Microw. J.*, **32**: 161–163, 1989.
59. A. A. M. Saleh and M. F. Wazowicz, Efficient, linear amplification of varying-envelope signals using FET's with parabolic transfer characteristics, *IEEE Trans. Microw. Theory. Tech.*, **MTT-33**: 703–710, 1985.
60. B. Cheng, Signal generator spectral purity consideration in RF communications testing, *Microw. J.*, **42**: 22–32, December 1999.
61. S. Ciccarelli, Predict receiver IM in the presence of LO phase noise, *Microw. & RF*, **35**: 86–90, 1996.
62. A. M. Rudkin (ed.), *Electronic Test Equipment*, London: Granada, 1981, Chapter 2.
63. Yu. M. Bruk and V. V. Zakharenko, Broadband compensation for dynamic-range measurements by intermodulation, *Instrum. Exp. Tech.*, **36** (Part 1, No. 4), 557–562, 1993.
64. Spectrum Analyzer Series, Hewlett-Packard Appl. Note 150-11, 1976.
65. V. G. Frenkeland and M. S. Shterengas, Auxiliary unit for a spectrum analyzer when measuring intermodulation distortion, *Meas. Tech.*, **32**: 385–387, 1989.

MUHAMMAD TAHER ABUELMA'ATTI
King Fahd University of
Petroleum and Minerals

MEASUREMENT OF FREQUENCY, PHASE NOISE AND AMPLITUDE NOISE

Frequency metrology has the highest resolution of all the measurement sciences. Simple systems readily achieve a fractional frequency resolution of 1 ppm (part per million) and some elaborate systems achieve 1 part in 10^{17} or less. Because of the readily achieved resolution, the growing trend is to convert the measurement of many different parameters to the measurement of frequency or frequency difference.

In the following we describe the basic ideas and definitions associated with various aspects of the specification and measurement of frequency, phase (or time), and the two components of spectral purity—phase modulation (PM) noise and amplitude modulation (AM) noise. The three topics are fundamentally linked, since as we will see below, frequency is proportional to the rate of change of phase with time, and the degree to which we can specify or measure signal phase, amplitude, or frequency is limited by the spectral purity of the signals.

BASIC CONCEPTS AND DEFINITIONS

Figure 1 shows the output voltage signal of an ideal sinusoid oscillator as a function of time. The maximum value V_0 is the nominal amplitude of the signal. The time required for the signal to repeat itself is the period T of the signal. The nominal frequency ν_0 of the signal is the reciprocal of the period, $1/T$. This voltage signal can be represented mathematically by a sine function,

$$v(t) = V_0 \sin \theta = V_0 \sin(2\pi \nu_0 t) \quad (1)$$

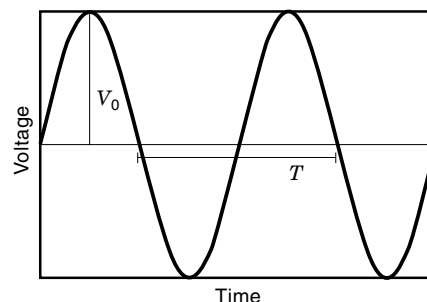


Figure 1. Output voltage of an ideal sinusoidal oscillator.

where the argument $\theta = 2\pi \nu_0 t$ of the sine function is the nominal phase of the signal. The time derivative of the phase θ is $2\pi \nu_0$ and is called the nominal angular frequency ω_0 . In the frequency domain, this ideal signal is represented by a δ function located at the frequency of oscillation.

In real situations, the output signal from an oscillator has noise. Such a noisy signal is illustrated in Fig. 2. In this example we have depicted a case in which the noise power is much less than the signal power. Fluctuations in the peak values of the voltage result in AM noise. Fluctuations in the zero crossings result in PM noise. Fractional frequency modulation (FM) noise refers to fluctuations in the period of the signal. Since the period (and thus the frequency) of the signal is related to the phase of the signal, FM noise and PM noise are directly related.

Figure 3 shows the power spectrum of a noisy signal (power as a function of frequency) as measured by a spectrum analyzer. Although the maximum power occurs at the frequency of oscillation, other peaks are observed at frequencies of $2\nu_0$, $3\nu_0$, $\dots$, $n\nu_0$. These frequencies are called *harmonics* of the fundamental frequency ν_0 ; $2\nu_0$ is the second harmonic, $3\nu_0$ is the third harmonic, and so on. The power at these harmonic frequencies will depend on the design of the source. The spectrum around the fundamental frequency displays power sidebands at frequencies above the carrier (upper sideband) and at frequencies below the carrier (lower sideband). These power sidebands are the result of PM and AM noise in the signal. While the power spectrum gives an idea of the total noise of a signal, it does not give information about the relative magnitude of the PM and AM noise. Furthermore, at frequencies close to ν_0 , it is difficult to separate noise power from the power of the fundamental frequency. Therefore, special measurement techniques are needed to measure PM and AM noise in oscillators.

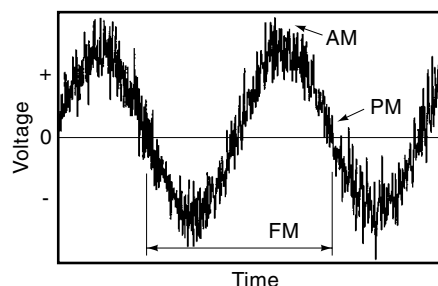


Figure 2. Output voltage of a noisy sinusoidal oscillator.

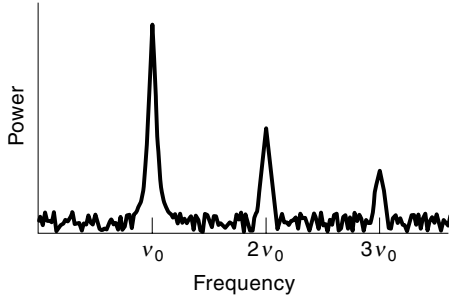


Figure 3. Power spectrum of a noisy signal.

Characterization of Frequency Stability and Amplitude Stability of a Signal

A noisy signal can be mathematically represented by

$$v(t) = [V_0 + \epsilon(t)] \sin[2\pi\nu_0 t + \phi(t)] \quad (2)$$

where $\epsilon(t)$ represents amplitude fluctuations (amplitude deviation from the nominal amplitude V_0) and $\phi(t)$ represents phase fluctuations (phase deviation from the nominal phase $2\pi\nu_0 t$) (1). The instantaneous frequency of this signal is defined as

$$\nu(t) = \frac{1}{2\pi} \frac{d}{dt}(\text{phase}) = \nu_0 + \frac{1}{2\pi} \frac{d}{dt} \phi(t) \quad (3)$$

Frequency fluctuations refer to the deviation of the instantaneous frequency from the nominal frequency: $\nu(t) - \nu_0$. Fractional frequency fluctuations, denoted as $y(t)$, refer to frequency fluctuations normalized to ν_0 , that is,

$$y(t) = \frac{\nu(t) - \nu_0}{\nu_0} = \frac{1}{2\pi\nu_0} \frac{d}{dt} \phi(t) \quad (4)$$

Equation (4) indicates that there is a direct relation between phase fluctuations and fractional frequency fluctuations. Therefore if the PM noise of a signal is measured, the FM noise can be easily obtained and vice versa. The time deviation or fluctuation $x(t)$ of a signal is equal to the integral of $y(t)$ from 0 to t . This relation can be expressed as

$$y(t) = \frac{d}{dt} x(t) \quad (5)$$

Units of Measure for PM Noise, FM Noise, and AM Noise

Phase fluctuations in the frequency domain are characterized by the spectral density of the phase fluctuations $S_\phi(f)$, given by

$$S_\phi(f) = \text{PSD}[\phi(t)] = [\phi(f)]^2 \frac{1}{\text{BW}} \quad (6)$$

where PSD refers to power spectral density, $[\phi(f)]^2$ is the mean-squared phase deviation at an offset frequency f from the frequency ν_0 (called the carrier in this context), and BW is the bandwidth of the measurement system (1–4). The offset frequency f is also called Fourier frequency. The units for $S_\phi(f)$ are rad^2/Hz . Equation (6) is defined for $0 < f < \infty$; nevertheless it includes fluctuations from the upper and lower sidebands and thus is a double sideband unit of measure.

The PM noise unit of measure recommended by the IEEE (1,2,4) is $\ell(f)$, defined as

$$\mathcal{L}(f) = \frac{S_\phi(f)}{2} \quad (7)$$

At Fourier frequencies far from the carrier frequency, where the integrated PM noise from ∞ to f (the Fourier frequency) is less than 0.1 rad^2 , $\ell(f)$ can be viewed as the ratio of phase noise power in a single sideband to power in the carrier (single-sideband phase noise). When $\ell(f)$ is expressed in the form $10 \log[\ell(f)]$ its units are dB below the carrier in a 1 Hz bandwidth ($d\beta c/\text{Hz}$).

Frequency fluctuations in the frequency domain are characterized by the spectral density of the fractional frequency fluctuations $S_y(f)$, given by

$$S_y(f) = \text{PSD}[y(t)] = [y(f)]^2 \frac{1}{\text{BW}} \quad (8)$$

where $y(f)^2$ represents the mean-squared fractional frequency deviation at an offset (Fourier) frequency f from the carrier. $S_y(f)$ is defined for Fourier frequencies $0 < f < \infty$, and its units are $1/\text{Hz}$.

The conversion between $S_y(f)$ and $S_\phi(f)$ can be obtained from Eq. (4). Applying the Fourier transform to both sides of Eq. (4), squaring, and dividing by the measurement bandwidth result in

$$S_y(f) = \left(\frac{1}{2\pi\nu_0} \right)^2 (2\pi f)^2 S_\phi(f) = \left(\frac{f}{\nu_0} \right)^2 S_\phi(f) \quad (9)$$

Amplitude fluctuations in the frequency domain are characterized by the spectral density of the fractional amplitude fluctuations $S_a(f)$, given by

$$S_a(f) = \text{PSD} \left(\frac{\epsilon(t)}{V_0} \right) = \left(\frac{\epsilon(f)}{V_0} \right)^2 \frac{1}{\text{BW}} \quad (10)$$

where $\epsilon(f)^2$ represents the mean-squared amplitude deviation at an offset frequency f from the carrier (1). $S_a(f)$ is defined for Fourier frequencies $0 < f < \infty$, and its units are $1/\text{Hz}$.

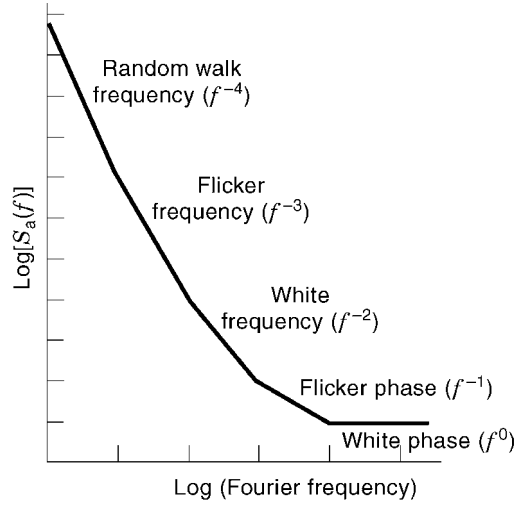
Figure 4(a) and (b) shows the common noise types characteristic of the PM noise and the AM noise of an oscillator (1,2,4–6).

Effects of Frequency Multiplication and Heterodyning on PM, FM, and AM Noise

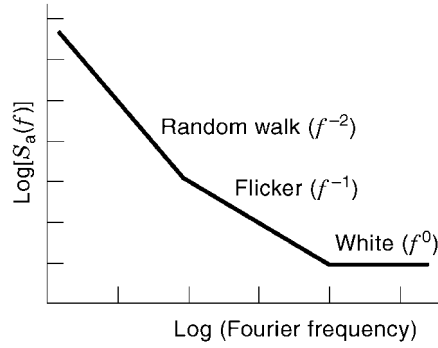
When the frequency of a signal is multiplied by N , the phase fluctuations are also multiplied by N , as shown in Fig. 5(a). The PM noise of the multiplied signal is given by

$$\begin{aligned} S_{\phi_2}(f) &= \frac{[\phi_2(f)]^2}{\text{BW}} + S_{\phi,M}(f) = \frac{N^2[\phi_1(f)]^2}{\text{BW}} + S_{\phi,M}(f) \\ &= N^2 S_{\phi_1}(f) + S_{\phi,M}(f) \end{aligned} \quad (11)$$

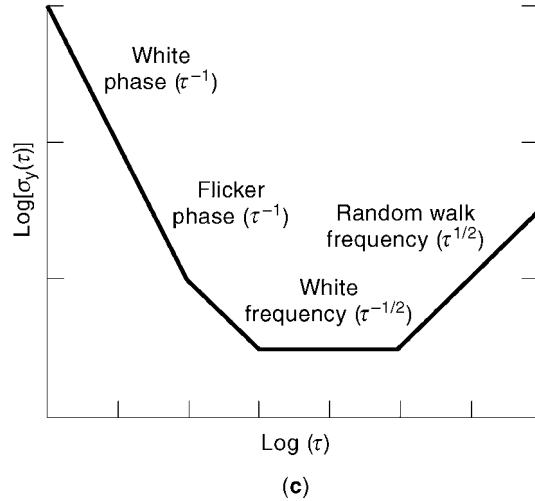
where $S_{\phi,M}(f)$ is the PM noise added by the frequency multiplier. Similarly, when the frequency of a signal is divided by N , the PM noise $S_\phi(f)$ of the divided signal is divided by N^2 . Frequency multiplication and frequency division do not alter the fractional FM noise $S_y(f)$ of a signal since both the frequency



(a)



(b)



(c)

Figure 4. Common types of noise (fluctuations) in oscillators: (a) PM noise; (b) AM noise; (c) $\tau_y(\tau)$ (1,2,4–6).

fluctuations and the nominal frequency are multiplied by N , and the ratio remains constant (7). Ideally, frequency multiplication or division should not have an effect on AM noise either. Nevertheless, the AM noise of the multiplied or divided signal can be affected and determined by the multiplication or division scheme.

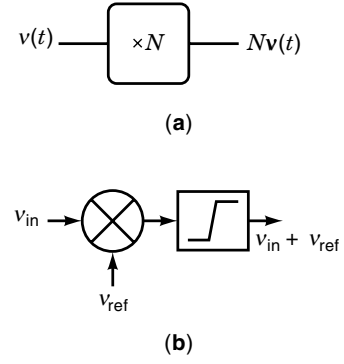


Figure 5. (a) Block diagram of a frequency multiplication system. (b) Block diagram of a frequency heterodyne or translation system.

A system that translates or shifts the frequency of an input signal by a fixed frequency is shown in Fig. 5(b). In this system, a mixer is used to multiply the input and reference signals. The output signal after the high-pass filter has a frequency of $v_{in} + v_{ref}$. (Alternately the lower sideband $v_{in} - v_{ref}$ could just as well have been chosen.) The input frequency has been shifted by the frequency of the reference. The PM noise of the output signal of a frequency translation or frequency heterodyne system is given by

$$S_{\phi,o}(f) = S_{\phi,in}(f) + S_{\phi,ref}(f) + S_{\phi,T}(f) \quad (12)$$

where $S_{\phi,in}(f)$ is the PM noise of the input signal, $S_{\phi,ref}(f)$ is the PM noise of the reference, and $S_{\phi,T}(f)$ is the PM noise of the translator (in this case the mixer and the high-pass filter). The AM noise of the output signal will depend on the details of the translation scheme.

Time-Domain Fractional Frequency Stability of a Signal

In the time domain, the fractional frequency stability of a signal is usually characterized by the Allan variance, a type of two-sample frequency variance given by

$$\sigma_y^2(\tau) = \frac{1}{2(N-2)\tau^2} \sum_{i=1}^{N-2} (x_{i+2} - 2x_{i+1} + x_i)^2 \quad (13)$$

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{i=1}^{M-1} (\bar{y}_{i+1} - \bar{y}_i)^2 \quad (14)$$

$$\sigma_y^2(\tau) = \frac{2}{(\pi v_0 \tau)^2} \int_0^\infty S_\phi(f) \sin^4(\pi f \tau) df \quad (15)$$

where N is the number of time deviation samples, x_i is the time deviation over the interval τ , $M = N - 1$ is the number of frequency samples, and $\bar{y}_i$ is the fractional frequency deviation for interval i (1–4). Equation (13) is used when time data are available, Eq. (14) is used when frequency data are available, and Eq. (15) is used to convert frequency domain data (PM noise) to the time domain. The squared root of the Allan variance, $\sigma_y(\tau)$, is generally used to specify the frequency stability of a source. Figure 4(c) shows the slopes of the common noise types characteristic of the $\sigma_y(\tau)$ of oscillators. If the dominant noise type in

short-term is flicker PM or white PM, the modified Allan variance given by Eq. (16), (17), or (18) can be used to improve the estimate of the underlying frequency stability of the sources (1,2,8):

$$\text{mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2 m^2 (N - 3m + 1)} \sum_{j=1}^{N-3m+1} \left(\sum_{i=j}^{m+j-1} (x_{i+2m} - 2x_{i+m} + x_i) \right)^2 \quad (16)$$

$$\text{mod } \sigma_y^2(\tau) = \frac{1}{2(N - 3m + 1)} \sum_{j=1}^{N-3m+1} (\bar{y}'_{j+m} - \bar{y}'_j)^2 \quad (17)$$

$$\text{mod } \sigma_y^2(\tau) = \frac{2}{m^4 (\pi \nu_0 \tau_0)^2} \int_0^\infty S_\phi(f) \frac{\sin^6(\pi \tau f)}{\sin^2(\pi \tau_0 f)} df \quad (18)$$

where

$$\bar{y}'_j = \frac{\bar{x}_{j+m} - \bar{x}_j}{\tau} \quad (19)$$

and

$$\bar{x}_j = \frac{\sum_{k=0}^{m-1} x_{j+k}}{m} \quad (20)$$

Here $\bar{x}_j$ is the phase (time) averaged over n adjacent measurements of duration τ_0 . Thus $\text{mod } \sigma_y(\tau)$ is proportional to the second difference of the phase averaged over a time $m\tau_0$. Viewed from the frequency domain, $\text{mod } \sigma_y(\tau)$ is proportional to the first difference of the frequency averaged over m adjacent samples.

The confidence intervals for $\sigma_y(\tau)$ and $\text{mod } \sigma_y(\tau)$ as a function of noise type and the number of samples averaged are discussed in Refs. 1 and 4.

MEASUREMENT SYSTEMS

Direct Measurements

Direct Measurements of Frequency and Frequency Stability Using a Counter. Figure 6 shows the timing diagram for the direct measurement of signal frequency relative to a reference frequency using a counter. The normal convention is to start with the signal under test and stop with the reference. The user typically chooses the nominal measurement period, which is an integral number of cycles of the reference. For example $\tau = N_{\text{ref}}/\nu_{\text{ref}} \cong 1$ s. The instrument counts N_{sig} , the nominal number of cycles of the signal under test that occur before the reference signal stops the count. The frequency of the signal averaged over a measurement interval τ is given by

$$\nu_{\text{sig}}(\tau) = N_{\text{sig}}(\tau) \frac{\nu_{\text{ref}}}{N_{\text{ref}}} \quad (21)$$

where ν_{ref} is the frequency of the reference. Since the measured signal frequency is proportional to the time base or ref-

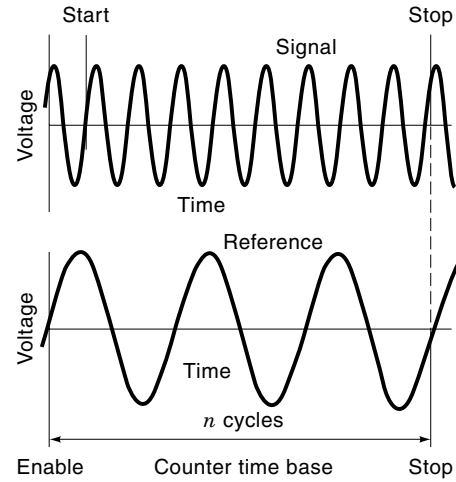


Figure 6. Timing diagram for a direct frequency measurement system.

erence frequency, an error in the time-base frequency leads to a proportional error in the determination of the signal frequency.

The intrinsic fractional frequency resolution of a simple counter is $\pm 1/N_{\text{sig}}$. Some sophisticated counters have interpolation algorithms that allow them to improve the intrinsic resolution by a factor of β , which can be 100 or more. If the frequency of the signal under test is less than the frequency of the reference, the resolution can often be improved by reversing the roles of the signal and reference. The counter reading can then be inverted to find the frequency of the signal.

The uncertainty in the frequency from a particular measurement taken over a measurement time τ is the intrinsic resolution plus the combined fractional instability of signal $\sigma_{y,\text{sig}}^2(\tau)$ and the reference $\sigma_{y,\text{ref}}^2(\tau)$. When these factors are independent, they are added in quadrature:

frequency uncertainty

$$= \frac{\Delta \nu(\tau)}{\nu} = \left[\left(\frac{1}{\beta N_{\text{sig}}} \right)^2 + \sigma_{y,\text{ref}}^2(\tau) + \sigma_{y,\text{sig}}^2(\tau) \right]^{1/2} \quad (22)$$

The time-domain fractional frequency stability of the signal $\sigma_{y,\text{sig}}^2(\tau)$ relative to the frequency stability of the reference can be estimated from a series of consecutive frequency measurements using Eq. (14) (1,2). There is often dead time between frequency measurements in the direct method, which leads to biases in the estimation of fractional frequency stability; these biases depend on noise type (9). More elaborate techniques that eliminate this bias and offer better intrinsic frequency resolution are described in later sections on heterodyne measurements.

Direct Measurements of Phase or Time Using a Counter. For sinusoidal signals, phase or time is usually referenced to the positive-going zero crossing of the signal. For digital signals, time is usually referenced to the mean of the 0 and the 1 states at the positive-going transition. Although the counter can be started with the signal or the reference, we usually start with the signal. Then advancing phase (time) corresponds to a

signal frequency that is higher than the reference. The instrument counts N_{ref} , the nominal number of cycles of the counter time base frequency ν_{ref} that occur before the reference signal stops the count. The phase of the signal relative to the reference is

$$\theta_{\text{sig}} = 2\pi N_{\text{ref}} \frac{\nu_{\text{sig}}}{\nu_{\text{ref}}} \quad (23)$$

where ν_{sig} is the frequency of the signal. The time of the signal relative to the reference is

$$t_{\text{sig}} = \frac{N_{\text{ref}}}{\nu_{\text{ref}}} \quad (24)$$

Since the measured phase (time) is proportional $1/(\nu_{\text{ref}})$, an error in the time base (reference) frequency leads to an error in the determination of the signal phase. There may also be phase errors or time errors in the measurement due to the voltage standing wave ratio (VSWR) on the transmission lines to the counter from the signal and the reference (10).

For a simple counter the intrinsic phase resolution is $2\pi \nu_{\text{sig}}/\nu_{\text{ref}}$, while the time resolution is $1/\nu_{\text{ref}}$. Some sophisticated counters have interpolation algorithms that allow them to improve the intrinsic resolution by a factor of β , which can be 100 or more.

The uncertainty in the measurement of phase $\Delta\phi$ or time ΔT using this approach is given by the intrinsic resolution plus the combined fractional instability of signal $\sigma_{y_{\text{sig}}}^2(\tau)$ and the reference $\sigma_{y_{\text{ref}}}^2(\tau)$. When these factors are independent, they are added in quadrature:

$$\Delta\phi = \theta_{\text{sig}} \left[\left(\frac{1}{N_{\text{ref}}\beta} \right)^2 + \sigma_{y_{\text{ref}}}^2(\tau) + \sigma_{y_{\text{sig}}}^2(\tau) \right]^{1/2} \quad (25)$$

$$\Delta T = t_{\text{sig}} \left[\left(\frac{1}{N_{\text{ref}}\beta} \right)^2 + \sigma_{y_{\text{ref}}}^2(\tau) + \sigma_{y_{\text{sig}}}^2(\tau) \right]^{1/2} \quad (26)$$

The time-domain fractional frequency stability of the signal $\sigma_{y_{\text{sig}}}^2(\tau)$ relative to the frequency stability of the reference can be estimated from a series of consecutive phase measurements separated by a time τ , using Eq. (13) (1,2). More elaborate techniques that offer better resolution are described in later sections on heterodyne measurements.

Heterodyne Measurements

Heterodyne techniques offer greatly improved short-term resolution over direct measurement techniques of frequency, phase, and time (see Fig. 7). In this technique the signal under test is heterodyned against the reference signal ν_{ref} and the difference frequency or beat frequency (lower curve) ν_b measured. The frequency resolution is improved by a factor ν_{ref}/ν_b over direct measurements.

Heterodyne Measurements of Frequency. Using the heterodyne method, the frequency of the signal source is

$$\nu_{\text{sig}} = \nu_{\text{ref}} \pm \nu_b \quad (27)$$

Additional measurements are required to determine the sign of the frequency difference. The usual method is to change the

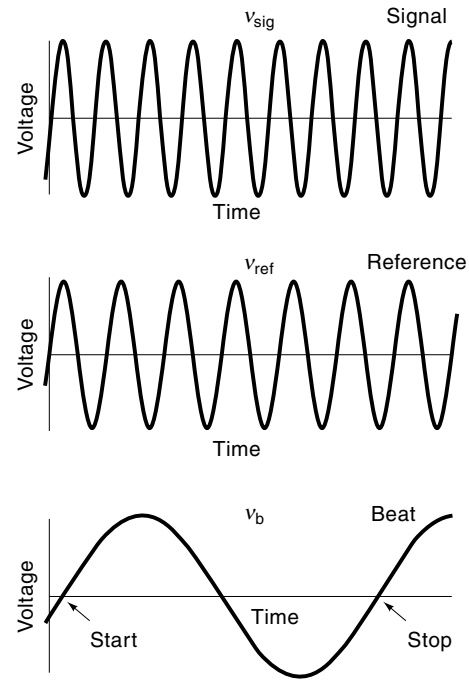


Figure 7. Timing diagram of a heterodyne time measurement system.

frequency of the reference by a known amount and determine whether the beat becomes smaller or larger. The resolution for frequency measurements is given by

$$\frac{\Delta\nu}{\nu_{\text{ref}}} = \Delta t \frac{\nu_b^2}{\nu_{\text{ref}}} \quad (28)$$

where Δt is the timing resolution. The uncertainty is limited by the frequency stability of the reference and the phase variations of the phase detector. The minimum time between data samples is $1/\nu_b$ for clocks that are nearly at the same frequency (a limitation in many situations).

Fifty percent of the time this approach is insensitive to the phase fluctuation between the signal and the reference. This down time, called “dead time” (5,9), biases the calculation of $\sigma_y(\tau)$ and $\text{mod } \sigma_y(\tau)$ by an amount that depends on the noise type and the duration of the dead time. Bias tables as a function of noise type and percent dead time are given in Ref. 9. The dead-time limitation can be circumvented by using two counters triggered on alternate cycles of ν_b .

Heterodyne Measurements of Phase (Time). The resolution for heterodyne measurements of time or phase is increased to

$$\Delta\tau = \frac{\nu_b}{\nu_{\text{ref}}} \Delta t \quad (29)$$

where Δt is the timing resolution of the counter. To avoid ambiguity ν_b should be larger than the frequency fluctuations of both the reference and the source under test. Additional measurements are required to determine if the frequency of the source is higher or lower than the reference. The phase of the beat signal goes through zero when the phase difference between the two signals is $\pm(2n + 1) \times 90^\circ$ where $n = 0, 1, 2, 3, \dots$. The time of the zero crossing is biased or in error by $\Delta\phi$ due

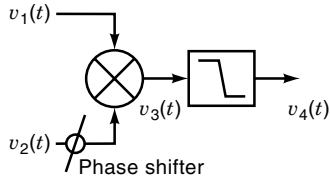


Figure 8. Basic building block for AM and PM noise measurement systems.

to imperfections in the symmetry of the phase detector and/or the VSWR in the reference and signal paths (10). Timing errors due to VSWR effects and typical temperature coefficients for mixer biases are given in Ref. 10 for frequencies of 5 and 100 MHz. These errors generally scale as $1/\nu_{\text{sig}}$.

The time of the source under test is

$$t_{\text{sig}} = T_{\text{ref}} \pm \frac{n}{\nu_{\text{ref}}} \pm \Delta\phi \quad (30)$$

where n is the number of beat cycles that have occurred since the original synchronization. The minimum time between data samples is $1/\nu_b$. For clocks that are at nearly the same frequency, this limitation can be very restrictive.

The time-difference data can be used to characterize the fractional frequency stability of the sources using Eq. (13). The resolution for short-term time domain frequency stability (τ less than 0.1 s) is typically much worse than that obtained from integrating the phase noise using Eq. (15) (8,11)

Basic Configuration of PM and AM Noise Measurement Systems

Figure 8 shows the basic building block used in PM and AM noise measurement systems. It consists of a phase shifter, a mixer, and a low-pass filter. The two input signals to the mixer can be represented as

$$\begin{aligned} v_1(t) &= A_1(t) \cos[2\pi\nu_0 t + \Phi_1(t)] \\ &= [V_1 + \epsilon_1(t)] \cos[2\pi\nu_0 t + \phi_1 + \phi_1(t)] \end{aligned} \quad (31)$$

$$\begin{aligned} v_2(t) &= A_2(t) \cos[2\pi\nu_0 t + \Phi_2(t)] \\ &= [V_2 + \epsilon_2(t)] \cos[2\pi\nu_0 t + \phi_1 + \phi_2(t)] \end{aligned} \quad (32)$$

where $\epsilon_1(t)$ and $\epsilon_2(t)$ represent the amplitude fluctuations of the signals and $\phi_1(t)$ and $\phi_2(t)$ represent the phase fluctuations of the signals. For an ideal mixer, the signal $v_3(t)$ is equal to the product of signals $v_1(t)$ and $v_2(t)$. After some algebraic manipulation, we have

$$\begin{aligned} v_3(t) &= \frac{A_1(t)A_2(t)}{2} \{ \cos[4\pi\nu_0 t + \Phi_1(t) + \Phi_2(t)] \\ &\quad + \cos[\Phi_1(t) - \Phi_2(t)] \} \end{aligned} \quad (33)$$

At the output of the low-pass filter, the signal reduces to

$$v_4(t) = \frac{A_1(t)A_2(t)}{2} \{ \cos[\Delta\phi + \Delta\phi(t)] \} \quad (34)$$

where $\Delta\phi$ is the difference between the phase angles of the two input signals ($\phi_1 - \phi_2$) and $\Delta\phi(t) = \phi_1(t) - \phi_2(t)$. When $\Delta\phi$ is approximately equal to an odd multiple of $\pi/2$ and the

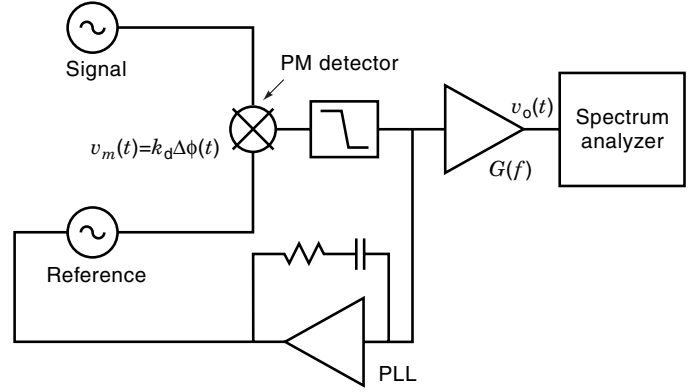


Figure 9. Block diagram for a heterodyne (two-oscillator) PM noise measurement system.

integrated phase noise $[\int \Delta\phi(f)^2 df]$ does not exceed 0.1 rad^2 , the two signals are in quadrature and the output voltage of the mixer is proportional to the difference of the phase fluctuations in the signals $v_1(t)$ and $v_2(t)$. When a double-balanced mixer is used, the amplitude fluctuations are suppressed by 25 to 40 dB (11,12). This is also called suppression of AM to PM conversion in a double-balanced mixer.

When the two input signals to the mixer are in phase or $\Delta\phi \cong 0$, the output voltage of the mixer is proportional to the amplitude fluctuations of the signals. The suppression of phase fluctuations when $\Delta\phi \cong 0$ is higher than 90 dB (11,12). The setup in Fig. 8 can be used in either AM noise or PM noise measurement systems by adjusting the phase shifter.

PM Noise Measurement Systems For Oscillators and Amplifiers

Heterodyne Measurements of PM Noise in Signal Sources (Two-Oscillator Method). Figure 9 shows a heterodyne PM noise measurement system for an oscillator. In this system, the signals from the test oscillator and a reference oscillator of similar frequency are fed into a double-balanced mixer. A phase-locked loop (PLL) is used to lock the reference frequency to the test oscillator frequency and to maintain quadrature between the two input signals to the mixer (13). The output voltage of the mixer is proportional to the difference between the phase fluctuations of the two sources. This voltage is amplified and its PSD is measured with a spectrum analyzer. Often this spectrum analyzer is of the fast Fourier transform (FFT) type. The voltage at the output of the amplifier is

$$v_o(t) = k_d G \Delta\phi(t) \quad (35)$$

where k_d is the mixer's phase-to-voltage conversion factor (or mixer sensitivity), G is the gain of the amplifying stage, and $\Delta\phi(t)$ is the difference between the phase fluctuations of the test oscillator and the reference $[\phi_A(t) - \phi_B(t)]$. The PM noise can be obtained from

$$S_{\phi}(f) = \frac{\text{PSD}[v_o(t)]}{(k_d G)^2} \quad (36)$$

The calibration factor or mixer sensitivity k_d can be found by turning off the PLL to obtain a beat frequency signal at the mixer output. The slope at the zero crossing and the period of

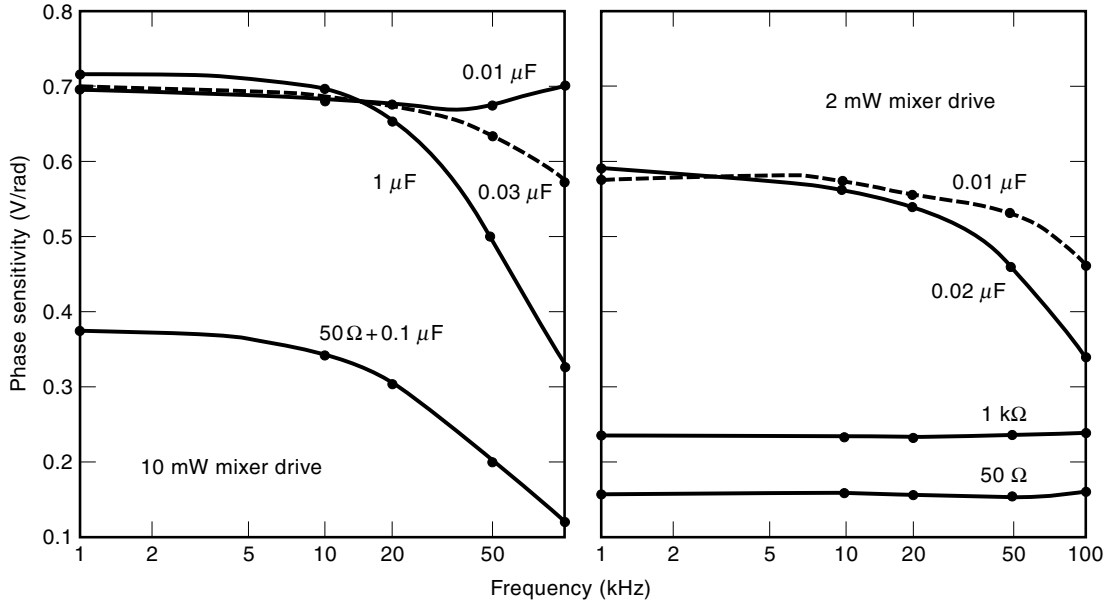


Figure 10. Sensitivity of a low-level double-balanced mixer at 5 MHz as a function of the intermediate frequency (IF) port termination for radio frequency (RF) and local oscillator (LO) inputs of +2 and +10 dBm (11).

this signal can be measured with an oscilloscope or another recording device. The calibration factor k_d in V/rad is

$$k_d = (\text{slope}) \times \frac{T}{2\pi} \quad (37)$$

Ideally this measurement should be made at the output of the amplifier because the measurement then yields $k_d G$, which includes the effect of amplifier input impedance on the performance of the mixer. The calibration of this PM noise measurement system using the beat-frequency method can introduce errors in the measurement if the mixer and the amplifier gains are frequency dependent, as is often the case. Figure 10 shows the variation of k_d with Fourier frequency for different mixer terminations. Capacitive terminations improve the mixer sensitivity, thereby improving the noise floor. However, the frequency response is not nearly as constant as the response obtained with resistive terminations, thus increasing the measurement error (11).

A calibrated Gaussian noise source centered about the carrier frequency can be used to calibrate the frequency-dependent errors (14,15). In this technique Gaussian noise is added to the reference signal by means of a low-noise power summer. Since the Gaussian noise is independent of the reference noise, equal amounts of PM and AM noise are added to the reference signal. When the Gaussian noise is “on,” the PSD of the output noise voltage $v_o(t)$ is equal to

$$\text{PSD}[v_{o,\text{on}}(t)] = (k_d G)^2 S_{\text{PMcal}} \quad (38)$$

where S_{PMcal} is the PM noise added by the Gaussian noise source. The calibration factor as a function of frequency is obtained by dividing $\text{PSD}[v_{o,\text{on}}(t)]$ by S_{PMcal} . A PSD measurement is then made with the Gaussian noise “off.” The calibrated PM

noise is obtained from

$$S_\phi(f) = \frac{\text{PSD}[v_{o,\text{off}}(t)]}{(k_d G)^2} = S_{\text{PMcal}} \frac{\text{PSD}[v_{o,\text{off}}(t)]}{\text{PSD}[v_{o,\text{on}}(t)]} \quad (39)$$

This approach greatly reduces the uncertainty of the measurement because it automatically takes into account the frequency-dependent errors. This approach also reduces the time necessary to make routine PM noise measurements as compared to traditional methods since the measurement is now reduced to a simple ratio measurement between noise on and noise off (14,15). The use of this calibration technique is illustrated in the cross-correlation measurements discussed in the following and shown in Fig. 11.

The confidence intervals for spectrum analyzers (swept and FFT) as a function of the number of measurements can be estimated from Table 1 (4,16–18). Biases in commonly used FFT window functions are discussed in Ref. 4.

Table 1. Approximate Statistical Confidence Interval for FFT and Swept Spectrum Analyzers (16–18). $\beta = N$ (the Number of Averages) for FFT Analyzers and $\beta = N(\text{RBW}/\text{VBW})$ for Swept Spectrum Analyzers. (RBW Refers to the Resolution Bandwidth of the Spectrum Analyzer and VBW refers to the video bandwidth).

β	69% Confidence Interval (dB)		95% Confidence Interval (dB)	
4	–2	+3	–3	+6
6	–1.5	+2.3	–2.5	+5
10	–1.2	+1.7	–2	+4
30	–0.72	+0.88	–1.3	+1.8
100	–0.41	+0.46	–0.76	+0.92
200	–0.3	+0.32	–0.54	–0.63
1,000	–0.13	+0.13	–0.25	–0.27
3,000	–0.08	+0.08	–0.15	+0.15
10,000	–0.04	+0.04	–0.08	+0.08

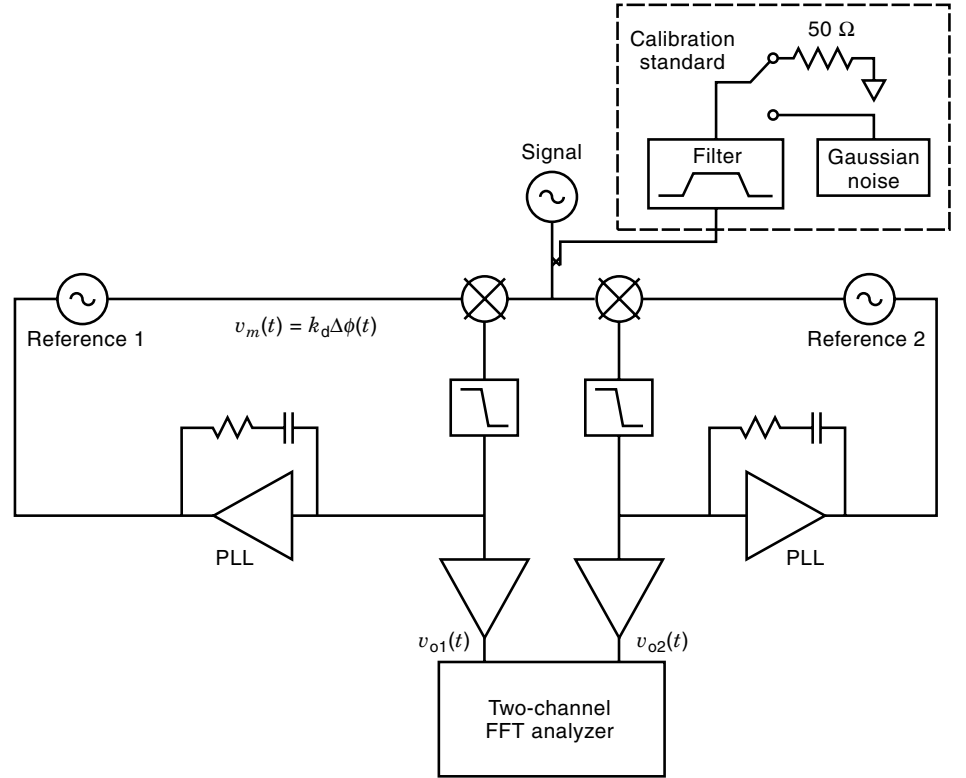


Figure 11. Block diagram of a two-channel cross-correlation system for measuring PM noise in an oscillator. Calibration of the system is accomplished using a PM and AM noise standard (14,15).

One of the shortcomings of the single-channel, two-oscillator measurement system is that the measured noise includes noise contribution from the test source as well as that from the reference source. The noise terms included in $S_\phi(f)$ are

$$S_\phi(f) = S_{\phi,A}(f) + S_{\phi,B}(f) + \frac{v_n^2(f)}{k_d^2 BW} + S_{a,A}(f)\beta_A^2 + S_{a,B}(f)\beta_B^2 \quad (40)$$

where $S_{\phi,A}(f)$ is the PM noise of the test source, $S_{\phi,B}(f)$ is the PM noise of the reference, $v_n(f)$ is the noise added by the mixer, the amplifier, and the spectrum analyzer, $S_{a,A}(f)$ and $S_{a,B}(f)$ refer to the AM noise of the test source and the reference, and β_A and β_B are the factors by which the AM noise is suppressed. If the PM noise of the reference is higher than the PM noise of the test source, then the PM noise of the source cannot be measured accurately.

The last three terms of Eq. (40) constitute the noise floor of the measurement system. The noise floor can be estimated by using a single source (test source or reference) to feed the two inputs of the mixer. A phase shifter placed in one of the channels is used to adjust the phase difference to an odd multiple of 90° . The PM noise of the driving source is mostly canceled and the measured noise at the output is

$$S_\phi(f) = \frac{v_n^2(f)}{k_d^2 BW} + 2S_{a,A}(f)\beta_A^2 + \eta(f)S_{\phi,A}(f) \quad (41)$$

where the factor $\eta(f)$ is due to decorrelation of the source noise and is much smaller than 1 for small Fourier frequencies. [See the section on delay line measurements, especially Eq. (52), for a discussion of this effect.] Equation (41) is approximately

equal to the noise-floor components in Eq. (40). Equation (40) indicates that the AM noise of the source and the reference can affect PM noise measurements; thus sources with low AM noise should be used.

Cross-Correlation Heterodyne Measurements of PM Noise in Signal Sources. One of the limitations of the single-channel, two-oscillator method is that the PM noise of the reference contributes to the measured noise. If three different oscillators are available (A, test source; B, reference 1; C, reference 2), PM noise measurements of three different pairs of oscillators can be made and the PM noise of the source can be approximated by

$$S_{\phi,A}(f) \cong \frac{1}{2} [S_{\phi,AB}(f) + S_{\phi,AC}(f) - S_{\phi,BC}(f)] - 2S_{a,A}(f)\beta_A^2 - \frac{v_n^2(f)}{k_d^2 BW} \quad (42)$$

where $S_{\phi,AB}(f)$ includes the PM noise of sources A and B, $S_{\phi,AC}(f)$ includes the PM noise of sources A and C, and $S_{\phi,BC}(f)$ includes the PM noise of sources B and C. One problem with this approach is that small errors in any of the three measurements taken separately can result in large overall errors. Another problem is that the noise of the measurement system still contributes to the noise floor.

A more effective way of eliminating the PM noise from the reference is to use cross-correlation PM noise measurements. Figure 11 shows a two-channel, cross-correlation PM noise measurement system that uses two reference oscillators and two PM noise detectors operating simultaneously. Each individual channel is a simple heterodyne measurement system. Therefore the noise terms in $\text{PSD}[v_{o1}(t)]$ and $\text{PSD}[v_{o2}(t)]$ divided

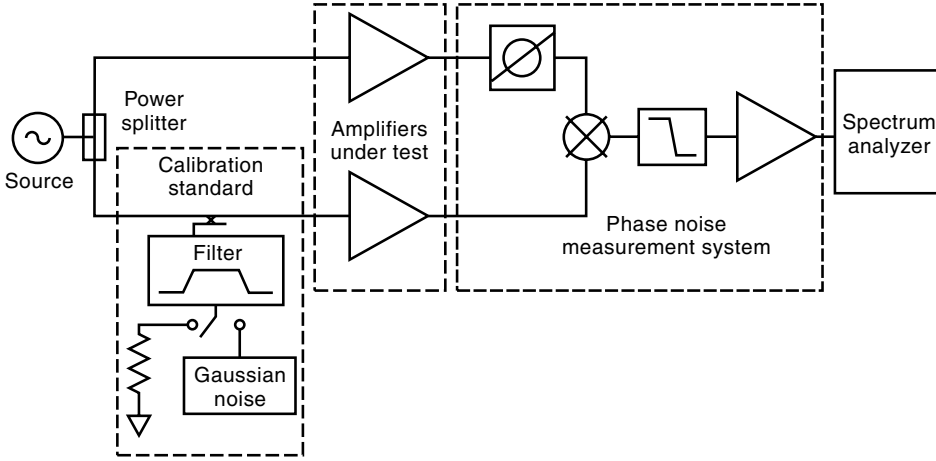


Figure 12. Block diagram of a single-channel system for measuring PM noise in a pair of amplifiers.

by the respective calibration factors are described by Eq. (40). The PSD of the cross-correlation of the noise voltages $v_{o1}(t)$ and $v_{o2}(t)$ divided by the calibration factor is

$$S_{\phi}(f) = S_{\phi,A}(f) + S_{a,A}(f)\beta_A^2 + \frac{1}{\sqrt{N}} \left(S_{\phi,B}(f) + S_{\phi,C}(f) + \frac{v_n^2(f)}{k_d^2 BW} \right) + S_{a,B}(f)\beta_B^2 + S_{a,C}(f)\beta_C^2 \quad (43)$$

where $S_{\phi,A}(f)$ is the PM noise of the test source, $S_{\phi,B}(f)$ and $S_{\phi,C}(f)$ are the PM noise of the references, and N is the number of averages in the measurement (11,14,15,19,20). The contribution of the PM noise of the references and the noise in the detectors and amplifiers are reduced by $\sqrt{N}$, because they are independent. This results in a reduction of the unwanted PM noise in the references and the measurement system of 10 dB for 100 averages and 20 dB for 10,000 averages. This powerful measurement technique makes it possible to obtain an accurate measurement of the PM noise of a source that has lower noise than the references if the AM noise of the source can be neglected.

Heterodyne Measurements of PM Noise in Amplifiers. Figure 12 shows the block diagram of a PM noise measurement system for a pair of amplifiers. In this system an oscillator signal is split using a reactive power splitter. The outputs of the splitter drive two test amplifiers, and their outputs feed a double-balanced mixer. The mixer output is then amplified and measured by a spectrum analyzer. A phase shifter in one of the channels is used to maintain quadrature. The PM noise of the amplifier pair is obtained dividing the PSD of the noise voltage $v_o(t)$ by the calibration factor $(k_d G)^2$,

$$S_{\phi}(f) = \frac{\text{PSD}[v_o(t)]}{(k_d G)^2} = S_{\phi,\text{amp}}(f) + \frac{v_n^2(f)}{k_d^2 BW} + S_a(f)\beta^2 \quad (44)$$

where $S_{\phi,\text{amp}}(f)$ is the PM noise of the amplifier pair and $S_a(f)$ is the AM noise of the source. This system assumes that the noise of the amplifiers is higher than the noise floor of the measurement system. If this is not the case, cross-correlation measurement systems should be used. The calibration factor can be easily obtained by adding a Gaussian noise source to one of the channels and making measurements with the noise on and

the noise off, as discussed previously. Calibration can also be achieved by using a second source to drive one of the amplifiers to obtain a beat signal at the output of the mixer. An oscilloscope can then be used to measure the zero-crossing slope and the period of the beat signal, and the calibration factor can be computed using Eq. (37).

A similar measurement system can be used to measure the PM noise of a single amplifier if the delay across the amplifier is so small that decorrelation of the source noise is not important to the measurement. See Eq. (52) and associated text for a discussion of decorrelation effects.

Cross-Correlation Heterodyne Measurements of PM Noise in Amplifiers. Figure 13 shows a cross-correlation PM noise measurement system for a pair of amplifiers. It consists of two channels, each a separate heterodyne measurement system. The PSD of the cross-correlation of the noise voltages $v_{o1}(t)$ and $v_{o2}(t)$ divided by the calibration factor is

$$S_{\phi}(f) = \frac{\text{PSD}[v_{o1}(t)v_{o2}(t)]}{(k_d G)^2} = S_{\phi,\text{amp}}(f) + \frac{1}{\sqrt{N}} \frac{v_n^2(f)}{(k_d G)^2 BW} + S_{a,A}(f)\beta_A^2 \quad (45)$$

The noise added by the phase detectors is reduced by $\sqrt{N}$. The calibration factor $k_d G$ can be obtained by adding a calibrated Gaussian noise source about the carrier frequency as shown in Fig. 13. In this setup the noise floor is generally limited by the AM noise of the source. It is therefore important to select a source with low AM noise and to operate the mixer at the maximum point of AM rejection.

Delay-Line Measurements of PM Noise in Signal Sources. A PM noise measurement system that does not need a second source is the delay-line system shown in Fig. 14 (11,21). In this setup the oscillator signal is split, a delay line of time delay τ_d is placed in one channel, and a phase shifter is placed in the other channel. The two channels are fed into a double-balanced mixer. The phase fluctuations of the combined signal at the mixer output are given by

$$\phi_m(f) = [2 - 2 \cos(2\pi f \tau_d)]^{1/2} \phi(f) \quad (46)$$

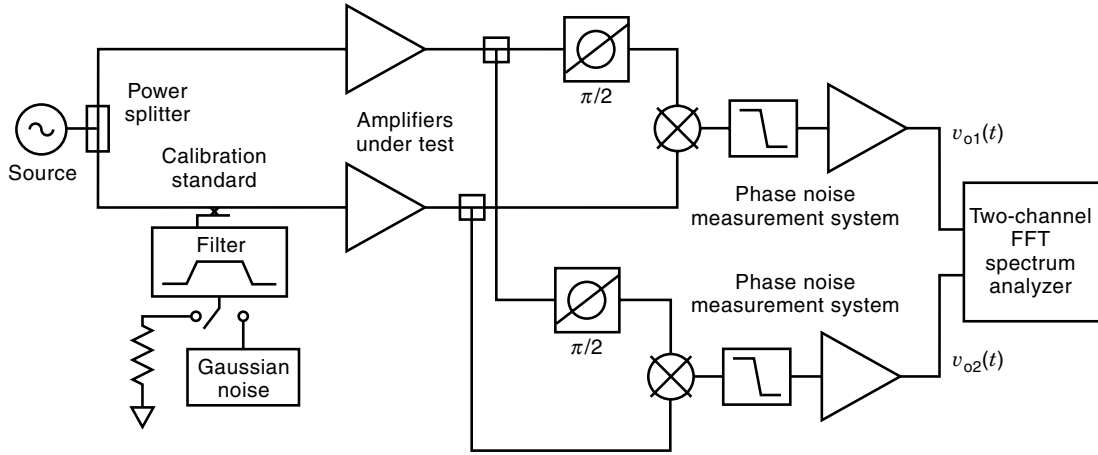


Figure 13. Block diagram of a two-channel cross-correlation system for measuring PM noise in a pair of amplifiers.

where $\phi(f)$ are the rms phase fluctuations of the source at an offset frequency f (21,22). If the phase shifter is adjusted so that the phase difference between the two input signals is an odd multiple of 90° , then the output voltage of the mixer is proportional to the phase fluctuations of the source. For $2\pi f\tau_d \ll 1$,

$$v_m(f) \cong k_d \left[2 - 2 \left(1 - \frac{(2\pi f\tau_d)^2}{2} \right) \right]^{1/2} \phi(f) = k_d 2\pi f\tau_d \phi(f) \quad (47)$$

where $v_m(f)$ is the output voltage of the mixer at an offset frequency f . Equation (47) can also be expressed in terms of $y(f)$ by multiplying the right side by v_0/f :

$$v_m(f) \cong k_d 2\pi v_0 \tau_d y(f) \quad (48)$$

Equation (48) can be equivalently expressed in the time domain as

$$v_m(t) \cong k_d 2\pi v_0 \tau_d y(t) \quad (49)$$

Equation (49) indicates that the output voltage of the mixer is proportional to the frequency fluctuations in the source, and thus this system measures the FM noise of the test source. The

voltage at the output of the amplifier is given by

$$v_o(t) \cong k_d (2\pi v_0 \tau_d) G y(t) = v_0 k_v G y(t) \quad (50)$$

where k_v is the mixer sensitivity to frequency fluctuations and G is the voltage gain of the amplifier. The FM and PM noise of the source can then be obtained:

$$S_y(f) \cong \frac{\text{PSD}[v_o(t)]}{(2\pi v_0 \tau_d k_d G)^2} = \frac{\text{PSD}[v_o(t)]}{(v_0 k_v G)^2} \quad (51)$$

$$S_\phi(f) \cong \frac{\text{PSD}[v_o(t)]}{(2\pi f \tau_d k_d G)^2} = \left(\frac{1}{f} \right)^2 \frac{\text{PSD}[v_o(t)]}{(k_v G)^2} \quad (52)$$

Equations (51) and (52) are valid only for $f \ll 1/(2\pi\tau_d)$, where the approximation in Eq. (47) is valid. The calibration factor k_v can be found by stepping the source frequency up and down and measuring the corresponding voltage change at the output of the amplifier. The voltage change divided by the frequency change is equal to $k_v G$ in V/Hz. At $f > 1/(2\tau_d)$, the output of the mixer is approximately sinusoidal with maximums occurring at odd multiples of $1/(2\tau_d)$, and minimums occurring at even multiples of $1/(2\tau_d)$. The first minimum occurs at $f = 1/\tau_d$, and thus a measurement of this occurrence can be used to determine τ_d (21). A calibrated Gaussian noise source can be used to calibrate the system at Fourier frequencies

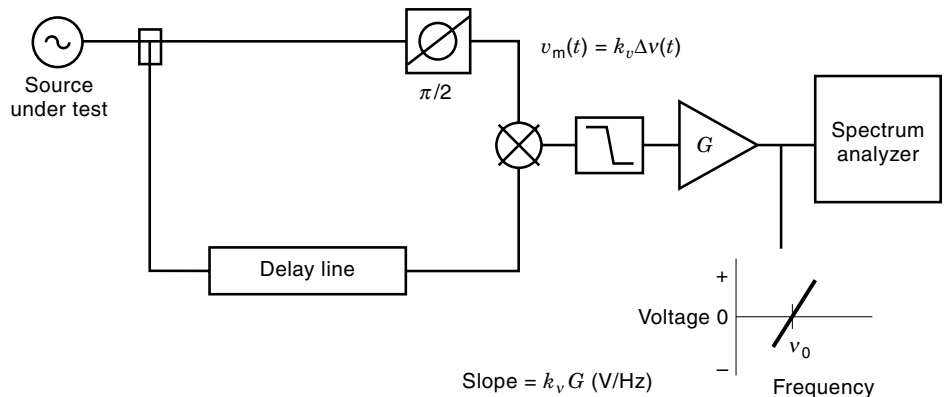


Figure 14. Block diagram of a PM noise measurement system that uses a delay line to measure PM noise in an oscillator.

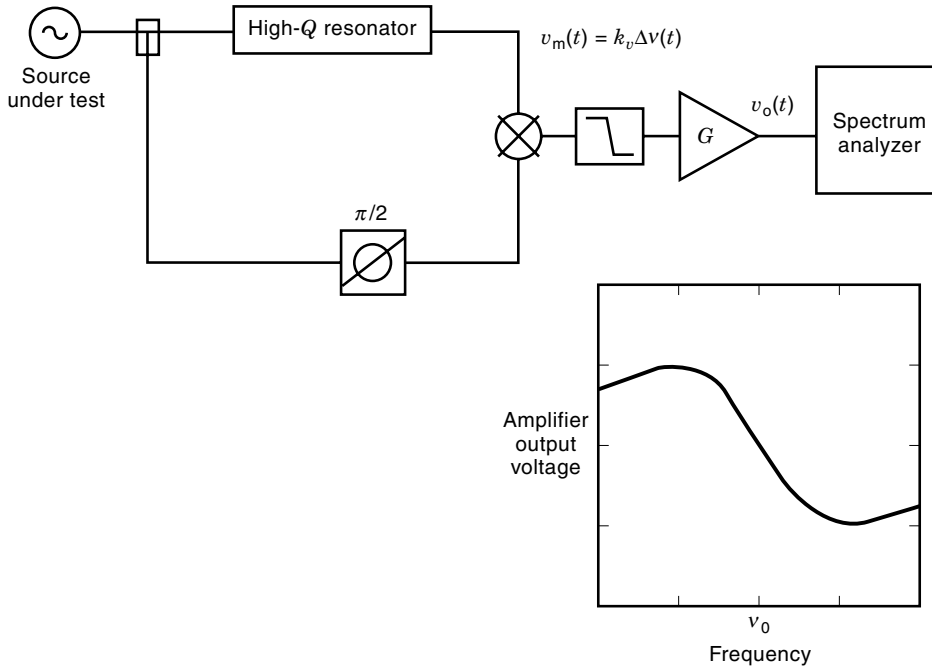


Figure 15. Block diagram of a PM noise measurement system that uses a high-Q-factor cavity to measure PM noise in an oscillator.

higher than $1/(2\tau_d)$, thus extending the frequency range of this system (14,15).

The close-in noise floor of this measurement system is much larger than that of the two-oscillator system. From Eq. (47), the effective phase sensitivity of the mixer is multiplied by a factor of $2\pi f\tau_d$ and thus is less than the phase sensitivity of the two-oscillator method (21). As discussed previously, the noise floor of the two-oscillator measurement system is given by

$$S_{\phi, \text{floor}}(f) = \frac{1}{k_d^2} \frac{v_n^2(f)}{\text{BW}} + S_{a,A}(f)\beta_A^2 + S_{a,B}(f)\beta_B^2 \quad (53)$$

For $f \ll 1/(2\pi\tau_d)$, the noise contributions of the mixer, the amplifier, and the spectrum analyzer to the noise floor increase by a factor of $(2\pi f\tau_d)^{-2}$. The noise floor for the delay line measurement system is given by

$$S_{\phi, \text{floor}}(f) = \frac{1}{(2\pi f\tau_d)^2 k_d^2} \left(\frac{v_n^2(f)}{\text{BW}} \right) + S_{a,A}(f)\beta^2 \quad (54)$$

The first term usually limits the noise floor of this system. The term in large parentheses usually follows a $1/f$ power law at frequencies close to the carrier; thus the overall noise floor at these frequencies follows a f^{-3} dependence on Fourier frequency. In addition, the noise floor is inversely proportional to τ_d^2 ; therefore longer delays will result in lower noise floors, but also a smaller Fourier frequency span in which Eqs. (51) and (52) are valid. At very long delays, the attenuation of the signal is so large that the noise floor is adversely affected. The noise floor of this measurement system cannot be measured directly since a source is needed and the noise from the source cannot be easily separated from the noise of the measurement system. If the noise contribution of the measurement system (mixer, amplifier, and spectrum analyzer) are known, then the noise floor can be approximated using Eq. (54).

Cavity Discriminator Measurements of PM Noise in Signal Sources. Figure 15 shows a cavity discriminator measurement system (11). This system is similar to the delay-line system, but uses a high-Q cavity in place of the delay line. If the signal frequency and the cavity resonance frequency are close, the cavity causes a phase delay proportional to the signal frequency. In the linear region around v_0 , the fractional frequency fluctuations of the source are converted to phase fluctuations according to

$$\phi(t) \cong 2Qy(t) \quad (55)$$

If the phase shifter is adjusted so that the two input signals to the mixer are in quadrature, then the output voltage of the mixer is proportional to phase fluctuations of the source. After amplification, the output voltage is given by

$$v_o(t) \cong k_d(2Q)Gy(t) = v_0 k_v G y(t) \quad (56)$$

where the mixer sensitivity to frequency fluctuations k_v is equal to $2Qk_d/v_0$. (Figure 15 shows the output voltage of the amplifier as a function of input frequency.) The FM and PM noise of the source are therefore

$$S_y(f) \cong \frac{\text{PSD}[v_o(t)]}{(2Qk_dG)^2} = \frac{\text{PSD}[v_o(t)]}{(v_0 k_v G)^2} \quad (57)$$

$$S_\phi(f) \cong \left(\frac{v_0}{f} \right)^2 \frac{\text{PSD}[v_o(t)]}{(2Qk_dG)^2} = \left(\frac{1}{f} \right)^2 \frac{\text{PSD}[v_o(t)]}{(k_v G)^2} \quad (58)$$

Equations (57) and (58) are valid only for $f \ll v_0/(2Q)$, where the phase-to-frequency relation of the cavity is linear and k_v is a constant. The value of k_v can be found by stepping the frequency of the source up and down and measuring the corresponding voltage change at the output of the amplifier. The voltage change divided by the frequency change is equal to $k_v G$ in volts per hertz. At higher Fourier frequencies k_v changes

according to (11)

$$k_v \propto \frac{1}{1 + \left(\frac{2Qf}{v_0}\right)^2} \quad (59)$$

A calibrated Gaussian noise source can be added to the source to calibrate the system at Fourier frequencies higher than $v_0/(2Q)$, thus extending the range of this measurement system.

As in the delay-line measurement system, the effective phase sensitivity of the mixer is less than the phase sensitivity of the two-oscillator method. The noise floor is thus given by

$$S_{\phi, \text{floor}}(f) = \frac{1}{(2Qf)^2 k_d^2} \frac{v_n^2(f)}{\text{BW}} + S_{a,A}(f) \beta^2 \quad (60)$$

Close to the carrier the noise floor follows a f^{-3} dependence on Fourier frequency and is higher than the noise of the two-oscillator measurement system. In addition, the noise floor is inversely proportional to Q^2 . Therefore, higher Q cavities will result in lower noise floors, but also a smaller Fourier frequency span in which Eqs. (57) and (58) are valid. Other resonant circuits, such as multiple-pole filters, can be used in place of the cavity. The resonant circuit used can add noise to the system and thus limit the resolution or noise floor of the measurement system.

AM Noise Measurement Systems For Oscillators and Amplifiers

Simple AM Noise Measurements For Oscillators. Figure 16 shows a single-channel AM noise measurement system for a source. In this system a source drives an AM detector. The output voltage of the detector is then amplified and fed into a swept or FFT spectrum analyzer. AM detectors commonly used are the mixer detector discussed previously or a diode detector.

The voltage at the output of the mixer is given by

$$v_m(t) \cong k_a \frac{\epsilon(t)}{V_0} \quad (61)$$

where k_a is the detector's sensitivity to fractional amplitude fluctuations. The AM noise of the source is then

$$S_a(f) = \frac{\text{PSD}[v_o(t)]}{(k_a G)^2} \quad (62)$$

where $v_o(t)$ is the voltage into the spectrum analyzer and G is the gain of the amplifier. The sensitivity k_a can be measured by replacing the test source with a source with amplitude modulation capability (12,20). The power of the (calibration) source should be the same as the power of the test source, and it should be adjusted so that the dc voltage at the output of the mixer is

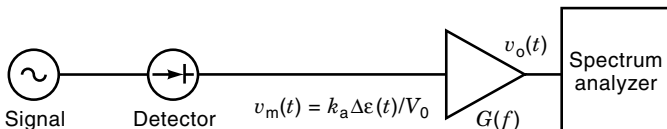


Figure 16. Block diagram of a single-channel AM noise measurement system for measuring AM noise in an oscillator.

the same as when the test source is used. The amplitude modulated input signal is given by

$$v(t) = V_0(1 + \text{AM}_{\text{in}} \cos \Omega t) \cos(2\pi v_0 t) \quad (63)$$

where AM_{in} is the peak fractional amplitude modulation and Ω is the modulation frequency. The magnitude of the amplitude modulation $S_a(f)$ at the input signal is $\frac{1}{2}(\text{AM})^2$, where AM is the modulation selected. The amplitude modulation at the output (AM_{out}) is then measured with the spectrum analyzer and the calibration factor is given by

$$(k_a G)^2 = \frac{2\text{AM}_{\text{out}}}{\text{AM}_{\text{in}}} \quad (64)$$

This measurement assumes that the calibration factor is a constant independent of the Fourier frequency. Many times k_a and G vary with frequency, and thus errors are introduced in the calibration. This is especially important for f higher than 100 kHz. To avoid this problem and speed the measurement of $k_a G$, a Gaussian noise source added to the test source can be used to calibrate the system (12,14,15). To calibrate the system, the PSD of $v_o(t)$ with the Gaussian noise on is measured with the spectrum analyzer. This curve will show any variation of $k_a G$ with Fourier frequency if it exists. The values of $(k_a G)^2$ as a function of Fourier frequency are obtained by dividing the measured $v_o(f)$ by the known calibrated noise power.

The confidence intervals for spectrum analyzer (swept and FFT) measurements, as a function of the number of measurements, can be estimated from Table 1 (4,16–18). Biases in commonly used FFT window functions are discussed in Ref. 4.

One problem of this simple measurement system is that the measured noise, given by Eq. (62), includes the AM noise of the source in addition to noise added by the detector, the amplifier, and the analyzer (noise floor of the system). The noise components included in $S_a(f)$ are

$$S_a(f) = S_{a,\text{src}}(f) + \frac{1}{(k_a)^2} \frac{v_n^2(f)}{\text{BW}} \quad (65)$$

where $S_{a,\text{src}}(f)$ is the AM noise of the source, and the second term represents the noise floor of the system. It is therefore difficult to separate the AM noise of the source from the noise floor.

Cross-Correlation AM Noise Measurements For Oscillators. The noise floor of the AM noise measurement system discussed

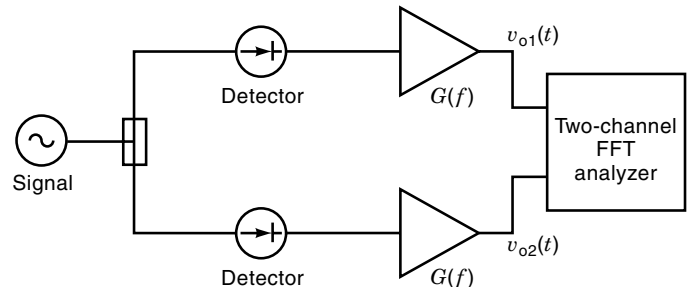


Figure 17. Block diagram of a two-channel cross-correlation system for measuring AM noise in an oscillator.

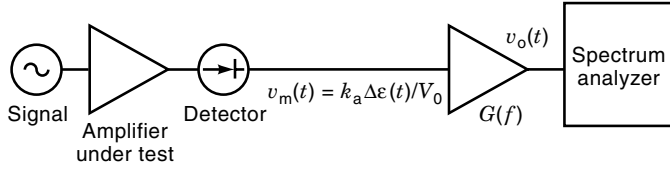


Figure 18. Block diagram of a single-channel system for measuring AM noise in an amplifier.

previously can be considerably reduced by using two-channel cross-correlation techniques as shown in Fig. 17. In this system $\text{PSD}[v_{o1}(t)]$ includes AM noise of the source plus noise in channel 1 (detector 1, amplifier 1). $\text{PSD}[v_{o2}(t)]$ includes AM noise of the source plus noise in channel 2 (detector 2, amplifier 2). The PSD of the cross spectrum divided by the calibration factor mainly includes the AM noise of the source since the noise that is not common between the two channels is reduced proportionally to $1/\sqrt{N}$, so

$$\frac{\text{PSD}[v_{o1}(t)v_{o2}(t)]}{(k_a G)^2} = S_{a,\text{src}}(f) + \frac{1}{\sqrt{N}} \left(\frac{v_{n1}^2(t)}{(k_{a1})^2 \text{BW}} + \frac{v_{n2}^2(t)}{(k_{a2})^2 \text{BW}} \right) \quad (66)$$

This measurement technique is very useful for separating the AM noise of the source from the system noise. This system can be calibrated with a source with AM capability or by using a calibrated Gaussian noise source (12,14).

Simple AM Noise Measurements For Amplifiers. Figure 18 shows a single-channel AM noise measurement system for an amplifier. A source is used to drive the amplifier under test, and the output signal of the amplifier is fed into an AM detector. The signal is amplified, and the output voltage is measured with a spectrum analyzer. The PSD of the noise voltage $v_o(t)$ divided by the calibration factor $(k_a G)^2$ is

$$\frac{\text{PSD}[v_o(t)]}{(k_a G)^2} = S_{a,\text{src}}(f) + S_{a,\text{amp}}(f) + \frac{1}{(k_a)^2} \frac{v_n^2(t)}{\text{BW}} \quad (67)$$

The first term in Eq. (67) represents the AM noise of the source, the second term is the noise of the test amplifier, and the third term is the noise floor of the measurement system. The calibration factor $k_a G$ can be obtained by using a source with AM capability or by using a calibrated Gaussian noise source to calibrate the system (12,14,20). If the AM noise of the source

or the system noise floor is comparable to the amplifier noise, accurate measurements of the amplifier AM noise cannot be made using this system.

Cross-Correlation AM Noise Measurements For Amplifiers.

Figure 19 shows a cross-correlation AM noise measurement system for an amplifier that reduces the noise floor of the measurement system. A reference source drives the test amplifier, the amplifier output is split, and each channel is fed into an AM detector. The output signals of the detectors are then amplified and measured with a two-channel FFT spectrum analyzer. The PSD of the cross-correlation $[v_{o1}(t)xv_{o2}(t)]$ divided by the calibration factor is

$$\frac{\text{PSD}[v_{o1}(t)v_{o2}(t)]}{(k_a G)^2} = S_{a,\text{src}}(f) + S_{a,\text{amp}}(f) + \frac{1}{\sqrt{N}} \left(\frac{v_{n1}^2(t)}{(k_a)^2 \text{BW}} + \frac{v_{n2}^2(t)}{(k_a)^2 \text{BW}} \right) \quad (68)$$

Equation (68) indicates that part of the noise floor in the system is reduced by a factor of $\sqrt{N}$. If the AM noise of the source dominates the measured noise, a limiter can be placed after the source to reduce its AM noise (6).

CARRIER SUPPRESSION MEASUREMENT SYSTEMS

The concept of carrier suppression was first introduced by Sann for measuring noise in amplifiers (23). Carrier suppression measurement systems use a bridge network to cancel the power at the carrier frequency, effectively enhancing the noise of the device under test (23–26). This results in a reduction of the noise floor of the measurement system, or in other words, in an increase of the detector sensitivity.

Carrier Suppression AM Noise Measurement Systems for Amplifiers

Figure 20 shows an AM noise measurement system for amplifiers that uses carrier suppression (25). The source signal is split, and the amplifier under test is placed in one of the channels. The variable attenuator is used to match the magnitudes of the two channels. The phase shifter is used to adjust the phase difference between the two channels to 0. When the two signals are combined in the hybrid, the carrier power is mostly cancelled. The degree of cancellation depends on how well the two channels are matched. An additional amplifier is placed at the output of the power summer to increase the power input to the AM detector. The effective AM noise of the device (as seen

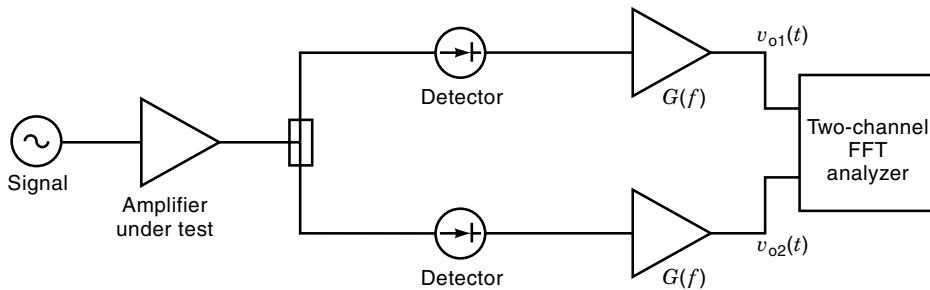


Figure 19. Block diagram of a two-channel cross-correlation system for measuring AM noise in an amplifier.

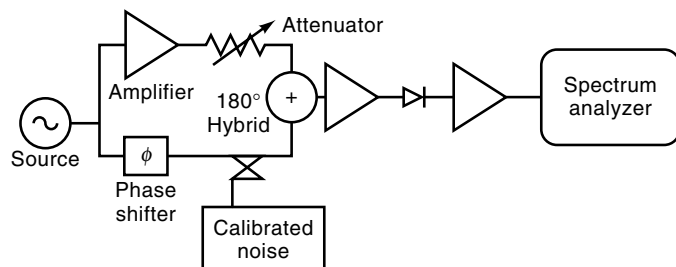


Figure 20. Carrier suppression AM noise measurement system for amplifiers.

by the AM detector) is increased by the amount of carrier suppression. An AM noise standard is used to calibrate the gain of the system.

Carrier Suppression PM Noise Measurement System for Amplifiers

Figure 21 shows a carrier suppression PM noise measurement system for amplifiers (25,26). A bridge circuit is used to effectively increase the PM noise of the amplifier with respect to the noise floor of the system. A PM noise standard is used to calibrate the gain of the system.

Carrier Suppression PM Noise Measurement Systems for Oscillators

Carrier suppression can also be used in PM measurement systems for oscillators. Figure 22 shows a two-oscillator PM noise measurement system with carrier suppression (25). In this system a bridge is used to raise the magnitude of the PM noise of the two oscillators with respect to the noise in the phase noise detector. A PM noise standard is used to calibrate the gain of the system. Carrier suppression can also be applied to delay line measurement systems (27).

The advantage of carrier suppression measurement systems over cross-correlation systems is that similar noise floors can be achieved at smaller measurement times. The disadvantage is that a very good amplitude and phase match of the bridge channels is required to suppress the carrier power. Furthermore, the match changes with temperature, and thus careful calibrations should be performed before and after a measurement to ensure

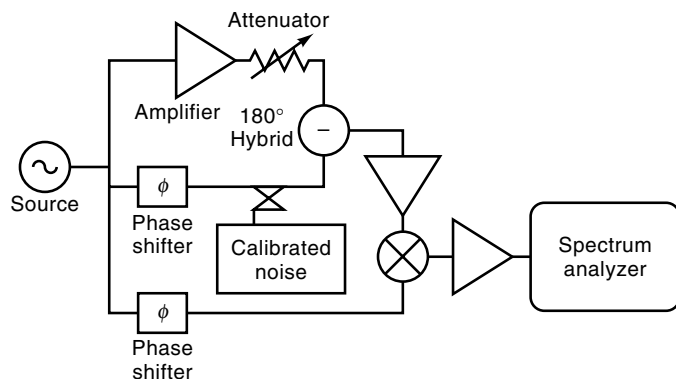


Figure 21. Carrier suppression PM noise measurement system for amplifiers.

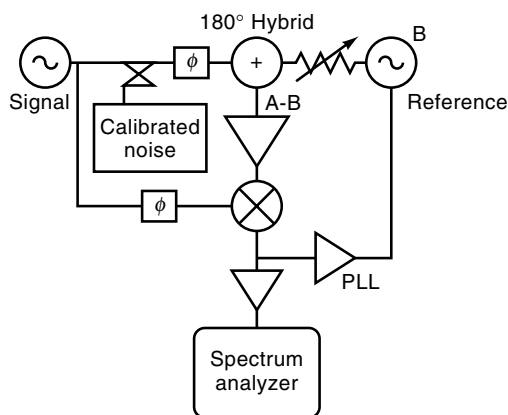


Figure 22. Two-oscillator PM noise measurement system with carrier suppression.

that the amount of suppression has not changed. The use of a PM/AM noise standard will ease the calibration process. In some specific systems, like AM measurement systems for amplifiers, carrier suppression systems are superior. The reason is that in these systems, the AM noise of an amplifier with lower or comparable noise to the source can be accurately measured. In cross-correlation measurement systems for amplifiers it is difficult to cancel the AM noise of the source, which is often comparable or larger than the noise in the amplifier under test.

BIBLIOGRAPHY

1. E. S. Ferre-Pikal et al., Draft revision of IEEE Std 1139-1988 standard definitions of physical quantities for fundamental frequency and time metrology—random instabilities, *Proc. 1997 IEEE Frequency Control Symp.* 1997, pp. 338–357.
2. D. W. Allan et al., Standard terminology for fundamental frequency and time metrology, *Proc. 42nd Annu. Symp. Frequency Control*, 1988, pp. 419–425.
3. J. A. Barnes et al., Characterization of frequency stability, *IEEE Trans. Instrum. Meas.*, **IM-20**: 105–120, 1971.
4. D. B. Sullivan et al., Characterization of clocks and oscillators, *National Institute of Standards and Technology, Technical Note No. 1337*, 1990.
5. D. A. Howe, D. W. Allan, and J. A. Barnes, Properties of signal sources and measurement methods, *Proc. 35th Annu. Symp. Frequency Control*, 1981, pp. A1–A47; also in Ref. 4.
6. T. E. Parker, Characteristics and sources of phase noise in stable oscillators, *Proc. 41st Annu. Symp. Frequency Control*, 1987, pp. 99–110.
7. F. L. Walls and A. DeMarchi, RF spectrum of a signal after frequency multiplication; measurement and comparison with a simple calculation, *IEEE Trans. Instrum. Meas.*, **IM-24**: 210–217, 1975.
8. F. L. Walls et al., Time-domain frequency stability calculated from the frequency domain: An update, *Proc. 4th Eur. Frequency Time Forum*, 1990, pp. 197–204.
9. J. A. Barnes and D. W. Allan, Variances based on data with dead time between the measurements, *NIST Tech. Note No. 1318*, 1990; also in Ref. 4.
10. L. M. Nelson and F. L. Walls, Environmental effects in mixers and frequency distribution systems, *Proc. IEEE Frequency Control Symp.*, 1993, pp. 831–837.

11. F. L. Walls et al., Extending the range and accuracy of phase noise measurements, *Proc. 42nd Annu. Symp. Frequency Control*, 1988, pp. 432–441; also in Ref. 4.
12. L. M. Nelson, C. W. Nelson, and F. L. Walls, Relationship of AM noise to PM noise in selected RF oscillators, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **41**: 680–684, 1994.
13. F. L. Walls and S. R. Stein, Servo techniques in oscillators and measurement systems, *National Bureau of Standards (US) Technical Note No. 692*, pp. 1–20, 1976.
14. F. L. Walls, Secondary standard for PM and AM noise at 5, 10 and 100 MHz, *IEEE Trans. Instrum. Meas.*, **IM-42**: 136–143, 1993.
15. F. L. Walls, Reducing errors, complexity, and measurement time for PM noise measurements, *Proc. 1993 Frequency Control Symp.*, 1993, pp. 81–86.
16. D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge, UK: Cambridge University Press, 1993.
17. B. N. Taylor and C. E. Kuyatt, *National Institute of Standards and Technology*, Technical Note No. 1297, 1993.
18. F. L. Walls, D. B. Percival, and W. R. Irelan, Biases and variances of several FFT spectral estimators as a function of noise type and number of samples, *Proc. 43rd Annu. Symp. Frequency Control*, 1989, pp. 336–341; also in Ref. 4.
19. W. F. Walls, Cross-correlation phase noise measurements, *Proc. 1992 IEEE Frequency Control Symp.*, 1992, pp. 257–261.
20. F. L. Walls et al., Precision phase noise metrology, *Proc. National Conf. Standards Laboratories (NCSL)*, 1991, pp. 257–275.
21. A. L. Lance, W. D. Seal, and F. Labaar, Phase noise and AM noise measurements in the frequency domain, *Infrared Millimeter Waves*, **11**: 239–289, 1984; also in Ref. 4.
22. Stanley J. Goldman, *Phase Noise Analysis in Radar Systems Using Personal Computers*, New York: Wiley, 1989, chap. 2, pp. 31–40.
23. K. H. Sann, Measurement of near-carrier noise in microwave amplifiers, *IEEE Trans. Microwave Theory and Techniques*, Vol. MTT-16, pp. 761–766, 1968.
24. G. J. Dick and J. Saunders, Method and apparatus for reducing microwave oscillator output noise, US Patent \$5,036,299, July 1991. See also D. G. Santiago and G. J. Dick, Microwave frequency discriminator with a cooled sapphire resonator for ultra-low phase noise, *Proc. 1992 IEEE Freq. Control Symp.*, pp. 176–182, 1992.
25. F. L. Walls, Suppressed carrier based PM and AM noise measurement techniques, *Proc. 1997 Intl. Freq. Control Symp.*, pp. 485–492, 1997.
26. C. McNeilage et al., Advanced phase detection technique for the real time measurement and reduction of noise in components and oscillators, *Proc. 1997 IEEE Intl. Freq. Control Symp.*, pp. 509–514, 1997.
27. F. Labaar, New discriminator boosts phase noise testing, *Microwaves*, Hayden Publishing Co., Inc., Rochelle Park, NJ, **21** (3): 65–69, 1982.

FRED L. WALLS
EVA S. FERRE-PIKAL
National Institute of Standards
and Technology

POWER AND ENERGY

POWER MEASUREMENT

Power measurement has a wide variety of applications. It is of importance primarily for the testing, monitoring, and maintenance of the energy supply network and of electrical equipment. It is also required in high-frequency and low-power circuits.

A wide variety of instruments and transducers are available for the measurement of power in ac and dc circuits. Analog instruments are available, but the recent trend has been toward digital instruments, because of their better performance and remote communication capabilities.

This article reviews the state and the trends of power measurement. The concept of electric power is first introduced; then common measurement methods and instruments adopted in different situations are discussed.

DEFINITION OF THE MEASURED QUANTITIES

Power can be defined as the time rate of energy transfer or energy dissipation in a load. It is dimensionally expressed as [J/s]. The electric power in a circuit, at any instant, equals the product of the current in the circuit and the voltage across its terminals at that instant. Dimensionally we have

$$[\text{J/s}] = [\text{J/C}][\text{C/s}] = [\text{V}][\text{A}] = [\text{W}] \quad (1)$$

Dc Circuits

In a dc circuit, if the voltage and current are constant, the electrical power P dissipated by a load (L) fed by a power supply (E) is the product of the voltage V_L across the load and the current I_L flowing in it:

$$P = V_L I_L \quad (2)$$

Ac Circuits

In the case of ac-fed circuits, the instantaneous rate of energy flow, or *instantaneous power*, is defined as the product of the instantaneous voltage drop across the load and the instantaneous current flowing through it:

$$p(t) = v(t)i(t) \quad (3)$$

It changes continuously, because of the cyclic variations in current and voltage. For this reason, its mean value over a defined time interval is generally considered.

Sinusoidal Circuits. The simplest case is that of a sinusoidal supply feeding a purely resistive load, where there is no phase difference between voltage $v(t)$ and current $i(t)$. The instantaneous power $p(t)$ is given by

$$p(t) = VI(1 - \cos 2\omega t) \quad (4)$$

in which V and I are the rms values of $v(t)$ and $i(t)$, respectively, and ω is the power supply angular frequency.

The instantaneous power is a unidirectional pulsating energy flow, given by a sum of two terms: a constant value VI ,

representing the average power, which gives the net energy transfer, and an oscillating component with a doubled angular frequency. In circuits fed by periodic ac voltages, it is then relevant to define the *average power* P dissipated in one period T (also called the *real* or *active power*):

$$P = \frac{1}{T} \int_0^T p(t) dt \quad (5)$$

In the case of purely reactive loads the phase difference between the voltage drop across the load and the current flowing through it is of 90° . The instantaneous power $p(t)$ is given by

$$p(t) = VI \cos 2\omega t \quad (6)$$

so there is no active power consumption in a reactive load. This is because energy is stored in the reactive load as electromagnetic or electrostatic energy for a certain time, and then returned to the power supply.

In the case of a sinusoidal supply, the resistive and purely reactive loads can be expressed by a real or a pure imaginary number, respectively. Since the load generally is not a pure element, it can be represented by a complex quantity, the impedance value. In this situation, the waveforms of the voltage $v(t)$, current $i(t)$, and instantaneous power $p(t)$ are shown in Fig. 1. The load impedance can also be represented by an equivalent circuit, such as a pure resistance R_{EQ} and a pure reactance X_{EQ} in series (Fig. 2). In this case the electrical power dissipated in the load Z_L can be expressed as the sum of the two power components dissipated on the resistance and reactance of the equivalent circuit. Because there is no active power dissipation in the reactance X_{EQ} , the average power is as follows:

$$P = V_{REQ} I_L = V_L I_L \cos \phi \quad (7)$$

The factor $\cos \phi$ appearing in Eq. (7) is referred as the *power factor*, because it yields the amount of power consumed as a fraction of $V_L I_L$. In fact, only a fraction of the voltage V_L contributes to the average power, because the component V_{XEQ} (the drop across the reactance) does not produce any active power dissipation, as it is orthogonal to the current I_L flowing into the load (Fig. 2).

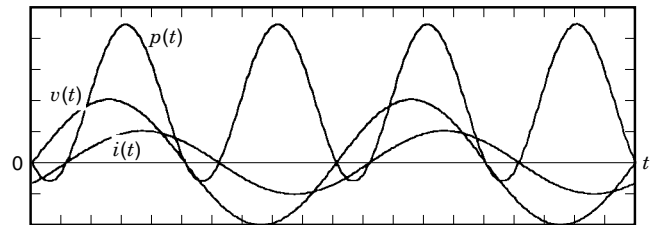


Figure 1. Waveforms of instantaneous power (p), voltage (v), and current (i).

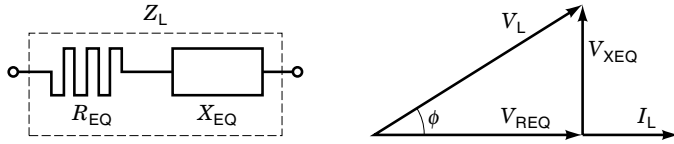


Figure 2. Voltage drop across the load and across its equivalent components.

The quantity

$$Q = V_{XEQ} I_L = V_L I_L \sin \phi \quad (8)$$

—called, by analogy, the *reactive power*—is introduced as a consequence of the voltage drop across a pure reactance. It is a term with zero mean value, representing an energy oscillating equally in both directions, and therefore does not give any contribution to the active power.

The product of the load voltage and current rms values,

$$P_A = V_L I_L \quad (9)$$

—called the *apparent power*—is a figure of merit representing the maximum energy transfer capability of a system and used to express the power consumption of various electrical devices.

Both apparent power and reactive power are dimensionally homogeneous with power and thus might be expressed in watt. But they have a different nature. While relationships involving instantaneous power and real power satisfy the principle of conservation of energy, apparent power and reactive power do not satisfy any conservation principles. To stress the different physical meaning of P_A and Q , their units are called voltampere (VA) and reactive voltampere (var), respectively.

The relationship existing between apparent power, active power and reactive power is given by

$$P_A = \sqrt{P^2 + Q^2} \quad (10)$$

Nonsinusoidal Circuits. A case of great interest is that of the power measurement on distorted waveforms. As an example, advanced systems for electric power production in industrial applications are commonly based on solid-state power components. They have proved to be extremely effective and useful, but, unfortunately, the waveforms that these devices produce on the power line have characteristics far different from the sinusoidal waveforms.

In this case, obviously, the above symbolic representation can no longer be applied. In the general case of distortion of both voltage and current waveforms, we have

$$\begin{aligned} v(t) &= \sum_{h=0}^M V_h \sin h\omega t \\ &= V_0 + V_1 \sin \omega t + V_2 \sin 2\omega t + V_3 \sin 3\omega t + \dots \\ i(t) &= \sum_{k=0}^M I_k \sin(k\omega t - \phi_k) \\ &= I_0 + I_1 \sin(\omega t - \phi_1) + I_2 \sin(2\omega t - \phi_2) \\ &\quad + I_3 \sin(3\omega t - \phi_3) + \dots \end{aligned} \quad (11)$$

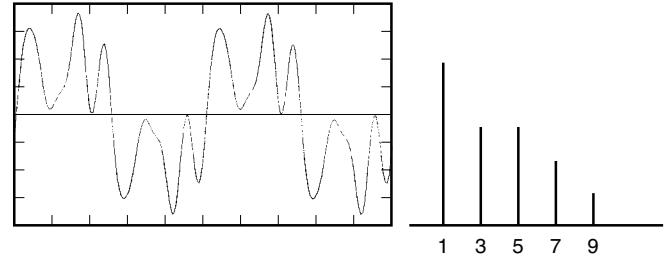


Figure 3. Voltage waveform in both time and frequency domains.

The active power, always defined as the mean power dissipated in one fundamental period, is then

$$\begin{aligned} P &= \frac{1}{T} \int_0^T \sum_{h=0}^M V_h \sin h\omega t \sum_{k=0}^M I_k \sin(k\omega t - \phi_k) dt \\ &= \frac{1}{T} \int_0^T \sum_{h=0}^M \sum_{k=0}^M V_h I_k \sin h\omega t \sin(k\omega t - \phi_k) dt \\ &= \frac{1}{T} \int_0^T \sum_{h=0}^M \sum_{k=0}^M V_h I_k \\ &\quad \times \frac{\cos[(h-k)\omega t + \phi_k] - \cos[(h+k)\omega t - \phi_k]}{2} dt \\ &= V_0 I_0 + \sum_{h=1}^M V_h I_h \cos \phi_h \\ &= V_0 I_0 + V_1 I_1 \cos \phi_1 + V_2 I_2 \cos \phi_2 + V_3 I_3 \cos \phi_3 + \dots \end{aligned} \quad (12)$$

—that is, the sum of the active powers of the various frequency components contained in the voltage and current. As an example, voltage and current waveforms in both time and frequency domains are shown in Figs. 3 and 4, respectively. Voltage components of order 1 (fundamental), 3, 5, 7, and 9 and current components of order 1, 5, and 11 are present. In this situation the active power is expressed as

$$P = V_1 I_1 \cos \phi_1 + V_5 I_5 \cos \phi_5$$

—that is, the actual power measurement is not affected by the 3rd, 5th, and 9th voltage harmonics or by the 11th current harmonic. In case of absence of distortion on either the voltage or the current, only the multiplication of the two fundamental waves must be performed.

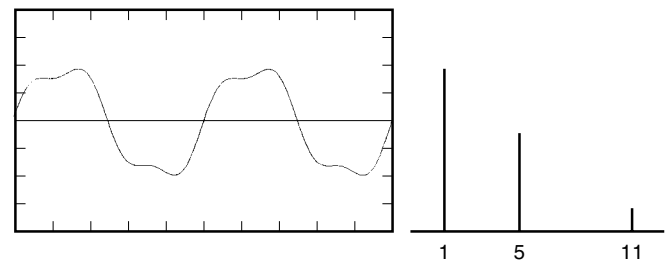


Figure 4. Current waveform in both time and frequency domains.

The apparent power of nonsinusoidal waves is the product of the load voltage and current rms values:

$$P_A = V_L I_L = \sqrt{\sum_{h=0}^M V_h^2 \sum_{k=0}^M I_k^2} = \sqrt{V_0^2 I_0^2 + \sum_{h=1}^M V_h^2 I_h^2 + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h^2 I_k^2} \quad (13)$$

where every voltage frequency component is separately combined with every current frequency component; the total apparent power then is not equal to the sum of the components related to the separate harmonics.

In the absence of the dc term, the difference $P_A^2 - P^2$ is

$$\begin{aligned} P_A^2 - P^2 &= \sum_{h=1}^M V_h^2 I_h^2 + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h^2 I_k^2 - \sum_{h=1}^M V_h^2 I_h^2 \cos^2 \Phi_h \\ &\quad - \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h I_h V_k I_k \cos \Phi_h \cos \Phi_k \\ &= \sum_{h=1}^M V_h^2 I_h^2 \sin^2 \Phi_h + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h^2 I_k^2 \\ &\quad - \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h I_h V_k I_k \frac{\cos(\Phi_h - \Phi_k) + \cos(\Phi_h + \Phi_k)}{2} \\ &= \sum_{h=1}^M V_h^2 I_h^2 \sin^2 \Phi_h + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h^2 I_k^2 \\ &\quad + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h I_h V_k I_k \\ &\quad \times \frac{\cos(\Phi_h - \Phi_k) - \cos(\Phi_h + \Phi_k) - 2 \cos(\Phi_h - \Phi_k)}{2} \\ &= \sum_{h=1}^M V_h^2 I_h^2 \sin^2 \Phi_h + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h^2 I_k^2 \\ &\quad + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M V_h I_h V_k I_k [\sin \Phi_h \sin \Phi_k - \cos(\Phi_h - \Phi_k)] \\ &= \left(\sum_{h=1}^M V_h I_h \sin \Phi_h \right)^2 \\ &\quad + \sum_{\substack{h=1 \\ (h \neq k)}}^M \sum_{k=1}^M [V_h^2 I_k^2 - V_h I_h V_k I_k \cos(\Phi_h - \Phi_k)] \quad (14) \end{aligned}$$

The first part of this relation contains terms involving the product of component of the same frequency; by analogy with the

sinusoidal case, they can be expressed as

$$Q = \sum_{h=1}^M V_h I_h \sin \Phi_h \quad (15)$$

and called reactive power. This relation has no physical meaning, but it is a convenient analytical component introduced for its duality with the physical relation Eq. (8). This mathematical decomposition can be given various interpretations. Many other analytical decompositions are possible.

The second part of the relation contains the summation of the products of current and voltage harmonics with different frequencies, expressed in terms of a further analytical expression

$$D = \sqrt{P_A^2 - (P^2 + Q^2)} \quad (16)$$

and called the *distortion power*.

While in sinusoidal conditions the reactive power is entirely related to the energy storage, and in a system their components can be summed algebraically, this is not true in nonsinusoidal conditions. In the case of voltage components of order 1, 3, 5 and current components of order 1, 5, and 11,

$$\begin{aligned} P_A^2 &= (V_1^2 + V_3^2 + V_5^2)(I_1^2 + I_5^2 + I_{11}^2) \\ P^2 &= V_1^2 I_1^2 \cos^2 \phi_1 + V_5^2 I_5^2 \cos^2 \phi_5 + 2V_1^2 I_1^2 V_5^2 I_5^2 \cos \phi_1 \cos \phi_5 \\ P_A^2 - P^2 &= V_1^2 I_1^2 + V_3^2 I_1^2 + V_5^2 I_1^2 + V_1^2 I_5^2 + V_3^2 I_5^2 + V_5^2 I_5^2 + V_1^2 I_{11}^2 \\ &\quad + V_3^2 I_{11}^2 + V_5^2 I_{11}^2 - V_1^2 I_1^2 \cos^2 \phi_1 - V_5^2 I_5^2 \cos^2 \phi_5 \\ &\quad - 2V_1^2 I_1^2 V_5^2 I_5^2 \cos \phi_1 \cos \phi_5 \\ &= V_1^2 I_1^2 \sin^2 \phi_1 + V_5^2 I_5^2 \sin^2 \phi_5 + V_3^2 I_1^2 + V_5^2 I_1^2 + V_1^2 I_5^2 \\ &\quad + V_3^2 I_5^2 + V_1^2 I_{11}^2 + V_3^2 I_{11}^2 + V_5^2 I_{11}^2 - 2V_1^2 I_1^2 V_5^2 I_5^2 \\ &\quad \times \cos \phi_1 \cos \phi_5 \end{aligned}$$

Voltage and current harmonics produce annoying effects in electrical power systems and also cause problems in industrial and communication apparatus. It is thus frequently necessary, or useful, to measure and analyze the power components in nonsinusoidal conditions.

High-Frequency Circuits. At low frequencies the power measurement is derived from voltage and current measurements. As the frequency increases, the wavelength become comparable to or lower than the circuit dimensions. In this case current and voltage become hard to measure, because they vary with the position along the transmission line, while the power remains constant. For this reason the power level is the most important factor in the design of high-frequency devices.

Power transmitted from a supply to a load at radio or microwave frequencies is generally measured by attaching a power sensor to the transmission-line port under measurement. The sensor output is connected to a power meter.

While an ideal sensor absorbs all the incident power (P_I), the actual impedance mismatch between the characteristic impedance of the radio frequency (RF) transmission line and the RF sensor input produces a partial reflection of the incoming power toward the supply. For this reason, the meter connected to the sensor does not account for the total amount of reflected power (P_R). The relationship among power dissipated

in the load (P_L), power incident, and power reflected is obviously $P_L = P_I - P_R$.

Directional couplers allow separate measurement of incident and reflected power by means of power meters applied, respectively, on the right and on the left side of the secondary waveguide. Directional couplers are also used to determine the reflection coefficient ρ of the sensor, which takes into account mismatch losses and is defined by

$$P_R = \rho^2 P_I \quad (17)$$

In order to take into account also the absorptive losses due to dissipation in the conducting walls of the sensor, leakage into instrumentation, power radiated into space, and so on, beside the reflection coefficient, the *effective efficiency* η_C of the sensor should also be considered. Generally, the reflection coefficient and effective efficiency are included in the *calibration factor* K , defined as

$$K = \eta_C(1 - \rho^2) \times 100\% \quad (18)$$

For instance, a calibration factor of 90% means that the meter will read 10% below the incident power. Generally calibration factors are specified by sensor manufacturers at different values of frequency.

GENERAL METHODS AND INSTRUMENTATION

As far as methods and instruments for power measurements are concerned, the problems involved change as the frequency of the power supply increases. Therefore, in the following, power measurement is discussed by considering general methods and instrumentation for dc and ac circuits. The ac circuits will be classified as (1) line-frequency circuits, (2) low- and medium-frequency circuits (up to a few megahertz), and (3) high-frequency circuits (up to a few gigahertz). Line-frequency circuits will be discussed separately from low-frequency circuits principally because of the existence of problems related specifically to three-phase power of the main.

Dc Circuits

A power measurement in a dc circuit can be generally carried out by using a voltmeter (V) and an ammeter (A), according to one of the connections shown in Fig. 5.

Both the connections introduce a systematic error, generally referred to as insertion error. In the connection of Fig. 5(a) the ammeter measures the current flowing into the voltmeter, I_V , as well as that flowing into the load, I_L . In the connection of Fig. 5(b) the voltmeter measures the voltage drop across the

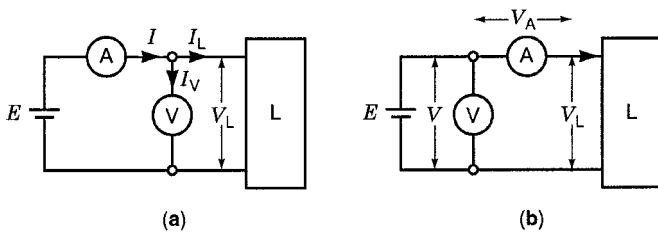


Figure 5. Two arrangements for dc power measurement circuits.

load, V_L , in addition to that dropping across the ammeter, V_A . So both connections give an excess of measured power, representing the power absorbed by the instrument connected closer to the load.

The analysis of the circuits of Fig. 5(a,b) provides the relations between the measurand, the electrical power P in the load, and the measured voltage V and current I , expressed as

$$P = V_L I_L = V(I - I_V), \quad P = V_L I_L = I(V - V_A) \quad (19)$$

By neglecting R_L (compared with the voltmeter internal resistance R_V) for the connection of Fig. 5(a) and the ammeter resistance R_A (compared with R_L) for the connection of Fig. 5(b), we have the respective approximations

$$I_V = I \frac{R_L}{R_V + R_L} \approx I \frac{R_L}{R_V}, \quad V_A = V \frac{R_A}{R_A + R_L} \approx V \frac{R_A}{R_L} \quad (20)$$

Consequently, the equations (19) become respectively

$$P = VI \left(1 - \frac{R_L}{R_V}\right) = VI \frac{R_V - R_L}{R_V} \quad (21)$$

$$P = VI \left(1 - \frac{R_A}{R_L}\right) = VI \frac{R_L - R_A}{R_L}$$

From Eqs. (20) and (21) analytical corrections of the insertion errors can be easily derived.

In a dc circuit the electrical power can also be measured by a wattmeter. The instrument most commonly used is the dynamometer. It is made of two fixed coils, positioned coaxially with space between them, and a moving coil, placed between the fixed coils and equipped with a pointer [Fig. 6(a)]. The fixed coils, commonly known as *current coils*, are connected in series with the load, to carry the load current. The moving coil (*voltage coil*), with a resistor R_V in series, is connected in parallel with the load R_L , to carry a current proportional to the voltage across it. When each coil carries a current, the interaction between the magnetic fields produced by the fixed and moving coils supplies a torque to move the pointer, which is proportional to the product of the two currents flowing into the coils. As a consequence, the deflection of the moving coil is proportional to the power dissipated in the load.

Insertion errors are also present in the dynamometer power measurement. In particular, on connecting the voltage coil

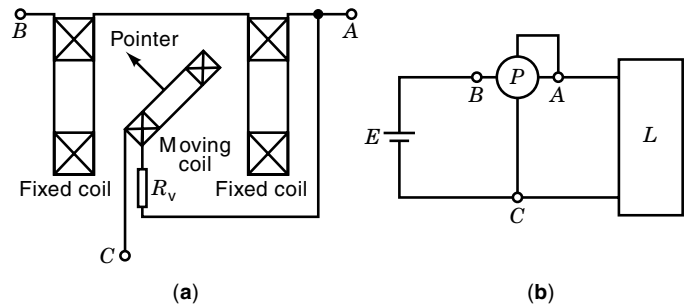


Figure 6. Power measurement with a dynamometer: (a) working principle, (b) measurement circuit.

between A and C [Fig. 6(b)], the current coils carry both the current through the load and that taken by the voltage coil. Consequently, the power P_L dissipated in the load can be obtained from the dynamometer reading P as

$$P_L = P - \frac{V^2}{R_V} \quad (22)$$

where $R_V = R_V + R_{VC}$ is the overall resistance of the voltage circuit (R_{VC} is the resistance of the voltage coil). By connecting the moving coil between B and C, this current error can be avoided, but now the voltage coil measures the excess voltage drop across the current coils. In this case the power in the load is

$$P_L = P - I^2 R_C \quad (23)$$

where R_C is the resistance of the current coil.

Ac Circuits

Active power in ac circuits can be measured by a dynamometer. Since its moving system cannot follow the rapid torque variations, it works as an intrinsic integrator of the instantaneous power, showing the average power according to Eq. (5). The insertion errors can be derived by simple considerations, in analogy with the previously discussed dc case, by replacing voltages and currents by the corresponding rms values.

However, a phase uncertainty due to the self-inductance of the voltage circuit arises in ac measurements. In sinusoidal conditions, if e_ω (in radians) is the phase defect angle (departure from exact quadrature) of the voltage circuit and $\cos \phi$ the load power factor, the relative uncertainty in active power measurements can be shown to be equal to $-e_\omega \tan \phi$. The phase uncertainty depends on the frequency. By using more complex circuits to compensate for this source of error, the frequency range of the dynamometer can be extended up to a few tens of kilohertz.

The reactive power measurement can be carried out by using special instruments (varmeters or reactive meters) with the same arrangements described for the active power. These are instruments whose response is proportional to $V_L I_L \sin \phi = V_L I_L \cos(90^\circ - \phi)$, where $\cos \phi$ is the load power factor. It can be measured with a wattmeter, by shifting the line voltage into quadrature.

In a polyphase system reactive power can be derived from the readings of wattmeters connected in a special way (Barbagelata's and Righi's insertions; see "Symmetrical Power Systems Supplying Unbalanced Loads" below). Also the apparent power, which is important for determining the line and supply capacity, can be measured by using a voltmeter and an ammeter. Electronic instruments, such as sampling wattmeters, can generally process the acquired data to supply all the different power quantities.

Line Frequency

For applications where the power is directly derived from the mains, the assumption of infinite power source can be reliably made and at least one of the two quantities voltage and current

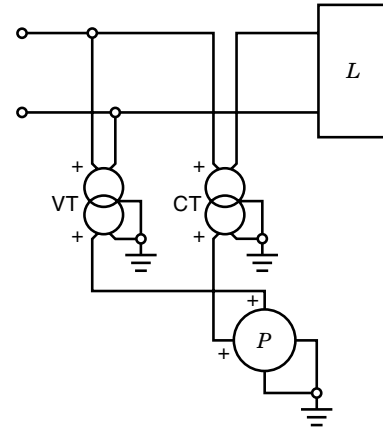


Figure 7. Single-phase power measurement with voltage (VT) and current (CT) transformers.

can be considered as sinusoidal. In this case only the power at the fundamental frequency should be examined (1).

Single-Phase Measurement. The power measurement at line frequency in single-phase circuits is generally carried out by a dynamometer, though the use of electronic instruments has become common. In practical applications, the case of a voltage greater than 1000 V is very common. Then the measurement must be carried out by using voltage and current transformers, inserted as shown in Fig. 7. In this case, the relative measurement uncertainty is equal to

$$\frac{\Delta P}{P} = (\eta_w + \eta_a + \eta_v) + (\epsilon_w + \epsilon_a + \epsilon_v) \tan \Phi_c \quad (24)$$

where η_w and ϵ_w are the instrumental and phase uncertainty of the wattmeter, η_a and η_v are the ratio uncertainties of the current (CT) and voltage (VT) transformers, and ϵ_a and ϵ_v are their phase uncertainties, respectively. If the load current exceeds the current range of the wattmeter, a current transformer has to be used, even in the case of low voltages.

Polyphase Power Measurement. In the following, the general case of power measurements on polyphase systems is analyzed. The common case of three-phase systems is treated, and different solutions are considered.

Measurements on Systems with Several Wires. Let a network with sinusoidal voltages and currents composed of n wires be considered. Among the currents flowing in such wires the following relation is established:

$$\sum_{i=1}^n I_i = 0 \quad (25)$$

The network can be thought as composed of $n-1$ single-phase independent systems, with the common return on one of the wires (say the s th wire). Then the total power can be measured as the sum of the readings of $n-1$ wattmeters, each one inserted with the current circuit in a different wire and the voltmeter

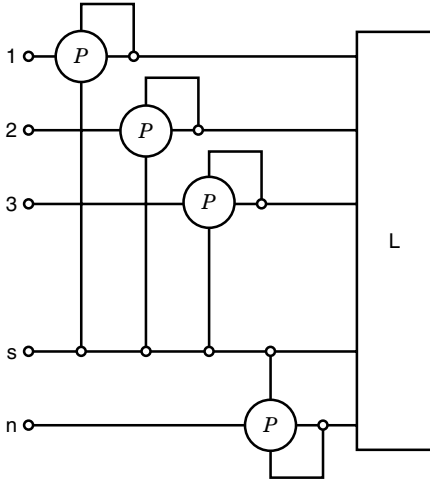


Figure 8. Power measurement on systems with several wires.

circuit between that wire and the s th one (Fig. 8):

$$P = \sum_{i \neq s}^n \dot{V}_{is} \dot{I}_i \quad (26)$$

The load power can also be measured by referring to a generic point O external to the network, as the sum of the readings of n wattmeters, each inserted with the ammeter circuit in a different wire and the voltmeter circuit connected between that wire and the point O :

$$P = \sum_{i=1}^n \dot{V}_{iO} \dot{I}_i \quad (27)$$

Power Measurements on Three-Wire Systems. The active power in a three-phase power system can generally be evaluated with three wattmeters connected as shown in Fig. 9. On each instrument, the current system is connected to carry the line current, while the voltmeter system is connected between the same wire and an artificial neutral point O , whose position is fixed by the voltmeter impedance of wattmeters or by suitable external impedances.

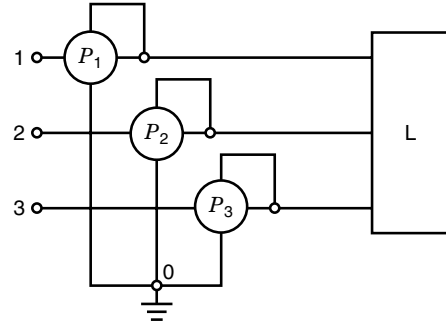


Figure 9. Insertion of three wattmeters for three-wire three-phase systems.

In these conditions the power in the load is the sum of the three wattmeters' indications:

$$P = \sum_{i=1}^3 \dot{V}_{iO} \dot{I}_i \quad (28)$$

If the three-phase system is provided by four wires (three phases with a neutral wire), the neutral wire is utilized as the common wire.

Analogously to the single-phase case, for medium-voltage circuits the three-wattmeter insertion is modified as in Fig. 10.

Symmetrical and Balanced Systems. The supply system is symmetrical and the three-phase load is balanced if

$$\begin{aligned} V_1 &= V_2 = V_3 \\ I_1 &= I_2 = I_3 \end{aligned} \quad (29)$$

In Fig. 11 the three possible kinds of insertion of an instrument S (active or reactive power meter) are illustrated, by using the following convention: S_{ijk} indicates a reading performed with the current leads connected to line i and the voltmeter leads connected between phases j and k . If i is equal to j , one is omitted, for example, P_{12} [Fig. 11(b)]. The active power related to a single phase is usually referred to as P_1 .

In the first case [Fig. 11(a)] S is inserted between the phase and the artificial point O . By Eq. (29), the overall active power is

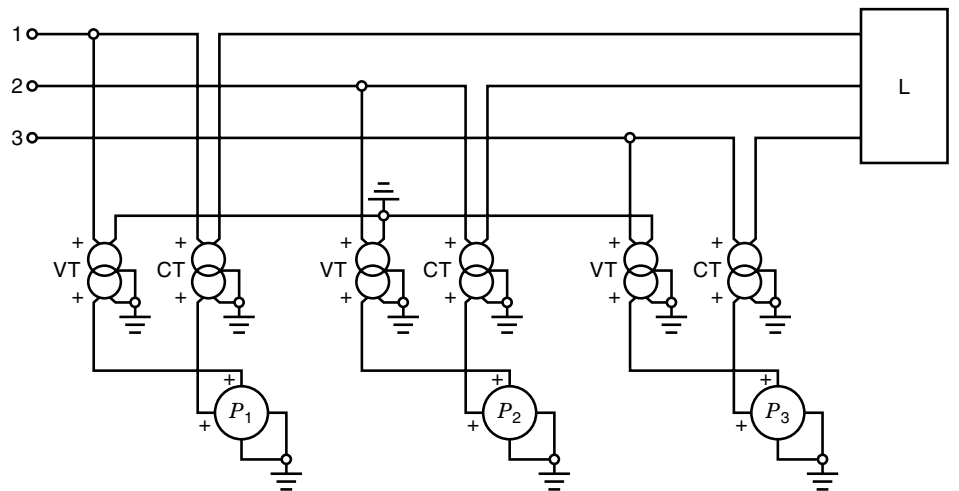


Figure 10. Insertion of three wattmeters for medium voltage.

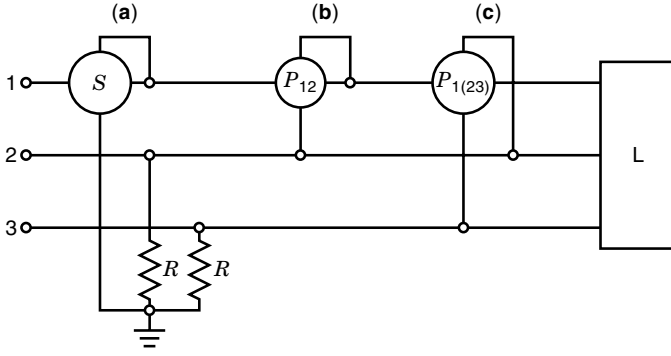


Figure 11. The three modes of insertion of a power meter.

given by three times the indication of the wattmeter S ; similarly for the overall reactive power if S is a reactive power meter. Notice that a pair of twin resistors with the same resistance R as that of the voltage circuit of S are placed on the other phases in order to balance the load.

The wattmeter reading corresponding to Fig. 11(c) can be expressed as

$$P_{1(23)} = I_1 \dot{V}_{23} \quad (30)$$

But $CIV_{12} + CIV_{23} + CIV_{31} = 0$ and $CIV_{13} = -CIV_{31}$; thus

$$P_{1(23)} = I_1(-\dot{V}_{12} - \dot{V}_{31}) = -I_1 \dot{V}_{12} + I_1 \dot{V}_{13}$$

where the last two terms are $I_1 CIV_{12} = P_{12}$ and $I_1 CIV_{13} = P_{13}$, which lead to the final relation, and to corresponding relations for the other phases, valid for any kind of supply and load:

$$\begin{aligned} P_{1(23)} &= P_{13} - P_{12} \\ P_{2(31)} &= P_{21} - P_{23} \\ P_{3(12)} &= P_{32} - P_{31} \end{aligned} \quad (31)$$

If the supply system is symmetrical, we have $P_{1(23)} = \sqrt{3} Q_1$. In fact, starting from the relationship

$$P_{1(23)} = I_1 \dot{V}_{23} = I_1 V_{23} \cos \beta \quad (32)$$

where $\beta = 90^\circ - \phi_1$ (Fig. 12), we have

$$P_{1(23)} = \sqrt{3} E_1 I_1 \sin \phi_1 = \sqrt{3} Q_1$$

with Q_1 the reactive power involved in phase 1. The three final relationships are

$$\begin{aligned} P_{1(23)} &= \sqrt{3} Q_1 = P_{13} - P_{12} \\ P_{2(31)} &= \sqrt{3} Q_2 = P_{21} - P_{23} \\ P_{3(12)} &= \sqrt{3} Q_3 = P_{32} - P_{31} \end{aligned} \quad (33)$$

Hence, in the case of symmetrical and balanced systems, the overall reactive power is given by

$$Q = 3Q_1 = 3P_{1(23)}/\sqrt{3} = \sqrt{3}P_{1(23)} \quad (34)$$

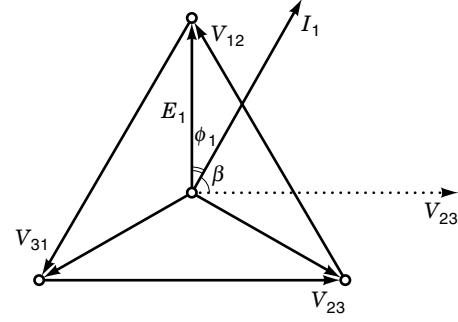


Figure 12. Phasor diagram for a three-phase symmetrical and balanced system.

Power Measurements Using Two Wattmeters. The overall active power in a three-wire system can be measured by using only two wattmeters. In fact, Aron's theorem states the following relationships:

$$\begin{aligned} P &= P_{12} + P_{32} \\ P &= P_{23} + P_{13} \\ P &= P_{31} + P_{21} \end{aligned} \quad (35)$$

Analogously, the overall reactive power can be measured by using only two reactive power meters:

$$\begin{aligned} Q &= Q_{12} + Q_{32} \\ Q &= Q_{23} + Q_{13} \\ Q &= Q_{31} + Q_{21} \end{aligned} \quad (36)$$

To prove one of the previous statements, say $P = P_{12} + P_{32}$, refer to Fig. 13, where the two wattmeters furnish P_{12} and P_{32} . Hence the sum of the two readings gives

$$\begin{aligned} P_{12} + P_{32} &= I_1 \dot{V}_{12} + I_3 \dot{V}_{32} = I_1(E_1 - E_2) + I_3(E_3 - E_2) \\ &= I_1 E_1 - I_1 E_2 + I_3 E_3 - I_3 E_2 \\ &= I_1 E_1 + I_3 E_3 - (I_1 + I_3) E_2 \\ &= I_1 E_1 + I_3 E_3 + I_2 E_2 = P_1 + P_2 + P_3 = P \end{aligned} \quad (37)$$

Provided that the system has only three wires, Aron's theorem applies to any kind of supply and load. In the case of symmetrical and balanced systems, it also allows the reactive power to

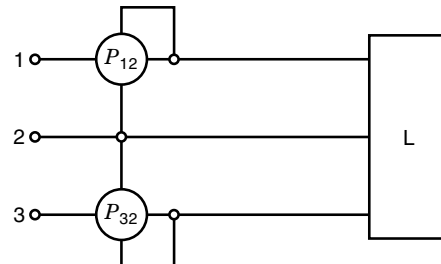


Figure 13. Power measurements using two wattmeters.

be evaluated as

$$Q = \sqrt{3}(P_{32} - P_{12}) \quad (38)$$

By using Eqs. (35) and (38) the power factor is seen to be

$$\begin{aligned} \cos \phi &= \frac{P_{12} + P_{32}}{\sqrt{(P_{12} + P_{32})^2 + 3(P_{32} - P_{12})^2}} \\ &= \frac{P_{12} + P_{32}}{\sqrt{4P_{12}^2 + 4P_{32}^2 - 4P_{12}P_{32}}} \\ &= \frac{1 + (P_{12}/P_{32})}{2\sqrt{(P_{12}/P_{32})^2 - (P_{12}/P_{32}) + 1}} \end{aligned} \quad (39)$$

Aron's insertion presents a limitation for low power factors. In fact, considering the functions

$$\begin{aligned} \cos(\phi + 30^\circ) &= \frac{P_{12}}{VI} \\ \cos(\phi - 30^\circ) &= \frac{P_{32}}{VI} \end{aligned} \quad (40)$$

(Fig. 14), it can be argued that for $\phi \leq 60^\circ$, P_{12} and P_{32} are both greater than zero, while for $\phi > 60^\circ$, $\cos(\phi - 30^\circ)$ is still greater than zero and $\cos(\phi + 30^\circ)$ is less than zero. The absolute error in the active power is

$$\Delta P = \frac{\partial(P_{12} + P_{32})}{\partial P_{12}} \Delta P_{12} + \frac{\partial(P_{12} + P_{32})}{\partial P_{32}} \Delta P_{32} = \Delta P_{12} + \Delta P_{32} \quad (41)$$

This corresponding relative error is greatest when P_{12} and P_{32} have values close to each other and opposite in sign; in particular, for $\cos \phi = 0$ ($\phi = 90^\circ$) the error is infinite.

If η_w and ϵ_w are the wattmeter amplitude and phase errors, the error in the active power is

$$\frac{\Delta P}{P} = \frac{(\eta_w + \epsilon_w \tan \phi_{12})P_{12} + (\eta_w + \epsilon_w \tan \phi_{32})P_{32}}{P_{12} + P_{32}} = \eta_w + \epsilon_w \frac{Q}{P} \quad (42)$$

Let two wattmeters with nominal values V_0 , I_0 , $\cos \phi_0$ and accuracy class c be considered. The maximum absolute error in

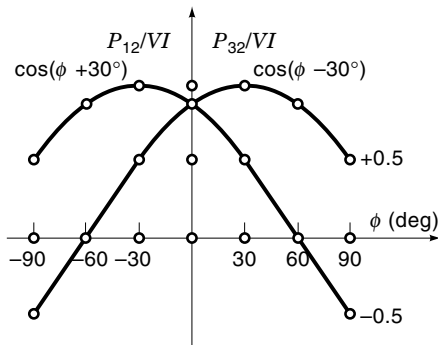


Figure 14. Sign of powers in Aron's insertion.

each reading is

$$\Delta P = \frac{c V_0 I_0 \cos \phi_0}{100} \quad (43)$$

Therefore, the percentage error related to the sum of the two indications is

$$\frac{\Delta P}{P} = \frac{c V_0 I_0 \cos \phi_0}{\sqrt{3} V I \cos \phi} = 1.11 \frac{c V_0 I_0 \cos \phi_0}{V I \cos \phi} \quad (44)$$

equal approximately to the error of only one wattmeter inserted in a single-phase circuit with the same values of I , V , and $\cos \phi$. Consequently, under the same conditions, the use of two wattmeters involves a measurement uncertainty rather lower than the use of three wattmeters. If the Aron's insertion is performed via current and voltage transformers, characterized by ratio errors η_a and η_v and phase errors ϵ_a and ϵ_v respectively, the active power error is

$$\begin{aligned} \frac{\Delta P}{P} &= \frac{(\eta_{TOT} + \epsilon_{TOT} \tan \phi_{12})P_{12} + (\eta_{TOT} + \epsilon_{TOT} \tan \phi_{32})P_{32}}{P_{12} + P_{32}} \\ &= \eta_{TOT} + \epsilon_{TOT} \frac{Q}{P} = \eta_{TOT} + \epsilon_{TOT} \tan \Phi_c \end{aligned} \quad (45)$$

where $\cos \phi_c$ is the conventional power factor, and

$$\eta_{TOT} = \eta_w + \eta_a + \eta_v$$

$$\epsilon_{TOT} = \epsilon_w + \epsilon_a + \epsilon_v$$

the error sums, with η_w , ϵ_w the wattmeter errors.

Symmetrical Power Systems Supplying Unbalanced Loads. If the load is unbalanced, the current amplitudes are different from each other and their relative phase is not equal to 120° . In this situation two wattmeters and one voltmeter have to be connected as proposed by Barbagelata (1) (Fig. 15). The first wattmeter gives P_{12} and P_{13} (changing the switch position), and the second one gives P_{31} and P_{32} .

From Aron's theorem the active power is

$$P = P_{12} + P_{32} \quad (46)$$

and then the reactive power Q is

$$Q = Q_1 + Q_2 + Q_3 = \frac{1}{\sqrt{3}}(P_{13} - P_{12} + \underline{P_{21}} - \underline{P_{23}} + P_{32} - P_{31}) \quad (47)$$

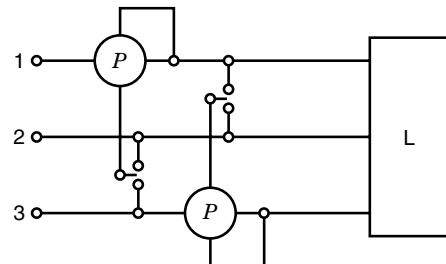


Figure 15. Barbagelata's insertion for symmetrical and unbalanced systems.

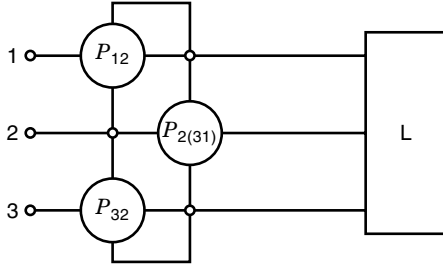


Figure 16. Righi's insertion for symmetrical and unbalanced systems.

For the underlined terms, from Aron's theorem it follows that

$$P = P_{13} + P_{23} = P_{12} + P_{32} = P_{21} + P_{31} \quad (48)$$

Then

$$P_{13} + P_{23} = P_{21} + P_{31} \quad \text{so} \quad P_{21} - P_{23} = P_{13} - P_{31}$$

Then we have

$$Q = \frac{1}{\sqrt{3}} [2(P_{13} - P_{31}) + P_{32} - P_{12}] \quad (49)$$

Hence, from only four power measurements, the overall active and reactive powers can be obtained.

The main disadvantage of this method is that the four measurements are not simultaneous; therefore, any load variations during the measurement will cause a loss in accuracy. In this case, a variation proposed by Righi (1) may be used, with three wattmeters connected as shown in Fig. 16 to measure P_{12} , P_{32} , and $P_{2(31)}$ simultaneously. The reactive power is

$$Q = \frac{1}{\sqrt{3}} (P_{13} - P_{12} + P_{21} - P_{23} + P_{32} - P_{31}) \quad (50)$$

As above, from Aron's theorem it follows that

$$P_{21} - P_{23} = P_{13} - P_{31} \quad \text{and} \quad P_{2(31)} = P_{21} - P_{23} = P_{13} - P_{31}$$

Then

$$Q = \frac{1}{\sqrt{3}} (P_{32} - P_{12} + 2P_{2(31)}) \quad (51)$$

For symmetrical and unbalanced systems another two-wattmeter insertion can be carried out (Fig. 17). The wattmeters give

$$\begin{aligned} P_{1(30)} &= E_3 I_1 = j \frac{\dot{V}_{12}}{\sqrt{3}} I_1 = -\frac{Q_{12}}{\sqrt{3}} \\ P_{3(10)} &= E_1 I_3 = j \frac{\dot{V}_{23}}{\sqrt{3}} I_3 = \frac{Q_{32}}{\sqrt{3}} \end{aligned} \quad (52)$$

Hence, the overall reactive power is

$$Q = Q_{12} + Q_{32} = \sqrt{3}(-P_{1(30)} + P_{3(10)}) \quad (53)$$

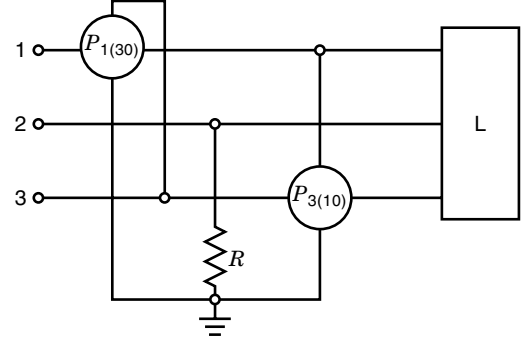


Figure 17. Insertion of two wattmeters for symmetrical and unbalanced systems.

Method Selection Guide. For three-wire systems, the flux diagram of Fig. 18 leads to selecting the most suitable method according to system characteristics.

Low- and Medium-Frequency Power Measurements

In the following, the main methods and instruments for power measurements at low and medium frequencies are considered.

Three-Voltmeter Method. The power dissipation in the load L can be measured by using a non-inductive resistor R and three voltmeters, as shown in Fig. 19 (2). Though one of the voltmeters may appear redundant on a first analysis of the circuit, in actual fact three independent data are needed in order to derive the power from Eq. (7). In particular, from the voltage drops v_{AB} and v_{BC} , the load current and load voltage can be directly derived; then, v_{AC} is used to retrieve information about their relative phase. If the currents drawn by the voltmeters are neglected and the current i_L flowing into the load L is assumed to be equal to that flowing into the resistor R, the statement can be demonstrated as follows:

$$\begin{aligned} v_{AC} &= Ri_L + v_L \\ v_{AC}^2 &= R^2 i_L^2 + v_L^2 + 2Rv_L i_L \end{aligned} \quad (54)$$

where the lowercase letters indicate instantaneous values. By computing rms values (indicated with capital letters) we obtain the power P_L :

$$\begin{aligned} \frac{1}{T} \int_0^T v_{AC}^2 dt &= \frac{1}{T} \int_0^T R^2 i_L^2 dt + \frac{1}{T} \int_0^T v_L^2 dt + \frac{1}{T} \int_0^T 2Rv_L i_L dt \\ V_{AC}^2 &= R^2 I_L^2 + V_L^2 + 2RP_L \\ P_L &= \frac{V_{AC}^2 - R^2 I_L^2 - V_L^2}{2R} = \frac{V_{AC}^2 - V_{AB}^2 - V_{BC}^2}{2R} \end{aligned} \quad (55)$$

The expression (55) is also applicable to dc on replacing rms values with dc ones. Since the result is obtained as a difference, problems arise from relative uncertainty when the three terms have a sum near zero.

Such a method is still used for high-accuracy applications.

Thermal Wattmeters. The working principle of thermal wattmeters is based on a pair of twin thermocouples whose output

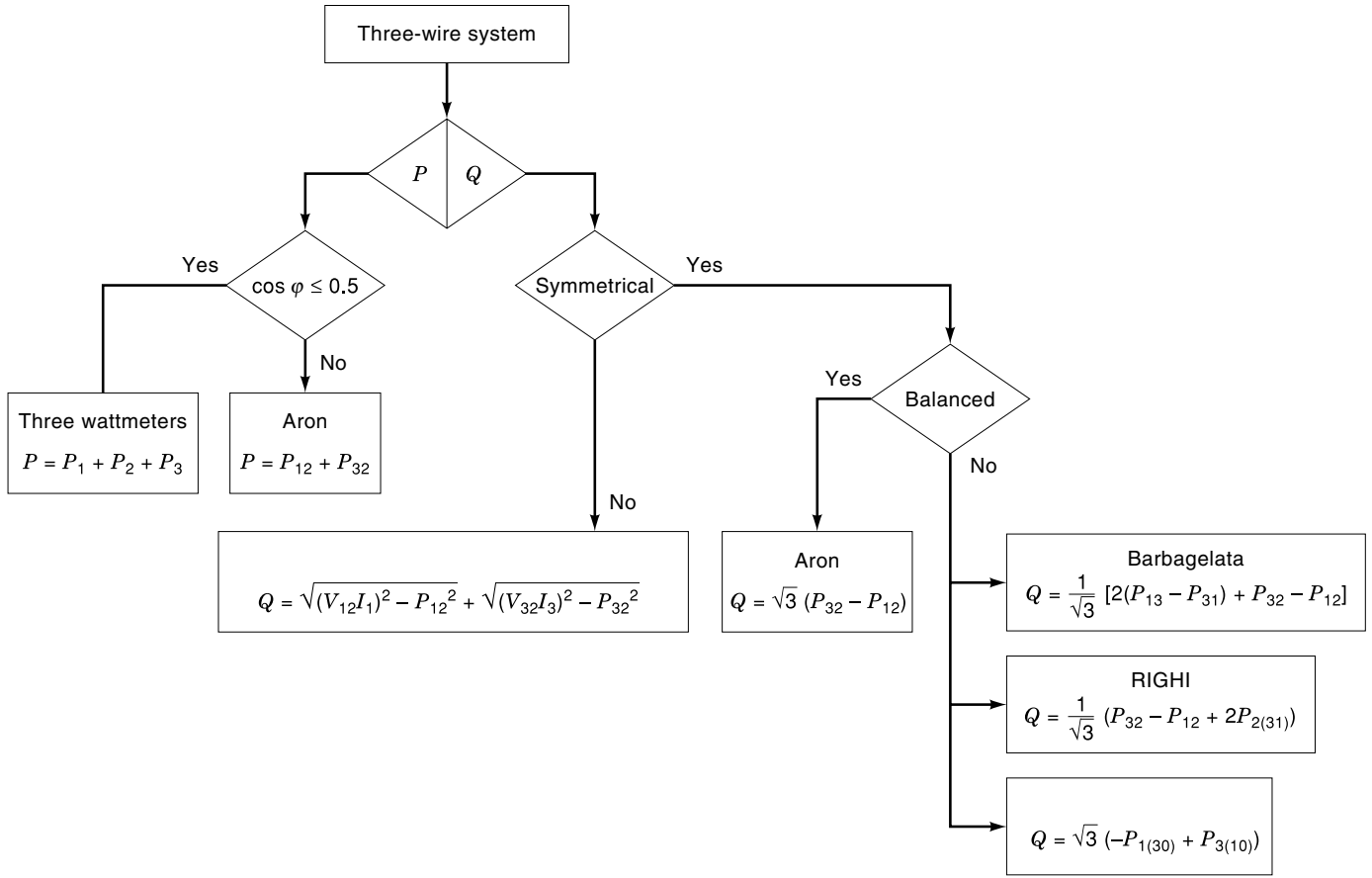


Figure 18. Method selection guide for power measurements on three-wire systems.

voltage is proportional to the square of the rms value of the currents flowing into the thermocouple heaters (3).

The essential circuit of a thermal wattmeter is shown in Fig. 20(a). Without the load, if the two heaters have equal resistance ($r_1 = r_2 = r$), in the limit $S \ll r$ they are connected in parallel and carry the same current i_p , proportional to the supply voltage:

$$i_1 = i_2 = i_p = \frac{v}{2R + r} \quad (56)$$

In this case, the output voltages of the two thermocouples turn out to be equal ($e_1 = e_2$), thus the voltage Δe measured by the voltmeter is null. In Fig. 20(b) this situation is represented by

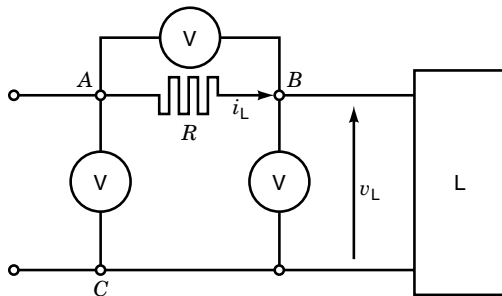


Figure 19. Three-voltmeter method.

the working point T , the same for both thermocouples. By applying a load L with a corresponding current i_L , a voltage drop across S arises, determining an imbalance between currents i_1 and i_2 (i_1 increases and i_2 decreases). In the limit $r \ll R$, the two heaters are in series; thus the current imbalance i_u through them is proportional to the load current:

$$i_u = i_1 - i_2 = \frac{S}{2R} i_L \quad (57)$$

Therefore, the working points of the two thermocouples change: to A for the thermocouple that carries the current i_1 , and to B for the other one [Fig. 20(b)].

Because the thermocouple response time is higher than the period of the heater current, it produces an emf proportional to the average squared value of this current. In this situation, in the above limits, the voltmeter measures a voltage imbalance Δe proportional to the active power absorbed by the load (except for the excess given by the powers dissipated in R , S , r_1 , and r_2):

$$\begin{aligned} \Delta e &= k(\overline{i_1^2} - \overline{i_2^2}) = k[\overline{(i_p + i_L)^2} - \overline{(i_p - i_L)^2}] = k \\ \overline{4i_p i_L} &= k_1 \overline{v(t) i(t)} = k_1 P \end{aligned} \quad (58)$$

where the overbars indicate time averages.

If the two thermocouples cannot be considered as matched and linear, the power measurement accuracy will obviously be

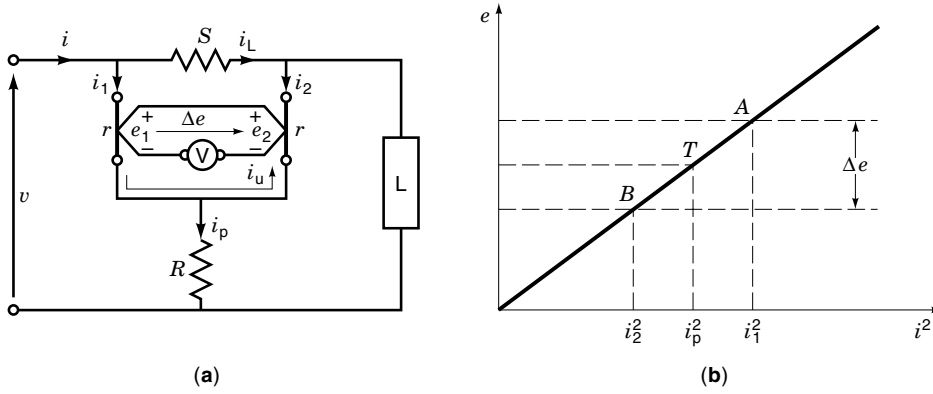


Figure 20. (a) Thermal wattmeter based on twin thermocouples; (b) working characteristic in ideal conditions.

compromised. This situation is shown in Fig. 21, where the two thermocouples are supposed to have two quite different nonlinear characteristics. In this case, the voltage measured by the voltmeter will be Δe_n instead of Δe .

Wattmeters based on the thermal principle allow high accuracy to be achieved in critical cases of highly distorted wideband signals.

Wattmeters Based on Multipliers. The multiplication and averaging processes involved in power measurements can be undertaken by electronic circuits (Fig. 22). Wattmeters of this kind fall into two categories, depending on whether the multiplication and averaging operations are performed in a continuous or a discrete way. In the first case multiplication is usually carried out by means of an analog electronic multiplier, while averaging is carried out by a low-pass filter or by an averaging reading instrument. In the other case (sampling wattmeters), simultaneous samples of voltage and current waveforms are taken, digitized, and then multiplied and averaged by using digital techniques.

In multiplier methods also, the resistances of the voltage and current circuits have to be taken into account, analogously to the case of dynamometers [see Eqs. (22) and (23)]. Furthermore, phase errors of both current circuits ($e_{\omega c}$) and voltage circuits ($e_{\omega v}$) increase the relative uncertainty of power measurement [in the case of sinusoidal conditions by $(e_{\omega c} - e_{\omega v}) \tan \phi$].

Wattmeters Based on Analog Multipliers. The main analog multipliers are based on a transistor circuit, such as a four-

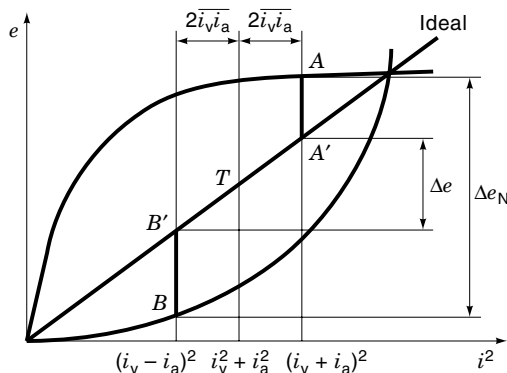


Figure 21. Ideal and actual characteristics of thermal-wattmeter thermocouples.

quadrant multiplier (4), which is fed by voltage and current signals to give the instantaneous power, and an integrator to provide the mean power (Fig. 23). More effective solutions are based on the time-division multiplier (TDM) and on Hall-effect-based multipliers.

TDM-Based Wattmeters. The block diagram of a wattmeter based on a TDM is shown in Fig. 24 (5). A square wave v_m [Fig. 25(a)] with constant period T_g , and with duty cycle and amplitude determined by $i(t)$ and $v(t)$, respectively, is generated. T_g is much smaller than the period of the measurands $v_x(t)$ and $v_i(t)$, so that one can consider these voltages constant during that time interval.

The duty cycle of v_m is set by the impulse duration modulator circuit (Fig. 24). The ramp voltage $v_g(t)$ [Fig. 25(b)], compared with the voltage $v_i(t)$ proportional to $i(t)$, produces a square-wave output, which drives the switch SW, fixing a time interval t_2 of the output $v_m(t)$ proportional to $v_i(t)$. If

$$v_g(t) = \frac{4V_{g0}}{T_g} t \quad \text{when } 0 \leq t \leq \frac{T_g}{4} \quad (59)$$

from simple geometrical considerations we obtain

$$t_2 = 2 \left(\frac{T_g}{4} - \frac{T_g}{4} \frac{v_i}{V_{g0}} \right) \quad (60)$$

and

$$t_1 - t_2 = \frac{T_g}{V_{g0}} v_i \quad (61)$$

The amplitude of $v_m(t)$ is set by the impulse amplitude

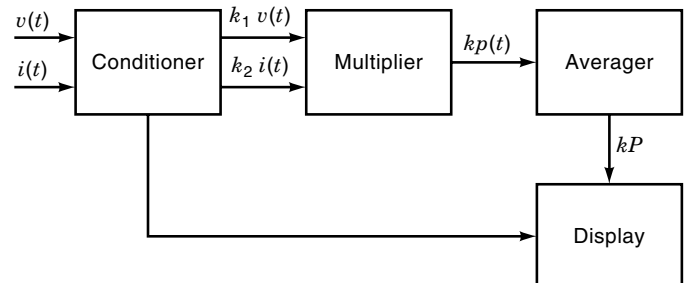


Figure 22. Block diagram of a multiplier-based wattmeter.

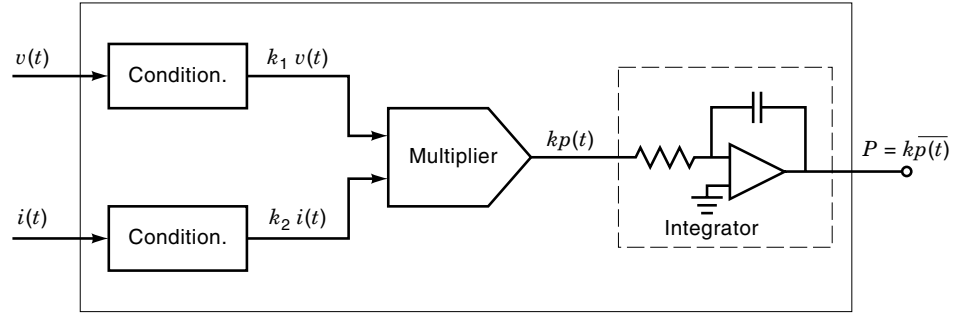


Figure 23. Block diagram of a four-quadrant multiplier-based wattmeter.

modulator circuit to be equal to $+v_x$ during the time interval t_1 , and to $-v_x$ during the time interval t_2 [Fig. 25(a)].

After an initial transient, the output voltage $v_{out}(t)$ of the low-pass filter (integrator) is the mean value of $v_m(t)$:

$$\begin{aligned} V_{out} &= \frac{1}{RC} \int_0^t v_m(t) dt = K' \left(\int_0^{t_1} v_x(t) dt - \int_{t_1}^{t_1+t_2} v_x(t) dt \right) \\ &= K' v_x(t_1 - t_2) = K v_x v_i \end{aligned} \quad (62)$$

The frequency range is generally between 200 Hz and 20 kHz, but can reach 100 kHz. Uncertainties are typically 0.01% to 0.02% (6).

Hall-Effect Based Wattmeters. As is known, in a Hall-effect transducer the output voltage $v_H(t)$ is proportional to the product of two time-dependent quantities, the current $i(t)$ through the transducer and the magnetic induction $B(t)$ (7):

$$v_H(t) = R_H i(t) B(t) \quad (63)$$

where R_H is the Hall constant. In the circuit of Fig. 26(a), the power P is determined by measuring $v_H(t)$ through a high-input-impedance averaging voltmeter, and by considering that $v_x(t) = ai(t)$ and $i_x(t) = bB(t)$, with a and b proportionality factors:

$$P = \frac{1}{T} \int_0^T v_x(t) i_x(t) dt = ab \frac{1}{T} \int_0^T i(t) B(t) dt = ab \frac{V_H}{R_H} \quad (64)$$

where T is the measurand period, and V_H the mean value of $v_H(t)$.

In the usual realization of the Hall multiplier (accuracy 0.1% up to a few megahertz), shown in Fig. 26(a), the magnetic induction is proportional to the load current i_L and the optimal polarizing current i_v is set by the resistor R_v . For frequencies over the megahertz range, an alternative arrangement is shown in Fig. 26(b), in which the load current i_L , or a suitable amount of it reduced by inserting the shunt R_s , flows directly into the Hall device, acting as a polarizing current, and the magnetic field is generated by the voltage v_x . In this way, the influence of temperature is reduced for line-frequency applications with constant-amplitude voltages and variable load currents.

Standard wattmeters for high frequency, use probes in waveguides with rectifiers.

Wattmeters Based on Digital Multipliers

Sampling Wattmeters. The most important wattmeter operating on discrete samples is the sampling wattmeter (Fig. 27). It is essentially composed of two input channels, each constituted by a conditioner (C), a sample-and-hold (S/H), and an analog-to-digital converter (ADC). Digital multiplier (MUL), summing (SUM), dividing (DIV), and displaying (DISP) units are also included. The architecture is handled by a processing unit, not shown in the figure.

The uniform sampling of input signals allows the active power to be evaluated as the mean of the sequence of

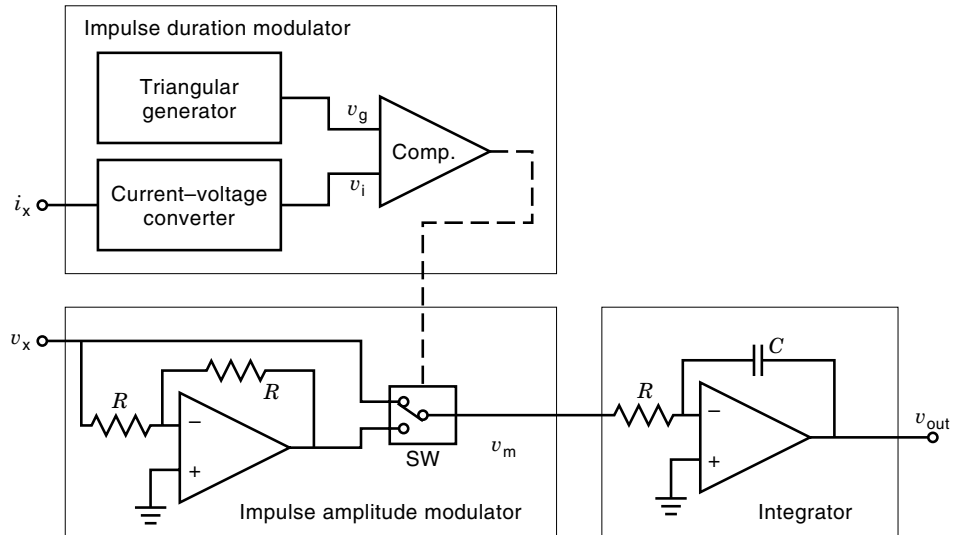


Figure 24. Block diagram of a TDM-based wattmeter.

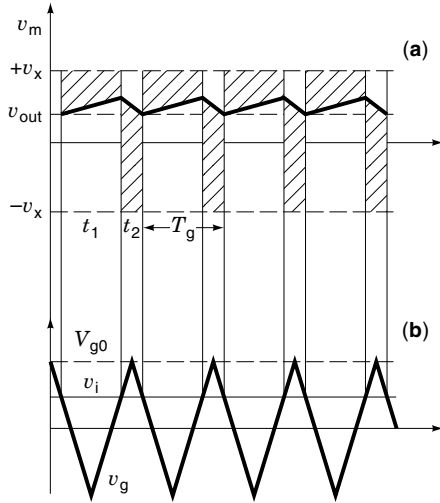


Figure 25. Waveform of TDM-based power measurement: (a) impulse amplitude-modulator output, (b) ramp-generator output.

instantaneous power samples $p_j(k)$:

$$P = \frac{1}{N} \sum_{k=0}^{N-1} p_j(k) = \frac{1}{N} \sum_{k=0}^{N-1} v_j(k) i_j(k) \quad (65)$$

where N is the number of samples in one period of the input signal. A previous estimation of the measurand fundamental period is made to adjust the summation interval of Eq. (65) and/or the sampling period in order to carry out synchronous sampling (8). The sampling period can be adjusted by using a frequency multiplier with a PLL circuit driven by the input signal (9). Alternatively, the contribution of the sampling error is reduced by averaging over a large number of periods of the input signal.

In the time domain, estimation of the period of highly distorted signals, as in pulse width modulation (PWM), is made difficult by the numerous zero crossings present in the waveform. Some types of digital filters can be used for this purpose. An efficient digital way to estimate the period is the discrete integration of the PWM signal. In this way, the period of the

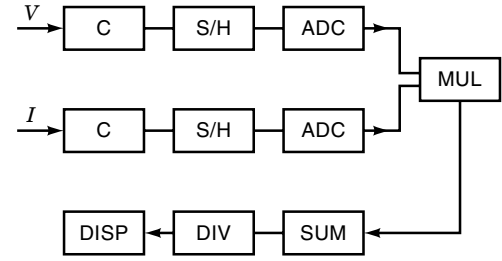


Figure 27. Block diagram of the sampling wattmeter.

fundamental harmonic is estimated by detecting the sign changes of the cumulative sum function (10):

$$S(k) = \sum_{i=1}^k v_i, \quad k = 1, 2, \dots, N \quad (66)$$

If the summation interval is extended to an integer number of periods of the function $S(k)$, then *quasisynchronous* sampling (11) is achieved through a few simple operations (cumulative summation and sign detection), and the maximum synchronization error is limited to a sampling period. Through relatively small increases in computational complexity and in memory size, the residual error can be further reduced through a suitable data-processing algorithm, that is, the multiple convolution in the time domain of triangular windows (10). Hence, the power measurement can be obtained as

$$P_{(B)} = \frac{1}{\sum_{k=0}^{2B(N-1)} w(k)} \sum_{k=0}^{2B(N-1)} w(k) p(k) \quad (67)$$

where $p(k)$ is the k th sample of the instantaneous power and $w(k)$ the k th weight corresponding to the window obtained as the convolution of B triangular windows (11).

Another way to obtain the mean power is through the consideration of the harmonic components of voltages and currents in the frequency domain by using the discrete Fourier transform (12). In particular, a fast Fourier transform algorithm is used in order to improve the efficiency. A two-step search for the harmonic peaks is then carried out: (1) the indices of the frequency samples corresponding to the highest spectral peaks

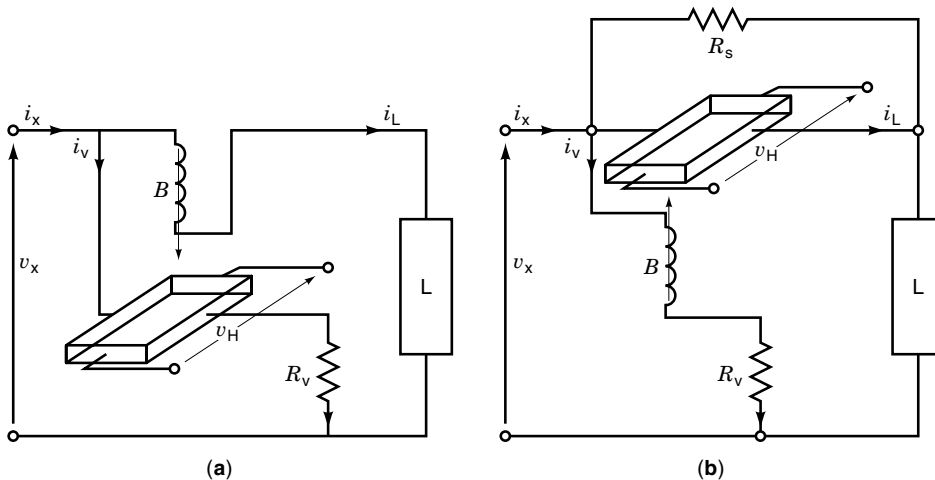


Figure 26. Configurations of the Hall-based wattmeter.

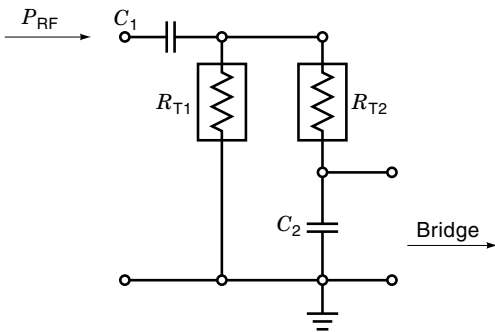


Figure 28. Working principle of the thermistor-based power meter.

provide a rough estimate of the unknown frequencies when the wideband noise superimposed on the signal is below threshold; (2) a more accurate estimate of the harmonic frequencies is carried out to determine the fractional bin frequency (i.e. the harmonic determination up to the frequency resolution); to this aim, several approaches such as zero padding, interpolation techniques, and flat-top window-based techniques can be applied (13).

High-Frequency Power Measurements

High-frequency power measurements can be classified as follows: *low power*, up to 10 mW; *medium power*, from 10 mW to 1 W; and *high power*, greater than 1 W. Meters used for power measurements at radio or microwave frequencies are generally classified as *absorption type* (containing their own load, generally 50 Ω for radio frequencies) and *transmitted or through-line type* (where the load is remote from the meter). Independently of the type, power meters are mainly based on thermistors, thermocouples, diodes, or radiation sensors.

Thermal Methods. In this sub-subsection, the main methods based on power dissipation will be examined, namely: (1) thermistor-based, (2) thermocouple-based, and (3) calorimetric.

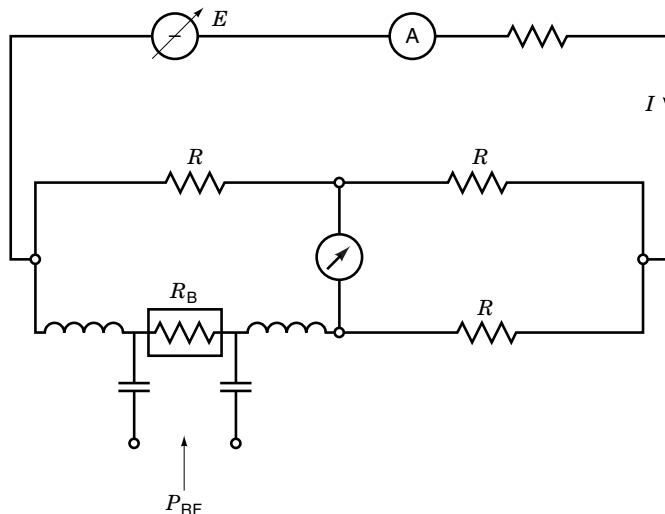


Figure 29. Manual bolometer.

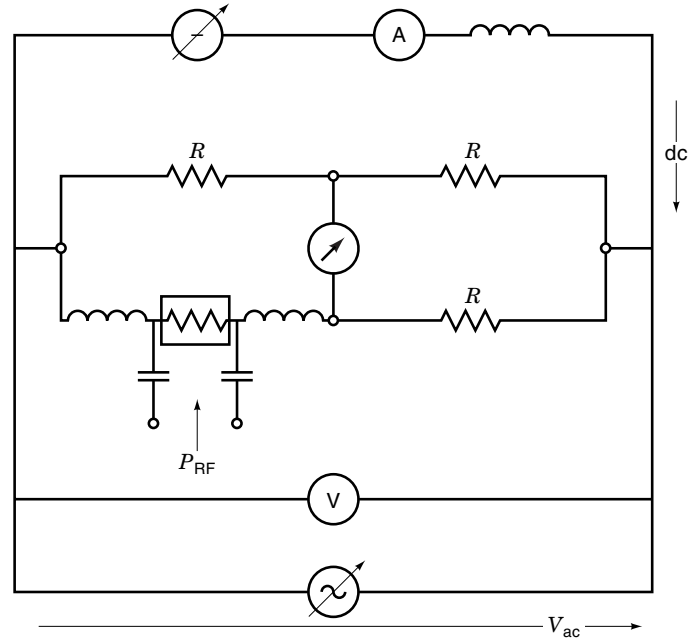


Figure 30. Manual bolometer with substitution method.

Thermistor-Based Power Meters. A thermistor is a resistor made up of a combination of highly temperature-sensitive metallic oxides (14). If it is used as a sensor in a power meter, its resistance becomes a function of the temperature rise produced by the applied power.

The working principle of the thermistor power meter is illustrated in Fig. 28 (15): two thermistors (R_{T1} and R_{T2}) are connected (1) in parallel, for measurand signals appearing at the RF input (P_{RF}) and (2) in series, for the following measuring circuit (e.g. a bridge). The capacitance C_1 prevents the dc power component from flowing to the thermistors; C_2 excludes the RF power from the bridge.

A bridge with a thermistor or a barretter in one arm is called a *bolometer*. Bolometer-based measurements are mainly performed with (1) a manual bolometer with variation of the bias current, (2) a manual bolometer with the substitution method, or (3) a self-balancing bolometer.

The *manual bolometer with a variation of the bias current* is illustrated in Fig. 29. Its working principle consists of two steps. In the first, no RF power is applied to the sensor; the equilibrium is obtained by varying the dc power supply E until the sensor resistance R_B , related to the dc power flowing in it, is equal to R . In this condition let the current I flowing into the

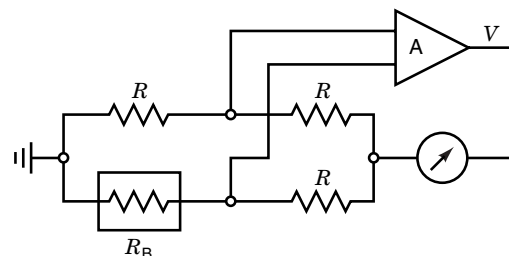


Figure 31. Self-balancing bolometer.

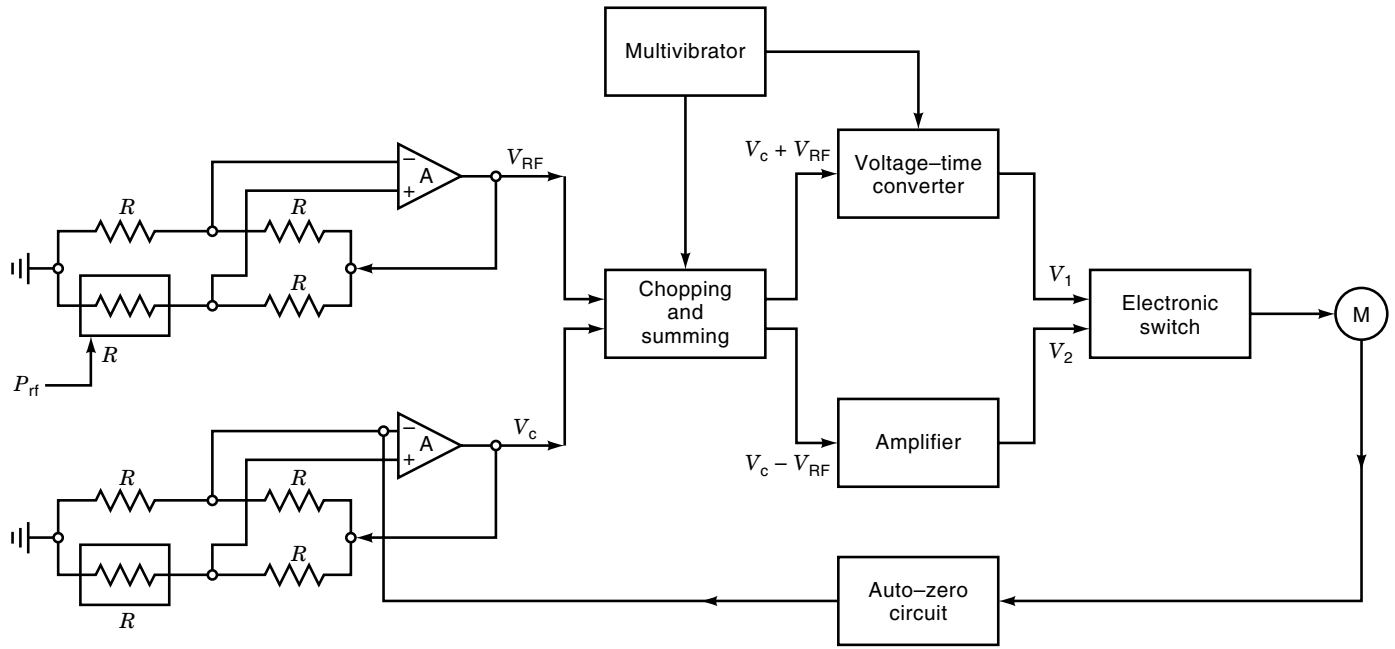


Figure 32. Power meter based on two self-balancing bridges.

circuit be equal to I_1 . In the second step, RF power P_{RF} is fed to the sensor; the power increase must be compensated by a dc power decrease, which is performed by lowering the bridge dc supply voltage E ; in this case, let I be equal to I_2 . Since the power dissipated in the sensor has been maintained constant in both steps, the power P_{RF} can be evaluated as

$$P_{RF} = \frac{R}{4}(I_1^2 - I_2^2) \quad (68)$$

The *manual bolometer with the substitution method* (Fig. 30) involves two steps. In the first, both RF power (P_{RF}) and dc power (P_{dc}) are present, and the power (P_d) necessary to bring the bridge to the equilibrium is

$$P_d = P_{dc} + P_{RF} \quad (69)$$

During the second step, P_{RF} is set to zero and an alternating voltage V_{ac} is introduced in parallel with the dc power supply. In this case, the power P_d necessary to balance the bridge,

$$P_d = P_{dc} + P_{ac} \quad (70)$$

is obtained by varying V_{ac} . Since P_d is the same in both cases,

the power supplied by the ac generator is equal to

$$P_{RF} = P_{ac} = \frac{V_{ac}^2}{4R} \quad (71)$$

Equation (71) implies that the RF power can be obtained by a voltage measurement.

The *self-balancing bolometer* (Fig. 31) automatically supplies a dc voltage V to balance the voltage variations due to changes in the sensor resistance R_B for an incident power P_{RF} . At equilibrium, R_B is equal to R , and the RF power will then be

$$P_{RF} = \frac{V^2}{4R} \quad (72)$$

As mentioned above, the thermistor resistance depends on the surrounding temperature. This effect is compensated in an instrument based on two self-balancing bridges (15). The RF power is input only to one of these, as shown in Fig. 32.

The equilibrium voltages V_c and V_{RF} feed a chopping and summing circuit, whose output $V_c + V_{RF}$ goes to a voltage-to-time converter. This produces a pulse train V_1 , whose width is proportional to $V_c + V_{RF}$. The chopping section also generates a signal with an amplitude proportional to $V_c - V_{RF}$ and

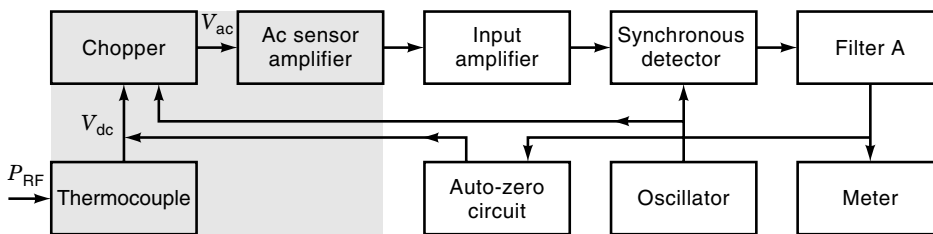


Figure 33. Power meter with thermocouple-based sensor.

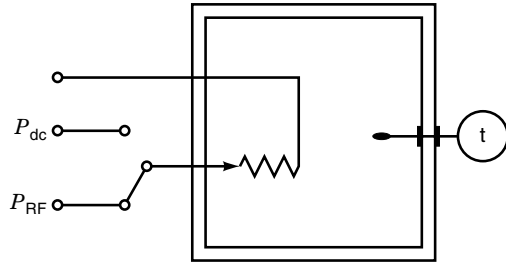


Figure 34. Calorimetric method based on a substitution technique.

a frequency of a few kilohertz, which is further amplified. The signals V_1 and V_2 enter an electronic switch, whose output is measured by a medium-value meter M . This measure is proportional to the RF power, because

$$P_{RF} = \frac{(V_c + V_{RF})(V_c - V_{RF})}{4R} = \frac{V_c^2 - V_{RF}^2}{4R} \quad (73)$$

Owing to the differential structure of the two bolometers, this device is capable of performing RF power measurements independent of the surrounding temperature. In addition, an offset calibration can be carried out when P_{RF} is null and V_c is equal to V_{RF} .

These instruments can range from 10 mW to 1 μ W and utilize sensors with frequency bandwidths ranging from 10 kHz to 100 GHz.

Thermocouple-Based Power Meters. Thermocouples (14) can be also used as RF power meters up to frequencies greater than 40 GHz. In this case the resistor is generally a thin-film type.

The sensitivity of a thermocouple can be expressed as the ratio between the dc output amplitude and the input RF power. Typical values are 160 μ V/mW for a minimum power of about 1 μ W.

The measurement of voltages of some tens of millivolt requires strong amplification, such that the amplifier does not introduce any offset. For this purpose, a chopper microvoltmeter is utilized (16), as shown in the Fig. 33. The thermocouple output voltage V_{dc} is chopped at a frequency of about 100 Hz; the resulting square wave is filtered to its mean value and then input to an ac amplifier to further reduce offset problems. A detector, synchronized to the chopper, and a low-pass filter transform the amplified square-wave voltage to a dc voltage, which is finally measured by a voltmeter.

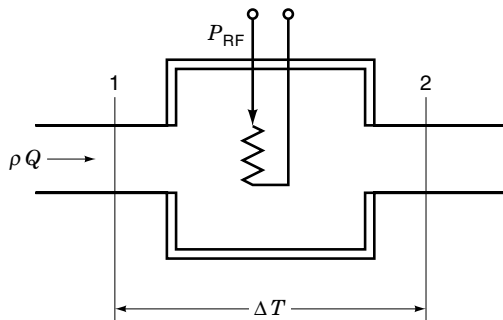


Figure 35. Calorimetric method based on a comparison technique.

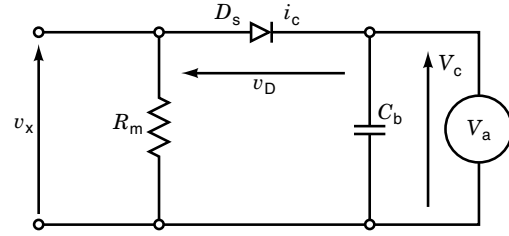


Figure 36. Circuit for diode-sensor-based power measurement.

Calorimetric Method. For high frequencies a substitution technique based on a calorimetric method is utilized (Fig. 34) (17). First, the unknown RF power P_{RF} is sent to the measuring device t , which measures the equilibrium temperature. Then, once the calorimetric fluid has been cooled to its initial temperature, dc power P_{dc} is applied to the device and regulated until the same temperature increase occurs in the same time interval. In this way, a thermal energy equivalence is established between the known P_{dc} and the measurand P_{RF} .

A comparison version of the calorimetric method is also used for lower-frequency power measurements (Fig. 35). One measures the temperature difference ΔT in a cooling fluid between the input (1) and the output (2) sections of a cooling element where the power P to be measured is dissipated. In this case, the power loss will correspond to P :

$$P = C_p \rho Q \Delta T \quad (74)$$

where C_p is the specific heat, ρ the density, and Q the volume flow of the refreshing fluid.

Diode-Sensor-Based Power Measurements. Very sensitive (down to 0.10 nW, -70 dBm) high-frequency (10 MHz to 20 GHz) power measurements are carried out through a diode sensor by means of the circuit in Fig. 36 (18). In particular, by a suitable selection of the components in this circuit, either true-average power measurements or peak power measurements can be performed.

The basic concept underlying *true-average power measurements* exploits the nonlinear (quadratic) region of the characteristic of a low-barrier Schottky diode (nonshaded area in Fig. 37). In this region, the current flowing through the diode is proportional to the square of the applied voltage; thus the diode acts as a quadratic-characteristic sensor.

In the essential circuit of diode sensor-based wattmeters shown in Fig. 36, the measurand v_x , terminated on the

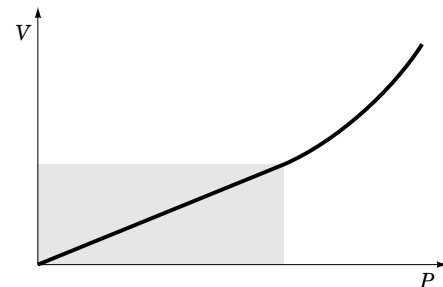


Figure 37. Characteristic of a low-barrier Schottky diode.

matching resistor R_m , is applied to the diode sensor D_s working in its quadratic region in order to produce a corresponding output current i_c in the bypass capacitor C_b . If C_b has been suitably selected, the voltage V_c between its terminals, measured by the voltmeter amplifier V_a , is proportional to the average of i_c , that is, to the average of the squares of instantaneous values of the input signal v_x , and hence to the true-average power.

In the true-average power measurement of nonsinusoidal waveforms having the biggest components at low frequency, as in RF amplitude modulation (AM), the value of C_b must also satisfy another condition. The voltage v_d on the diode must be capable of holding the diode switched on (conducting) even for the smallest values of the signal. Otherwise, in the valleys of the modulation cycle, the high-frequency modulating source is disconnected by the back-biased diode and the measurement is therefore misleading.

On the other hand, for the same signal but for a different selection of the C_b value, the circuit can act as a peak detector for *peak power measurements*. As a matter of fact, the voltage v_c on the bypass capacitor C_b during the peak of the modulation cycle is so large that in the valleys the high-frequency peaks are not capable of switching the diode into conduction; thus these peaks do not contribute to the measured power level.

If higher power levels have to be measured (10 mW to 100 mW), the sensing diode is forced to operate outside the quadratic region, in the linear region (shaded area in Fig. 37). In this case, the advantage of true-average power measurements for distorted waveforms is lost, and for peak power measurements, since the diode input-output characteristic is nonlinear and the output is quadratic, spectral components different from the fundamental introduce significant measuring errors.

BIBLIOGRAPHY

1. G. Zingales, Measurements on steady-state circuits, in *Methods and Instruments for Electrical Measurements*, (in Italian), Torino: UTET, 1980, Chap. VI.
2. G. Zingales, Power measurements on single-phase ac circuits, in *Methods and Instruments for Electrical Measurements* (in Italian), Torino: UTET, 1980, Chap. VI, Sec. 6.2.
3. G. Korányi, Measurement of power and energy, in L. Schnell (ed.), *Technology of Electrical Measurements*, Chichester: Wiley, 1993.
4. J. Milmann and C. C. Halkias, *Integrated Electronics: Analog and Digital Circuits and Systems*, New York: McGraw-Hill, 1972.
5. M. L. Sanderson, Power measurement, in *Instrumentation Reference Book*, Butterworths, 1988, Part 3, chap. 1.4.
6. P. S. Filipski, A TDM wattmeter with 0.5 MHz carrier frequency, *IEEE Trans. Instrum. Meas.*, **IM-39**: 15–18, 1990.
7. J. R. Carstens, *Electrical Sensors and Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1992.
8. Z.-L. Lu, An error estimate for quasi-integer-period sampling and an approach for improving its accuracy, *IEEE Trans. Instrum. Meas.*, **IM-23**: 337–341, 1984.
9. V. Haasz, The error analysis of digital measurements of electrical power, *Measurement*, Vol. 6, No. 4, Oct.–Dec., 1988.
10. P. Arpaia et al., Real-time algorithms for active power measurements on PWM-based electric drives, *IEEE Trans. Instrum. Meas.*, **IM-45**: 462–466, 1996.
11. X. Dai and R. Gretschi, Quasi-synchronous sampling algorithm and its applications, *IEEE Trans. Instrum. Meas.*, **IM-43**: 204–209, 1994.
12. M. Bellanger, *Digital Processing of Signals: Theory and Practice*, Chichester: Wiley, 1984.
13. M. Bertocco, C. Offelli, and D. Petri, Numerical algorithms for power measurements, *Eur. Trans. Electr. Power*, **ETEP 3**: 91–101, 1993.
14. H. N. Norton, Thermometers, in *Handbook of Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1989, Chap. 19.
15. Anonymous, Thermistor mounts and instrumentation, *Fundamentals of RF and Microwave Power Measurements*, Application Note 64-1, Hewlett Packard, 1978, Chap. II.
16. R. E. Pratt, Power measurements in C. F. Coombs (ed.), *Handbook of Electronic Instruments*, New York: McGraw-Hill, 1995, Secs. 15.1–15.16.
17. F. F. Mazda, High-frequency power measurements, in *Electronic Instruments and Measurement Techniques*, Cambridge, UK: Cambridge University Press, 1987, Chap. VIII, Sec. 8.4.
18. Anonymous, Diode detector power sensors and instrumentation, in *Fundamentals of RF and Microwave Power Measurements*, Application Note 64-1, Hewlett Packard, 1978, Chap. IV.

Reading List

- Anonymous, *Fundamentals of RF and Microwave Power Measurements*, Application Note 64-1, Hewlett Packard, 1978. Though not very recent, this is a valid and comprehensive reference for main principles of high-frequency power measurements.
- P. Arpaia et al., An expert system for the optimum design of measurement systems, *IEEE Proc. A*, **142**: 330–336, 1995. Reports on an artificial intelligence tool for the automatic design of power-measuring systems.
- F. Avallone, C. De Capua, and C. Landi, Measurement station performance optimization for testing on high efficiency variable speed drives, *Proc. IEEE IMTC/96*, Brussels, 1996, pp. 1098–1103. Proposes an analytical model of uncertainty arising from power measurement systems working under highly distorted conditions.
- F. Avallone, C. De Capua, and C. Landi, Measurement and reconstruction techniques applied to power measurements on high efficiency variable speed drives, *Proc. XIV IMEKO World Congress*, Tampere, 1997. Proposes a technique to improve accuracy of power measurements under highly distorted conditions.
- F. Avallone, C. De Capua, and C. Landi, A digital technique based on real-time error compensation for high accuracy power measurement on variable speed drives, *Proc. IEEE IMTC/97*, Ottawa, 1997. Reports on a real-time technique for error compensation of transducers working under highly distorted conditions.
- F. Avallone, C. De Capua, and C. Landi, Metrological performance improvement for power measurements on variable speed drives, *Measurement*, vol. 21, no. 1/2, May/June 1997. Shows how to compute the uncertainty of measuring-chain components for power measurements under highly distorted conditions.
- G. Bucci, P. D'Innocenzo, and C. Landi, A modular high-speed DSP-based data acquisition apparatus for on-line quality analysis on power systems under non-sinusoidal conditions, *Proc. 9th IMEKO TC4 Int. Symp.*, Budapest, 1996, pp. 286–289. Gives a strategy of measurement system design for power measurements under nonsinusoidal conditions.
- K. K. Clarke and D. T. Hess, A 1000 A / 20 kV / 25 kHz–500 kHz volt-ampere-wattmeter for loads with power factors from 0.001 to 1.00, *IEEE Trans. Instrum. Meas.*, **IM-45**: 142–145, 1996. Provides information on the implementation of an instrument to perform an accurate measurement of currents (1 A to 1000 A), voltages (100 V to 20 kV), and powers (100 W to 20 MW) over the frequency range from 25 kHz to 500 kHz.
- J. J. Clarke and J. R. Stockton, Principles and theory of wattmeters operating on the basis of regularly spaced sample pairs, *J. Phys. E*.

- Sci. Instrum.*, **15**: 645–652, 1982. Gives the basis of regularly spaced sample pairs, *J. Phys. E. Sci. Instrum.*, **15**: 645–652, 1982. Gives the basics of synchronous sampling for digital wattmeters.
- J. W. Gardner, *Microsensors: Principles and Applications*, Chichester: Wiley, 1994.
- S. L. Garverick et al., A programmable mixed-signal ASIC for power metering, *IEEE J. Solid State Circuits*, **26**: 2008–2015, 1991. Reports on a programmable mixed analog–digital integrated circuit based on six first-order sigma–delta ADCs, a bit serial DSP, and a byte-wide static RAM for power metering.
- F. K. Harris, The measurement of power, in *Electrical measurements*, Huntington, NY: Krieger, 1975, Chap. XI. A clear reference for line-frequency power measurements.
- J. K. Kolanko, Accurate measurement of power, energy, and true rms voltage using synchronous counting, *IEEE Trans. Instrum. Meas.*, **IM-42**: 752–754, 1993. Provides information on the implementation of a synchronous dual-slope wattmeter.
- F. F. Mazda, Pulse power measurement, in *Electronic Instruments and Measurement Techniques*, Cambridge, UK: Cambridge University Press, 1987, Chap. VIII, Sec. 8.5.
- J. C. Montano et al., DSP-based algorithm for electric power measurement, *IEE Proc. A*, **140**: 485–490, 1993. Describes a Goertzel FFT-based algorithm to compute power under nonsinusoidal conditions.
- H. N. Norton, Radiation pyrometers, in *Handbook of Transducers*, Englewood Cliffs, NJ: Prentice-Hall, 1989, Chap. 20.
- T. S. Rathore, Theorems on power, mean and rms values of uniformly sampled periodic signals, *IEE Proc. A*, **131**: 598–600, 1984. Provides fundamental theorems for effective synchronous sampling wattmeters.
- G. N. Stenbakken, A wideband sampling wattmeter, *IEEE Trans. Power Appar. Syst.*, **PAS-103**: 2919–2925, 1984. Gives the basics of asynchronous-sampling-based wattmeters and criteria for computing uncertainty in the time domain.

GIOVANNI BUCCI
CARMINE LANDI
University of Aquila
CLAUDIO DE CAPUA
Federico II University of Naples

POWER METERS

Several kinds of power meters are employed in electrical and electronics engineering. Depending on the specific application, a unique power meter must be used. A description of the main kinds of power meters (dc power meters, ac power meters for electric power applications, and power meters for high-frequency electronic systems) follows. In electronic circuits, for the frequency range of approximately 1 Hz to 500 kHz, depending on the nature of the signal, power may be measured with instruments implementing methods similar to those employed in ac power meters for electric power applications or power meters for high-frequency.

DC POWER METERS

The direct current (dc) power dissipated in a load resistance R_L may be calculated from measured voltage across R_L (with a dc voltmeter) and measured current flowing through R_L (with a dc ammeter). This task may be performed in the two circuit configurations that are shown in Fig. 1. In the circuit of Fig. 1(a), the voltmeter measures the voltage drop V_L across

R_L , while the ammeter measures the sum of two currents, I_L , flowing through R_L , and I_V , flowing through the voltmeter. If no correction associated with I_V is made, then the dc power is calculated as

$$P_{(a)} = IV_L = (I_L + I_V)V_L \quad (1)$$

If a correction for I_V is made, then the power dissipated in the load, P_L , is calculated as

$$P_L = I_L V_L = IV_L - I_V V_L = P_{(a)} - P_V \quad (2)$$

where $P_V = I_V V_L$ is the power dissipated in the voltmeter. The worst-case relative error with which the P_L is measured in the circuit of Fig. 1(a) while correction for I_V is made is

$$\delta_{PL(a)} = \frac{I}{I_L} \delta_I + \delta_{VL} + \frac{I_V}{I_L} \delta_{IV} \quad (3)$$

where δ_I is the relative measurement error of I (error of the ammeter), δ_{VL} is the relative measurement error of V_L (error of the voltmeter), and δ_{IV} is the relative error with which the I_V has been estimated (calculated from known voltmeter resistance R_V and measured V_L). If the power calculated from Eq. (1) is taken as the result of a measurement (no correction is applied), then the error of the P_L in the circuit shown in Fig. 1(a) is

$$\delta_{PL(a)} = \frac{I}{I_L} \delta_I + \delta_{VL} + \frac{I_V}{I_L} \quad (4)$$

If the circuit of Fig. 1(b) is used to measure P_L , then the ammeter measures the current I_L flowing through R_L , while the voltmeter measures the sum of two voltage drops, V_L across R_L , and V_A across the ammeter. If no correction related to V_A is made, then the dc power is calculated as

$$P_{(b)} = I_L V = I_L (V_L + V_A) \quad (5)$$

If a correction for V_A is made, then the power dissipated in the load, P_L , is calculated as

$$P_L = I_L V_L = I_L V - I_L V_A = P_{(b)} - P_A \quad (6)$$

where $P_A = I_L V_A$ is the power dissipated in the ammeter. The worst case relative error with which the P_L is measured in the circuit of Fig. 1(b) while correction for V_A is made is

$$\delta_{PL(b)} = \delta_{IL} + \frac{V}{V_L} \delta_V + \frac{V_A}{V_L} \delta_{VA} \quad (7)$$

where δ_{IL} is the relative measurement error of I_L (error of the ammeter), δ_V is the relative measurement error of V (error of the voltmeter), and δ_{VA} is the relative error with which the V_A has been estimated (calculated from known ammeter resistance R_A and measured I_L). If the power calculated from Eq. (5) is taken as the result of a measurement (no correction is applied) then the error of the P_L is

$$\delta_{PL(b)} = \delta_{IL} + \frac{V}{V_L} \delta_V + \frac{V_A}{V_L} \quad (8)$$

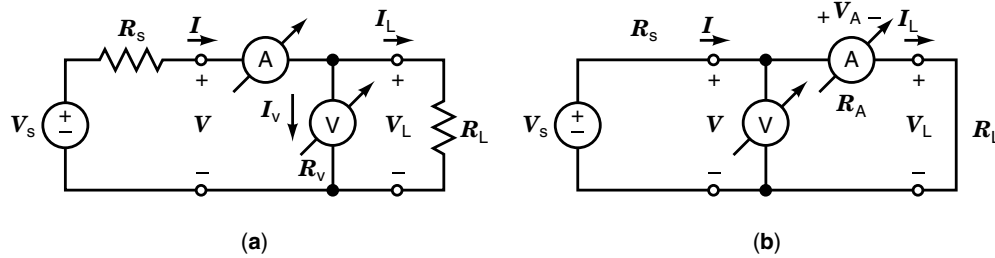


Figure 1. Dc power measurement using a voltmeter and an ammeter. There are two possible measurement circuits: (a) direct V_L measurement, where the voltage drop across the load is measured directly but the ammeter measures a sum of load current and voltmeter's current, and (b) direct I_L measurement. The result of power measurement needs to be computed from the measured voltage and current.

To decide which of the two circuits is preferable, errors $\delta_{PL(a)}$ and $\delta_{PL(b)}$ need to be compared. The circuit that gives a lower measurement error is preferred. The key criterion obtained from this comparison is: which of the two ratios is smaller, I_V/I_L or V_A/V_L ?

Power meters process the information of I_L and V_L and display the measured power with given accuracy. Block diagrams of two dc power meters are shown in Fig. 2. The circuit shown in Fig. 2(a) implements analog multiplication of signals proportional to V and I_L . The output voltage of an analog multiplier, V_O , is proportional to the product of two input voltages. In the circuit of Fig. 2(a), V_O is proportional to the power dissipated in R_L and may be converted to a digital form with an analog-to-digital converter (ADC), or displayed in an analog form. Analog multipliers are integrated circuits and they allow the presence

of a limited voltage at their inputs. Therefore, when the input voltage V is equal or exceeds the full scale value, a voltage attenuator needs to be employed. Or else, if V is too small to be processed with the smallest possible error, an amplifier needs to be used. In circuits of Fig. 2, current I_L is converted to voltage with a current sensor that has a low-pass frequency response (includes dc in its frequency range). The benefit of this type of current to voltage conversion is a small equivalent R_A resistance. Also, the current sensor may be of clamp-on type, that makes the connection of the power meter to the circuit under test very easy. Output voltage from current sensors is always a small value and requires amplification before it is applied to the input of the analog multiplier. Integrated circuits specially designed for applications in power measurement circuits are available. An example of such a device is the AD7750 product-to-frequency converter from the Analog Devices company (World Wide Web Site: <http://www.analog.com>).

The circuit depicted in Fig. 2(b) shows block diagram of a digital power meter. Instead of using the analog multiplier, voltages proportional to V and I_L are converted to a digital form with ADCs. Digital values D_V and D_I proportional to V and I_L are then processed by the digital part of the power meter. The circuit shown in Fig. 2(b) uses two ADCs, but sometimes one DAC and an analog multiplexer in front of it may be used.

Measurement errors of circuits shown in Fig. 2 are computed similarly as errors for the circuit shown in Fig. 1(b); see Eqs. (5) to (8). An equivalent voltage V_A (or equivalent R_A) may be specified for the current sensor as for an ammeter.

AC POWER METERS USED IN ELECTRIC POWER APPLICATIONS

Instantaneous power p is a product of current i and voltage v that are functions of time t

$$p(t) = i(t)v(t) \quad (9)$$

Fig. 3(a) shows sinusoidal $i(t)$ and $v(t)$ as functions of phase angle ωt . In the graph, the phase shift of v is assumed to be 0° and current i lags the voltage by 50° (inductive-resistive load). The power p for these v and i is shown in Fig. 3(b). In cases when the instantaneous power is the signal that carries the desired information, circuits like those shown in Fig. 2 are used. An example of using the instantaneous power as a medium for fault detection of electric motors is described in the paper of

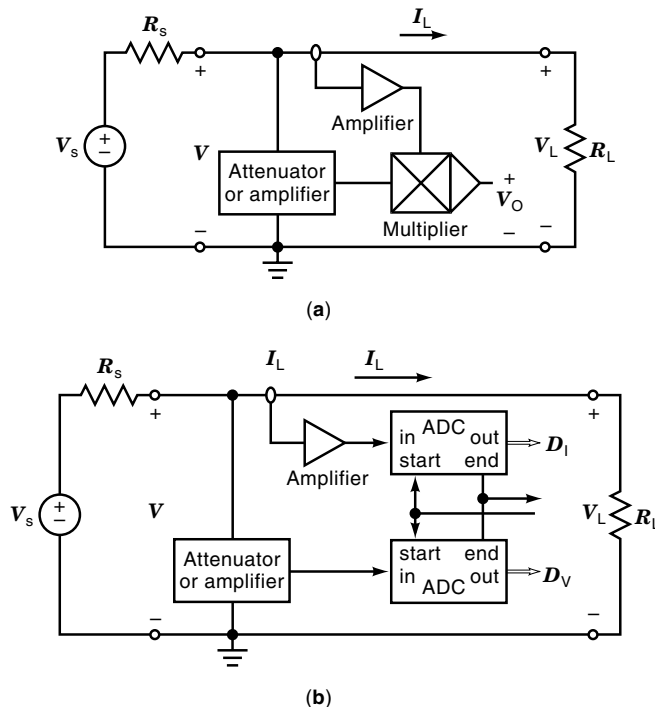


Figure 2. Dc power meters provide the value of measured power. One of two methods of computing the power may be employed: (a) analog power computation, (b) digital power computation.

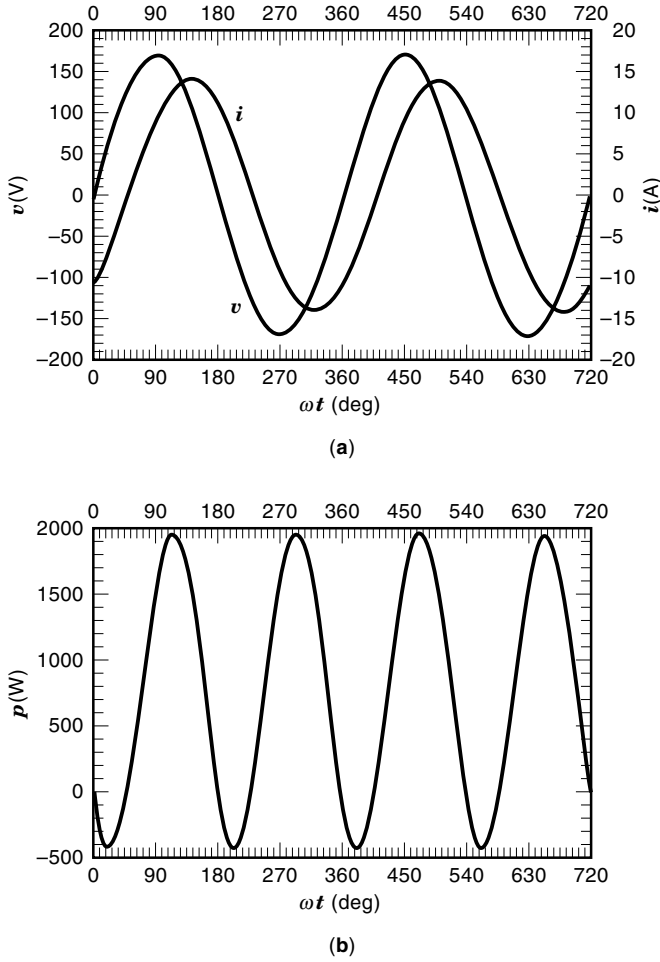


Figure 3. Ac voltage, current, and instantaneous power, (a) sinusoidal $v(t)$ and $i(t)$, (b) instantaneous power $p(t)$. Instantaneous power is a function of time and may take positive or negative values.

Legowski, Ula, and Trzynadlowski (1). In order to update the circuits shown in Fig. 2 to ac applications, an ac voltage source with Thévenin equivalent impedance $Z_S(s)$ needs to be used in place of the dc voltage source. As the load in ac circuits an impedance $Z_L(s)$, rather than R_L , needs to be used. All symbols of voltages and currents need to be changed to those that represent ac signals. In these modified circuits, the output voltage from the block “Attenuator or Amplifier” is proportional to $v(t)$ and the output voltage from the block “Amplifier” is proportional to $i_L(t)$. In the ac circuit similar to the one shown in Fig. 2(a), the output voltage from the analog multiplier is $v_o(t)$ and is proportional to $p(t)$. While using this circuit, the $v_o(t)$ is further processed in an analog circuit or sampled with an ADC and processed digitally. In the ac circuit similar to the one shown in Fig. 2(b), output voltages from the “Attenuator or amplifier” and “Amplifier” are sampled with ADCs at a sampling frequency suitable for the digital signal processing of the $p(t)$. In both circuits, the bandwidth of the current sensor, “Attenuator or amplifier,” and “Amplifier” must match the bandwidths of $v(t)$ and $i_L(t)$. In order to recall the other definitions of power in ac circuits, assume that the excitation is sinusoidal and the circuit is linear. Thus phasor analysis may be used. Definition

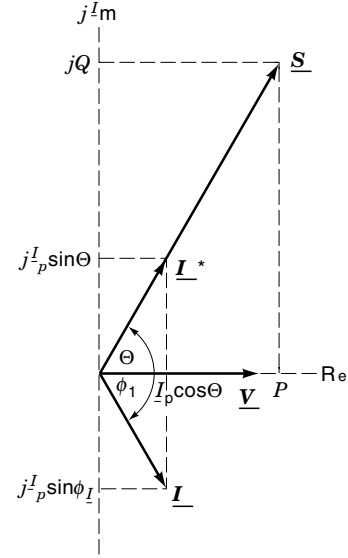


Figure 4. Phasor diagram of voltage, current, and power.

of the complex power is

$$\underline{S} = \frac{1}{2} \underline{V} \underline{I}^* \quad (10)$$

which results in

$$\underline{S} = P + jQ \quad (11)$$

where P is the active power (or real power, or average power), and Q is the reactive power. A phasor diagram that shows a superposition of current, voltage, and power phasors is shown in Fig. 4. This phasor diagram shows a specific case for which the phase shift of the voltage is equal zero. The active power and reactive power are

$$P = \frac{1}{2} V_p I_p \cos \Theta \quad [\text{W}] \quad (12)$$

and

$$Q = \frac{1}{2} V_p I_p \sin \Theta \quad [\text{VAR}] \quad (13)$$

where V_p is the peak value of the voltage, I_p is the peak value of the current, and Θ is

$$\Theta = \phi_V - \phi_I$$

In the last equation, ϕ_V is the phase shift of the voltage, and ϕ_I is the phase shift of the current. The apparent power is the magnitude of CBISul;

$$|\underline{S}| = \sqrt{P^2 + Q^2} = \frac{1}{2} V_p I_p \quad [\text{VA}] \quad (14)$$

A graph of various components of ac power in the time domain (as functions of phase shift ωt) is shown in Fig. 5. As before, $\phi_V = 0^\circ$ and $\phi_I = -50^\circ$ have been chosen for the figure. In this figure, two components of $p(t)$ are shown, $p_p(t)$ representing the energy

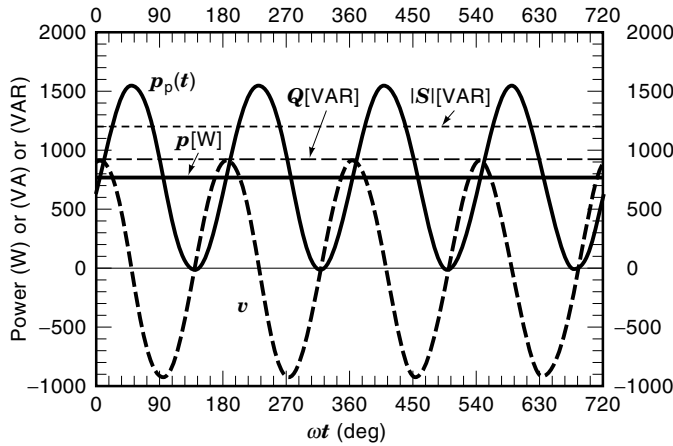


Figure 5. Components of ac power in time domain.

flow into the load impedance, and $p_Q(t)$ which characterizes the energy borrowed and returned by the load impedance. The real power P is at the same time the average value of $p(t)$ and also the average value of $p_p(t)$. The reactive power Q is the peak value of $p_Q(t)$, while the average value of $p_Q(t)$ equals zero. The apparent power $|S|$ equals one half of the peak-to-peak value of $p(t)$. A description of power in ac circuits may be found in the book by Cunningham and Stuller (2).

Ac power meters are designed to measure individual components of the ac power (including power factor, $PF = \cos\Theta$) or given sets of these components. The electrodynamic instrument (or dynamometer, or electrodynamicometer) is the moving coil instrument that measures active power. The main parts of its mechanism are shown in Fig. 6(a). Between 1843 to 1910, the work of researchers like Wilhelm E. Weber, Lord Kelvin, James P. Joule, Andre Marie Ampere, and the brothers Werner and William Siemens made possible the development of a reliable electrodynamic instrument that started to be mass manufactured. In it, the magnetic field is produced by a two-part fixed coil conducting current i_f . This magnetic field and current i_m in the moving coil produce a torque that turns the moving mechanism of the meter. Fig. 6(b) depicts a circuit in which ac power is measured [similarly as shown in Fig. 1(b)] with the electrodynamic instrument. The average value of the torque is

$$T_m = \frac{1}{T} \int_0^T i_m i_f \frac{dM}{d\alpha} dt \quad (15)$$

where M is the mutual inductance between the moving and fixed coils, α is the angular deflection of the moving coil, and T is the period of i_f and i_m . In series with the moving coil is connected a reference resistor R_V that makes the current i_m proportional to the voltage v . Current i_m is carried to the moving coil by two control springs (that produce the reference restoring torque) and the shaft of the moving mechanism. The position of the pointer that is attached to the shaft of the moving mechanism is read from a scale in watts. The electrodynamic instrument can measure dc power. When ac power is measured, the averaging effect of the T_m is produced by the inertia of the moving mechanism. Therefore, it is not possible to measure the power at frequencies below 10 Hz or so, because the averaging effect is not in place yet. On the other hand, the electrodynamic

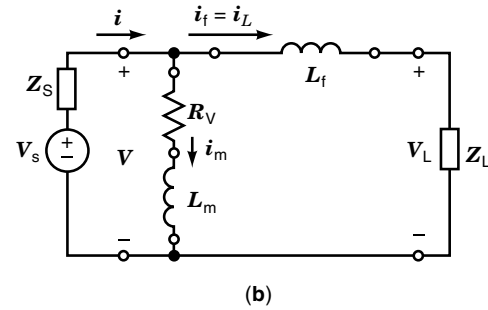
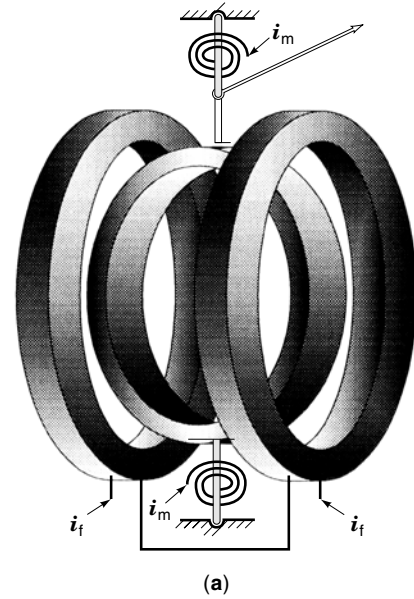


Figure 6. Electrodynamic power meter. Inductance L_f is formed by a series connection of two inductors. Rotating inductor L_m is placed in the middle of these two inductors.

instrument may be used for frequencies up to 400 Hz. These instruments are optimized for measurements at 50 Hz or 60 Hz and may be used for sine wave signals with limited nonlinear distortions (harmonics of the line current up to the seventh harmonic are included in their frequency range). Measurement error of these instruments is usually in the range from 1% to 5% when the measured power equals the full scale value. When smaller than full scale values are measured, the measurement error increases accordingly. Other types of power meters that convert electric power to deflecting torque, for example, electrostatic wattmeters or moving-iron wattmeters, were also developed. Meters of other components of electric power, for example, power-factor meters and reactive power meters (varmeters), have been developed too. More information on torque based measurement instruments may be found in books by Kidwell (3) and Kinnard (4).

Figure 7 shows the block diagram of single phase electronic power meter. This instrument measures a set of components of the ac power that includes true rms values of line current and line voltage and the active power. The electronic power meter shown in Fig. 7 uses analog integrated circuits to convert the line current and voltage to their rms values and to produce a signal proportional to the instantaneous power. The ADCs in this type of wattmeter are of the integrating type (similar to

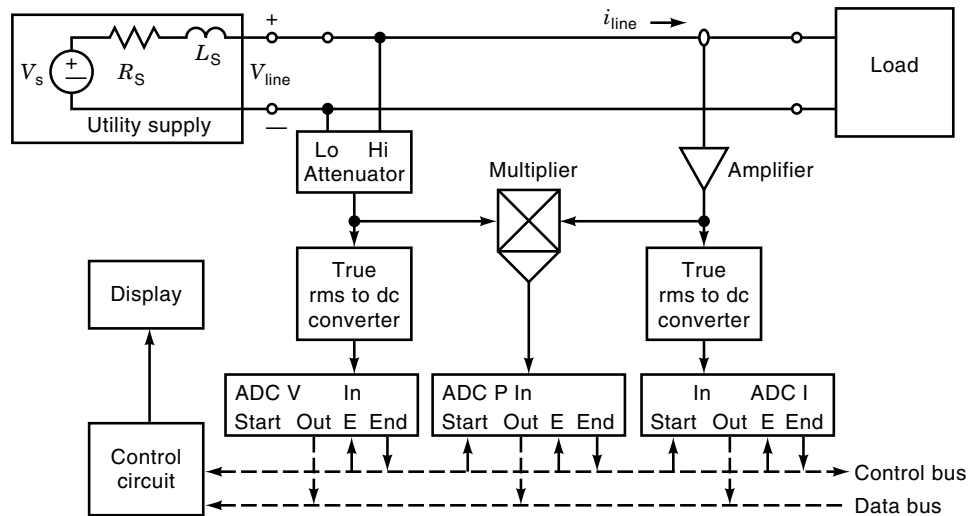


Figure 7. Single-phase electronic power meter with analog computation circuit. The output voltage from the multiplier is proportional to the instantaneous power. Digital output of the integrating analog-to-digital converter is proportional to the average power.

those used in digital multimeters). The integrating ADC converts the output voltage from the analog multiplier, that is, proportional to the instantaneous power, to its average value over the time of integration, that is, to the active power. The power factor may be computed from these measurements using Eq. (12) if v_{line} and i_{line} are sinusoidal. Also, reactive power and apparent power may be computed.

Electronic power meters are digital instruments. An example of this class of power meters are the instruments developed by Valhalla Scientific (5) and Fluke (6,7). They are more accurate than electrodynamic instruments, measure more than one component of power, and have several ranges of measured quantities. Electronic power meters may measure active power in the range from 1 mW to 12 kW, voltage from 1 V to 600 V, current from 0.1 mA to 20 A. They have much better frequency characteristic than the electrodynamic instrument. Electronic wattmeters can be used for measuring dc power and ac power from 20 Hz to 500 kHz. They have the total measurement uncertainty from 0.1% to 0.5% of reading for frequencies up to 5 kHz or 10 kHz. For frequencies larger than that the uncertainty deteriorates, and is from 2% to 5% at 20 kHz. It worsens

further for increasing frequencies. Electronic power meters may be used when the current's crest factor is in the range from 2.5:1 to 50:1, depending on the ratio of the rms value of the measured current to the full scale value. The smaller this ratio is, the larger the crest factor may be.

In the circuit of Fig. 7, a Hall effect current sensor is used. There are three types of current sensors: (1) resistive shunt (for currents lower than 500 A), (2) current transformer, and (3) Hall effect device. Two kinds of resistive shunts are made, for frequencies up to 100 Hz and for frequencies exceeding 500 kHz. Among the advantages of the resistive shunt current sensors are: they convert dc and ac current, are very reliable, do not produce offset voltage, and those made for frequencies below 100 Hz are not expensive. To the disadvantages of these current sensors belong: a lack of electrical isolation from the line voltage, significant power loss, they require amplification of the output voltage, those made for frequencies exceeding 500 kHz are expensive. Current transformer type sensors may sense currents up to 100 kA and their output current may be easily converted to voltage. Their advantages include: electrical isolation from the line voltage, they are very reliable, and

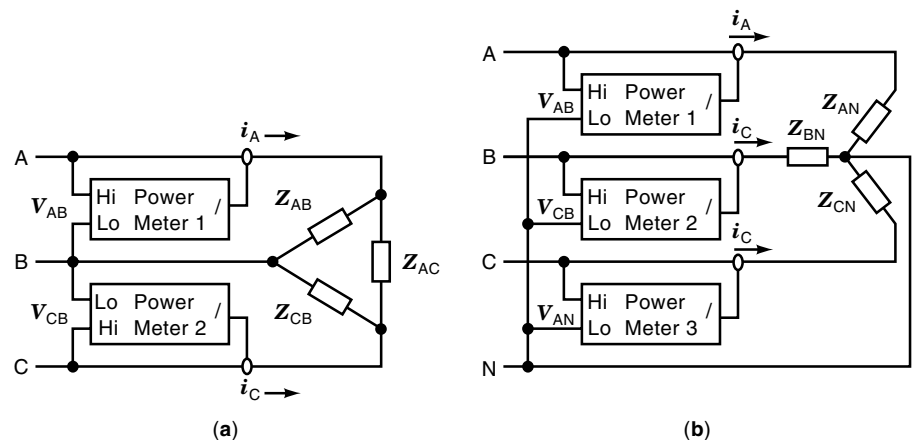


Figure 8. Three-phase power measurement circuits: (a) circuit for Δ connected load or two power meter circuit, and (b) circuit for Y connected load with neutral line or three power meter circuit.

low cost. The disadvantages are: they sense ac current only, the power loss is not negligible, their output current is frequency dependent, and they are susceptible to stray ac magnetic fields. The Hall effect current sensors can measure dc and ac currents. They provide electrical isolation from the line voltage, may measure large currents (3 kA or so), some have bandwidth from dc to 200 kHz, and they are very reliable. The disadvantages include: they require external power supply and the output signal includes an offset that needs to be compensated. An electronic power meter may include resistive shunt current sensors. In such a case, the range of the current input may be increased up to a few kiloamperes by using additional current transformers. However, use of current transformers increases the measurement error by about 1% or 2% and the frequency range will be limited to a range from 40 Hz to about 400 Hz. More information on current sensors is provided in the paper written by Drafts (8).

Figure 8 shows circuits for three phase power measurements with single-phase power meters. The circuit shown in Fig. 8(a) is used when the load is Δ connected and only three phase lines are used. In such a case two power meters are used. Inputs to "Power meter 1" are i_A and v_{AB} , while inputs to "Power meter 2" are i_B and v_{CB} . Power delivered to the three phase load is

$$P_{3\phi 3W} = P_1 + P_2 \quad [W] \quad (16)$$

where $P_{3\phi 3W}$ is the total active power measured in the three-phase, three-wire circuit, P_1 and P_2 are the active powers measured by "Power meter 1" and "Power meter 2", respectively. Figure 8(b) shows the measurement circuit used when in addition to the three phase lines the neutral conductor is used. Three power meters are used in such a case. Inputs to the "Power meter 1" are i_A and v_{AN} , inputs to the "Power meter 2" are i_B and v_{BN} , and inputs to the "Power meter 3" are i_C and v_{CN} . The power delivered to the three phase load is

$$P_{3\phi 4W} = P_1 + P_2 + P_3 \quad [W] \quad (17)$$

where $P_{3\phi 4W}$ is the total active power measured in the three-phase, four-wire circuit, P_1 , P_2 , and P_3 are the active powers measured by "Power meter 1", "Power meter 2", and "Power meter 3", respectively.

In many cases, for example, when adjustable speed drives are used to drive electric motors, the line current is not sinusoidal at all. An adjustable speed drive may be supplied from a single phase line (this is not a typical case), as it is shown in Fig. 9(a), or from a three-phase line, as shown in Fig. 9(b). An adjustable speed drive contains a rectifier with a huge capacitance at its output and an inverter that converts the dc supply to

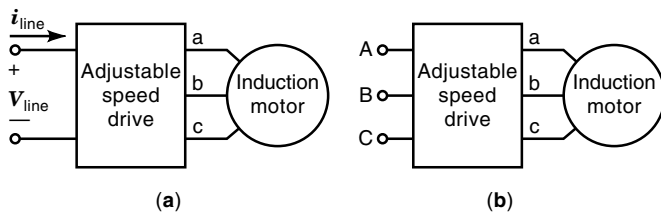


Figure 9. Adjustable speed motor drives: (a) motor drive with single-phase input or single-phase or three-phase converter, (b) motor drive with three-phase input or three-phase to three-phase converter.

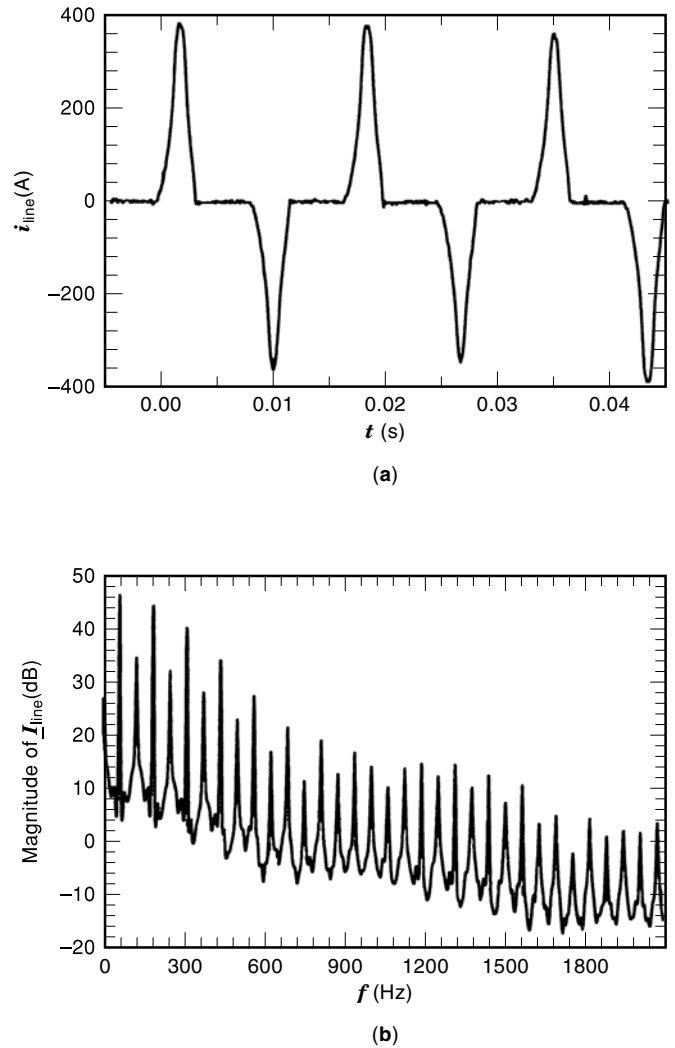


Figure 10. Line current of an adjustable speed drive, (a) timing diagram, (b) spectrum.

adjustable frequency three-phase supply. Because of this huge capacitance the line current i_{line} has a form of relatively narrow pulses. An oscillogram of i_{line} of a single phase inverter is shown in Fig. 10(a) and spectrum of it is shown in Fig. 10(b). Because i_{line} may be of the nature shown in Fig. 10(a), one of the important parameters of a power meter is the maximum acceptable crest factor of the current. A power meter usually has an indicator in the form of a light-emitting diode (LED) that warns the user when the current has too large of a crest factor and the measurements are unreliable. An example of waveforms on the output from the "Adjustable speed drive" that drove a 60 hp induction motor are shown in Fig. 11(a) and (b). The inverter used in the adjustable speed drive that produced these signals was a voltage source type and the direct torque control was used in it. Line-to-line voltages at the output of this drive are pulse trains, as shown in Fig. 11(a). Because an inductive load is driven by the drive, the current waveform is similar to the integral of the voltage and are sinusoidal with a ripple component, as shown in Fig. 11(b). Figure 11(c) shows the spectrum of the line current. An analysis of the graphs shown in Fig. 11 makes it clear that high-bandwidth instruments must be used

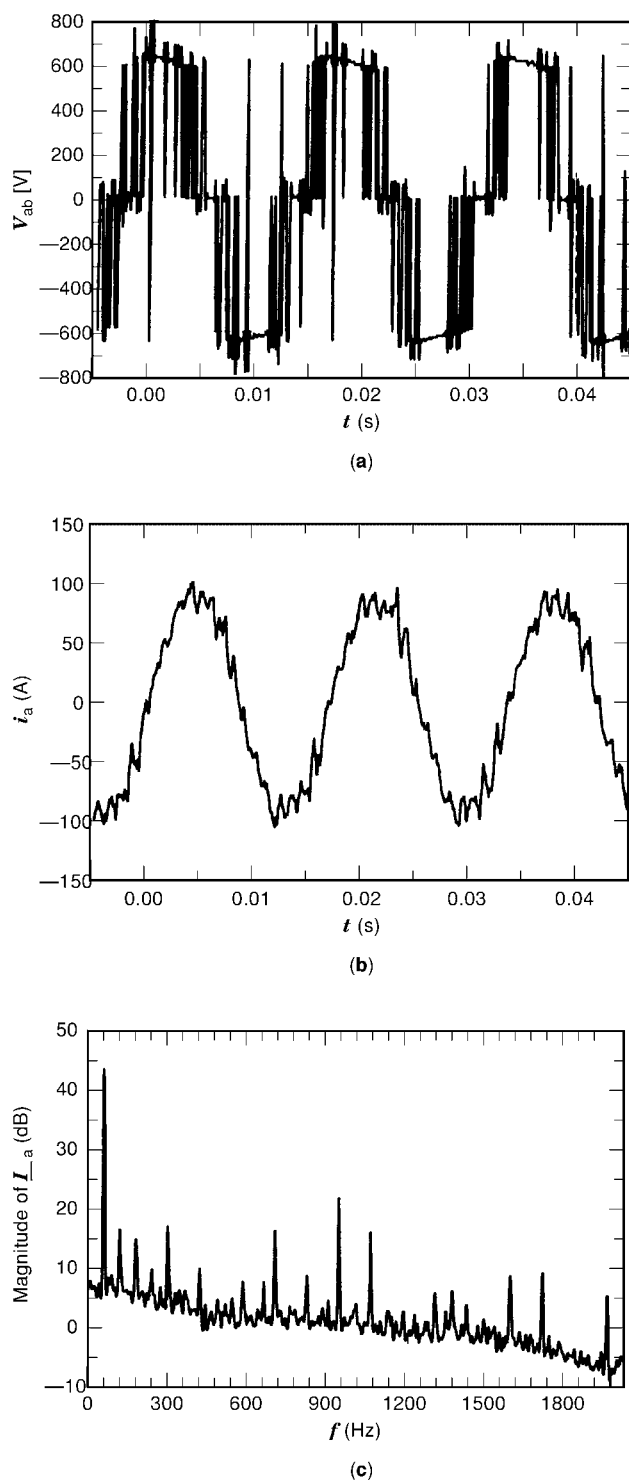


Figure 11. Line-to-line voltage and line current of an adjustable speed drive, (a) line-to-line voltage, (b) line current, (c) spectrum of line current.

to measure the output power of the “Adjustable speed drive” shown in Fig. 9. Electronic instruments called power analyzers are used to make these measurements. They are capable to measure power delivered to the “Adjustable speed drive” (rectifier with a large capacitor on its output), as well as the output power of the “Adjustable speed drive.” A block diagram

of a power analyzer is shown in Fig. 12. It has three input channels, but it may be used in any configuration, from single-phase power meter to three-phase, three-wattmeter measurement system. Current inputs accept various current sensors. Signal-processing methods are used to compute all components of three phase power, including spectrum analysis of line currents. Every channel has a “Programmable attenuator” and “Programmable amplifier” for matching the signal levels with the ranges of ADCs. These ADCs sample the signals with sampling frequency that is required for processing signals like these shown in Figs. 10 and 11. Every programmable attenuator/amplifier with its ADC is supplied by a floating power supply, because its input is connected to a three-phase system. Digital signals from every programmable attenuator/amplifier are transmitted through isolation circuits to the digital signal processing (DSP) system, where the signal processing takes place. Power analyzers may be used as instruments in a measurement system using the IEEE-488 Bus or connected to a computer using the RS-232-C transmission line. Power analyzers measure active power, reactive power, apparent power, power factor, rms value of voltages and currents, peak values of voltages and currents, crest factor of voltages and currents, peak inrush current, impedance, spectrum of voltages and currents (fundamental frequency, up to 99th harmonic, and total harmonic distortion), and integrals of active and reactive power (Whr and Varhr). Active power range is from 25 mW to 400 kWp, voltage range is from 0.5 V to 2 kVp, current range is from 50 mA to 200 Ap, frequency range from 0.1 Hz to 500 kHz. Measurement uncertainty is usually in the range from 0.05% to 0.2% of the reading. An example of this kind of instrument is the Voltech power analyzer (9).

Example of a measurement circuit for efficiency measurements of induction motors driven by the “Adjustable speed drive” is shown in Fig. 13. This kind of measurements must follow requirements described in appropriate standards, for example, IEEE Std. 112, and it is feasible to make these measurements in a computer-controlled system, as shown in Fig. 13. In this system, as the load of the motor under test a programmable dynamometer made of a dc generator with digitally controlled field current is used. The dynamometer is adjusted to the required power range by changing the load of the dc generator. The input power is measured with the power analyzer, while the mechanical power is measured with the torque measurement system. Measurement of temperature in up to 8 points of the motor under test is necessary in some measurement protocols.

MECHANICAL POWER METERS

A block diagram of a mechanical power meter used in the measurement system of Fig. 13 is shown in Fig. 14. A torque transducer is inserted in the shaft connecting the motor under test with the dynamometer. The torque transducer consists of a reference part of the shaft (reference diameter and material) with a strain gauge bridge (R_1 , R_2 , R_3 , and R_4) affixed to its surface in such a way that the output signal from the bridge is proportional to the torque. The bridge is used as an ac bridge at carrier frequency in the range from 1 kHz to 20 kHz. Supply voltage to this bridge and the output signal from the bridge are transmitted between the stationary measurement system and

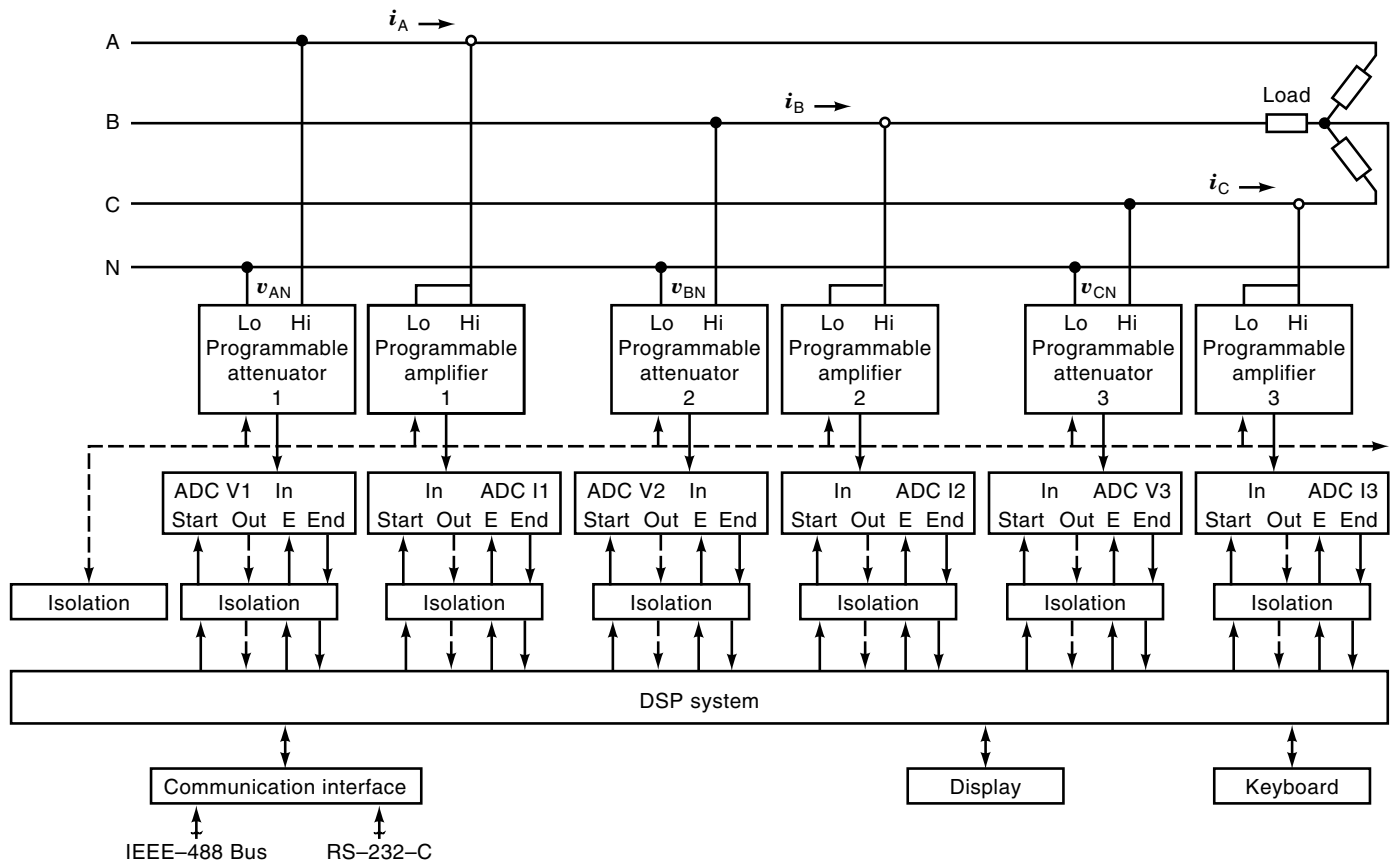


Figure 12. Three-phase electronic power analyzer using digital computation method. Successive approximation digital-to-analog converters are used to acquire samples of phase voltages and phase currents. Signal processing is used to compute parameters of a three-phase system.

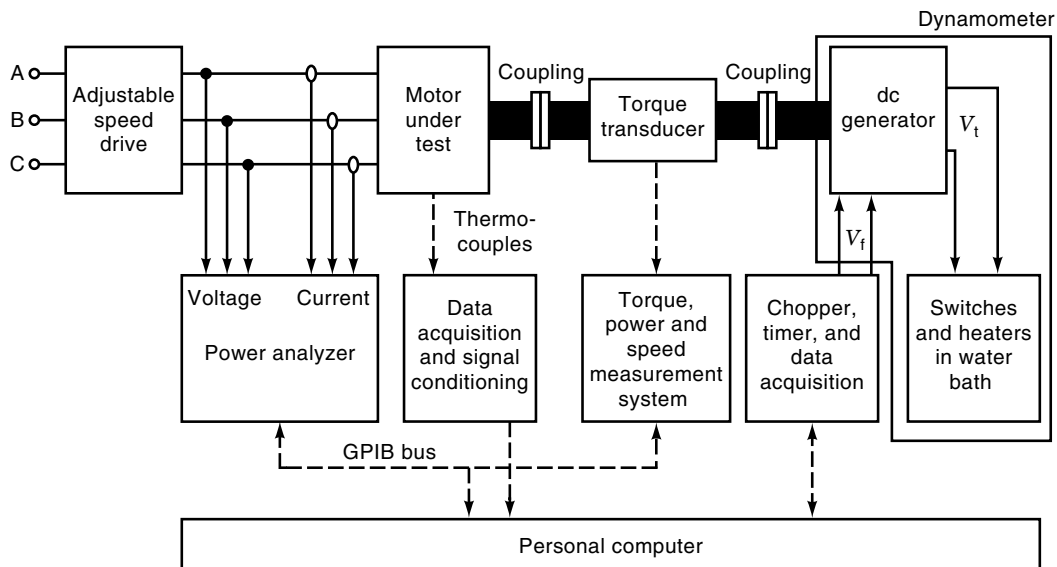


Figure 13. Automatic system for measuring efficiency of electric motors. The dc generator is used as a programmable load for the motor under test. Various test programs may be carried out in this system.

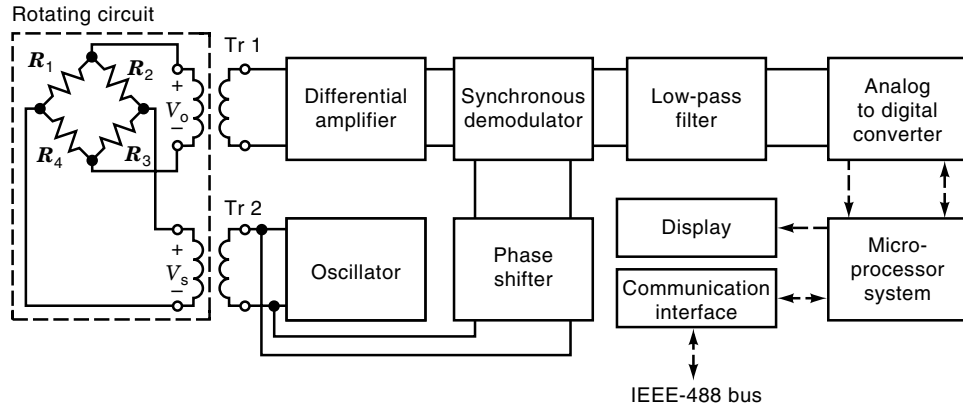


Figure 14. Mechanical power meter for measurement of power transmitted by a shaft. A torque transducer that includes a rotating strain gauge bridge needs to be installed in the shaft.

the rotating shaft with two rotary transformers. The ac bridge is used because of very large interference of noise to the sensing part of the meter. The output voltage from the bridge is a very low level signal, hence a large voltage gain amplifier must be used before the signal can be rectified with the synchronous demodulator. The synchronous demodulator uses a reference voltage from the same oscillator that produces the excitation for the bridge and has the ability of rejecting dc offset and the noise, and passing only the signal of frequency of the oscillator. As the synchronous demodulator the integrated circuit AD630 from Analog Devices, described in the Analog Devices databook (10), may be used. The signal from the oscillator that is used as the reference for the synchronous demodulator must have its phase shift adjusted accordingly with the phase of the output signal from the bridge. The dc value of the output voltage from the synchronous demodulator (average value of the output voltage) carries the information of changes of resistances of the bridge. The dc value of the output voltage from the synchronous demodulator is obtained by using a low-pass filter. The AD630 with ease allows the measurement of change of bridge resistances of 0.5 ppm, which corresponds to a 3.2 mV change in the output voltage from the low-pass filter. The dc output voltage from the low-pass filter is proportional to torque. This output voltage is converted to a corresponding digital value with the ADC. In addition to torque, the velocity of the shaft is measured and the results are stored in the microprocessor system. Mechanical power is computed in the microprocessor system based on the measured torque as

$$P_m = \frac{V_{avg} T}{63025} \quad (18)$$

where P_m is the mechanical power in horsepower, T is the torque in pounds per inch, and V_{avg} is the average shaft velocity in revolution per minute. An example of the mechanical power meter is the torque transducer and mechanical power instrument from the Himmelstain company (11).

POWER METERS FOR HIGH FREQUENCY (HF) ELECTRONIC SYSTEMS

Electronic systems employ a very wide frequency range, from dc to hundreds of gigahertz, and a great variety of waveforms,

ranging from continuous wave (CW unmodulated and modulated sinewaves) to complex pulse trains. Values of measured power range from tenths of picowatt to several kilowatts. A number of measurement methods are used to measure the power of these diverse signals. In electronic systems that operate in the radio frequency (RF) and microwave frequency ranges from about 100 kHz to 110 GHz, active power is the most frequently measured quantity. In this frequency range, power meters belong to the group of typical instruments used for evaluating electronic systems. An HF power meter consists of a power sensor, which converts active power of a RF or microwave signal to a dc or low-frequency signal proportional to the active power and a power meter (also called an RF meter). For a given kind of signal and its range of frequency, a right power sensor must be selected from a set of more than twenty different kinds of power sensors. One power meter may work with a few power sensors that were designed specifically for this power meter. Therefore, power sensors and the power meter must be of the same brand, say Hewlett-Packard or Rhode & Schwartz. Three methods of power sensing are used and they are described below. Results of power measurements are expressed as absolute power in watts (W) or as relative power in dBm. The definition of dBm is

$$\text{dBm} = 10 \log_{10} \left(\frac{P}{1\text{mW}} \right) \quad (19)$$

where the active power P is expressed in milliwatts. The active (or average) power in RF and microwave power measurements is averaged over many periods of the signal and is defined as

$$P_{avg} = \frac{1}{nT_1} \int_{t_0}^{t_0+nT_1} v(t)i(t) dt \quad (20)$$

where T_1 is the period of the lowest frequency component of $v(t)$ and $i(t)$. The averaging time is typically in the range from several hundredths of a second to a few seconds and is much greater than T_1 ; therefore it is not essential to integrate over an integer number of periods of $v(t)$ and $i(t)$. For pulse type signals,

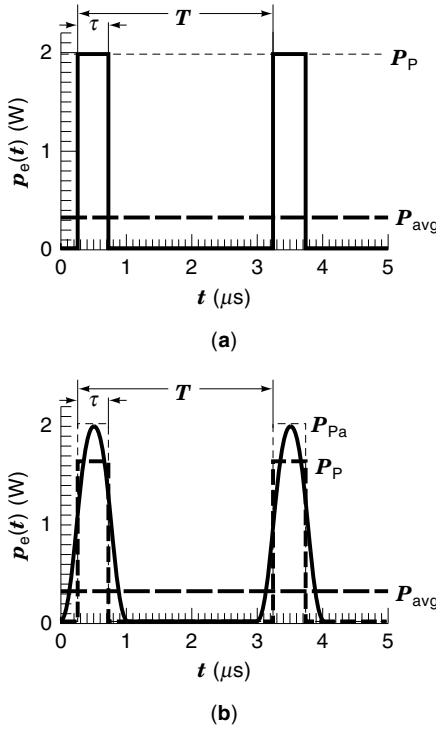


Figure 15. Envelopes of power pulses, (a) rectangular power pulses, (b) sinusoidal power pulses.

the pulse power is defined as

$$P_P = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} v(t)i(t) dt \quad (21)$$

where τ is the pulse width, defined as a time from the instance when the rising edge rises to 50% of the pulse amplitude to the instance when the falling edge falls to 50% of the pulse amplitude. Two envelopes of power pulses are shown in Fig. 15. The envelope of power shown in Fig. 15(b) represents a pulse train of single periods of the function $[1 - \cos(\omega t)]$. This shape of pulses has been chosen because it is easy to envision the three values of power for this signal [which are depicted in Fig. 15(b)]. Measurements of pulse power as defined by Eq. (21) are difficult to perform, therefore another definition of the pulse power is also used:

$$P_{Pa} = \frac{P_{avg}}{DT} \quad (22)$$

where DT is the duty cycle defined as

$$DT = \frac{\tau}{T} \quad (23)$$

Only for a rectangular power envelope $[p_e(t)]$ such as the one shown in Fig. 15(a) will the definitions described by Eqs. (21) and (22) give the same result. For the particular power envelope shown in Fig. 15(b), the peak value of $p_e(t)$ equals P_{Pa} . For many practical kinds of power envelopes, as for example, Gaussian pulse shape used in certain navigation systems, the peak value of $p_e(t)$ is not equal P_{Pa} . Power meters for many other kinds of complex signals, like fast digital phase-shift-keyed modulation

or multiple carrier signals, are also available. Proper selection of a power sensor and power meter adequate for the specific kind of measured signal is a critical factor in accurate power measurements.

The measured power may be the terminating power, that is the power absorbed in a load, or directional power, that is forward or reflected power. For measuring the terminating power, the measurement usually made, the power sensor is used in place of the load. The readout from the power meter is affected by errors related to inaccuracies of:

1. the HF part of the circuit, because the HF power dissipated in the power sensor is slightly different from the power that would be dissipated there if everything in the circuit is perfect (in the description that follows, the perfect parameters are called "ideal")
2. the circuit that produces dc or low frequency power equal to the HF power dissipated in the power sensor (this function is performed by a closed loop control system)
3. the measurement circuit of the dc or low frequency power

Denote the HF power actually dissipated in the load (the HF power dissipated in the power sensor) as P_1 and the ideal amount of power that would be dissipated there when the HF part of the circuit is perfect as $P_{l,ideal}$. The relationship between P_1 and $P_{l,ideal}$ is

$$P_1 = \frac{1 - |\Gamma_1|^2}{|1 - \Gamma_s \Gamma_1|^2} P_{l,ideal} \quad (24)$$

where Γ_1 is the reflection coefficient of the load (a complex number with magnitude ρ_1 and phase shift ϕ_1) and Γ_s is the reflection coefficient of the source (a complex number with magnitude ρ_s and phase shift ϕ_s). Denote the dc or low frequency power that substitutes the HF power as P_{sub} . Inaccuracy of the substitution process is represented by the effective efficiency η_e

$$\eta_e = \frac{P_{sub}}{P_1} \quad (25)$$

Recalling that

$$P_1 = P_i - P_r \quad (26)$$

where P_i is the incident power and P_r the reflected power, and that the calibration factor κ_{cal} is

$$\kappa_{cal} = \frac{P_{sub}}{P_i} \quad (27)$$

the following relationship between the effective efficiency and calibration factor may be written

$$\kappa_{cal} = \eta_e(1 - |\Gamma_1|^2) \quad (28)$$

If the result of the HF power measurement displayed by the power meter is denoted as P_m , its departure from the $P_{m,ideal} = P_{sub}$ is expressed by the equation

$$P_m = \xi P_{m,ideal} + P_{m,os} \quad (29)$$

where ξ represents the gain coefficient (ideally $\xi = 1$) and $P_{m,os}$ represents the offset error ($P_{m,os}$ may be a positive or negative number, ideally $P_{m,os} = 0$). Hence, the relationship between the readout from the power meter and the ideal value of HF power delivered to the power sensor is

$$P_m = \xi \kappa_{cal} \frac{P_{l,ideal}}{|1 - \Gamma_s \Gamma_l|^2} + P_{m,os} \quad (30)$$

A more detailed description of the accuracy of HF power measurement may be found in the Hewlett-Packard Application Note 64-1A (12).

Power sensor circuits are quite complex and only basic principles of their operation are described below. Every power sensor is characterized by a set of calibration coefficients that may be stored in an EEPROM installed in the power sensor unit. The power meter reads these calibration coefficients and automatically implements them in computations of the result of a measurement. Peak power analyzers for power measurements in pulsed RF and microwave electronic systems are also available. A broader description of RF and microwave power measurements may be found in Hewlett-Packard's Application Notes 64-1A (12) and 64-4A (13). Parameters of power sensors and power meters currently available from the Hewlett-Packard company may be found in the product catalog (14) and listed there publications. Power sensors and RF meters offered by the Rhode & Schwartz company may be found in the Tektronix company's catalog of measurement products (15). The author also referred to unpublished materials written by Barp (16).

Thermistor Power Sensors and Power Meters

Bolometers are power sensors that implement devices whose resistance is a function of temperature. The increment of temperature of such a device is proportional to the increment of active power dissipated in it, no matter what kind of signal (ac or dc) is delivering the power. Hence, bolometers possess a power substitution ability that is beneficial in transferring standards. There are two types of bolometers, barretters and thermistors. A barretter is a very small power (about 10 mW) metallic resistor. Power sensors using barretters are not manufactured any more because the allowed overload of a barretter is small and its damage may readily occur. Power sensors employing thermistors are still fabricated, but they are not typical power sensors now. Their frequency range is limited and their impedance matches are not good. The main use of thermistor power sensors is as power transfer standards to references, for example, at the National Institute of Standards and Technology (NIST) in Boulder, CO, and in the round robin procedures. Thermistors are mounted in either coaxial or waveguide structures. Figure 16 shows the circuit of the RF thermistor power sensor and Fig. 17 shows the measurement circuit of the power meter. Thermistors TRF1 and TRF2 are heated by the input RF signal, v_{RFin} , and a portion of the dc supply voltage of the measurement circuit, V_{RF} . The RF signal is applied through the coupling capacitor C_C to thermistors TRF1 and TRF2 connected in parallel (for the RF signal, the bypass capacitor C_{b1} connects the terminal RFB2 to the ground). Thermistors TRF1 and TRF2 are connected to the dc bridge measurement circuit using terminals RFB1 and RFB2, therefore in the dc bridge they are connected in series. Equivalent resistance of the parallel connection of R_{TRF1} and R_{TRF2} equals a matching impedance

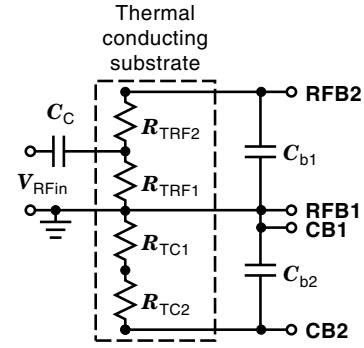


Figure 16. HF thermistor power sensor. Two thermistors TRF are used to convert HF power to an equivalent dc or low frequency voltage. Two thermistors TC are used to compensate for changes of ambient temperature.

for the transmission line (50Ω or 75Ω). This equivalent resistance is maintained constant (temperature of TRF1 and TRF2 is maintained constant) and independent of the amount of RF power delivered to it as well as of ambient temperature variations. The dc voltage across thermistors TRF1 and TRF2 equals $0.5V_{RF}$. The power meter keeps the sum of the RF power and dc power dissipated in R_{TRF1} and R_{TRF2} at a constant value. In order to simplify the description, first ignore the presence of R_{TC1} and R_{TC2} in the circuit. For $v_{RFin} = 0$, the V_{RF} has its largest value (the whole power dissipated in R_{TRF1} and R_{TRF2} is delivered from the dc voltage source) corresponding to the full scale power. When v_{RFin} produces in R_{TRF1} and R_{TRF2} one half of the full scale power, then V_{RF} delivers one half of the full scale power. When the v_{RFin} delivers the full scale power, then the $V_{RF} = 0$. The second pair of thermocouples, TC1 and TC2, is used to compensate for variations of ambient temperature. All four thermistors are mounted on a thermal conducting substrate and they are in the same temperature. The bypass capacitor C_{b2} ensures that there is no RF signal between terminals CB2 and CB1. Thermistors TC1 and TC2 are a part of the compensation bridge supplied by the dc voltage V_C . For $I_{AZ} = 0$, the voltage across thermistors TC1 and TC2 connected in series equals $0.5 V_C$. Voltage V_C changes when the ambient temperature changes. The resistance temperature coefficient of thermistors is negative, thus if the ambient temperature

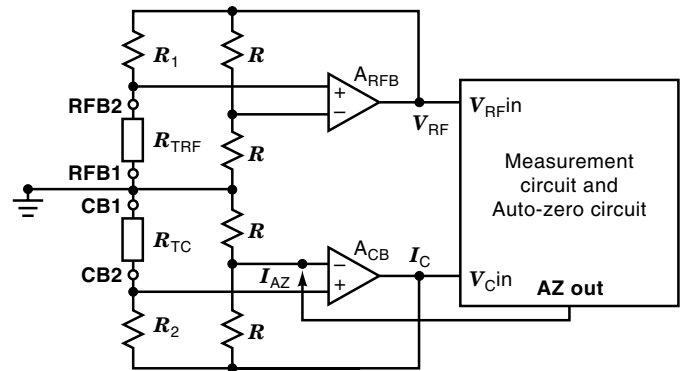


Figure 17. Thermistor HF power meter. This simplified circuit diagram shows a power measurement system formed by the power meter and the power sensor.

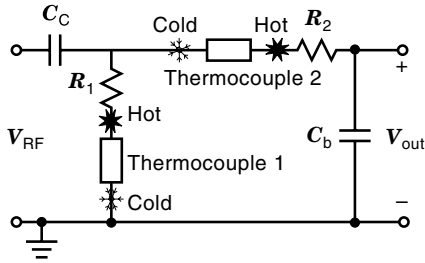


Figure 18. HF thermocouple power sensor. A pair of thermocouples is used to convert the HF power to an equivalent dc or low frequency voltage.

increases, the R_{TC1} and R_{TC2} would like to decrease. However, instead of that the V_C decreases, less power is delivered from the V_C and the R_{TC1} and R_{TC2} remain constant. The same mechanism changes V_{RF} when the ambient temperature varies. Because of that, V_{RF} must be larger than initially was assumed. The readout of the power meter is proportional to the difference ($V_C - V_{RF}$), hence V_{RF} may be made larger than the full scale V_{RFIn} (rms value of v_{RFIn}). To adjust the zero of the power meter, at $v_{RFIn} = 0$ the auto-zero circuit sets the auto-zero current I_{AZ} to a value for which the output voltage from the power meter, V_{OUT} , that is proportional to the difference ($V_C - V_{RF}$), equals zero. For increasing v_{RFIn} the V_{RF} decreases and V_{OUT} increases in proportion with the increase of V_{RFIn} .

Thermocouple Power Sensors and Power Meters

Thermocouples exhibit an inherent square-law transfer characteristic, where square-law means that the dc output voltage is proportional to the square of the input rms voltage across a reference resistance, that means to the input RF power. Thermocouple sensors are heat-based devices, consisting of a resistor that dissipates the measured power of any kind of an electric signal and the thermocouple that converts its temperature to a dc voltage. Thermocouple sensors allow power meters to be made with full scale power from $0.3 \mu\text{W}$ and with a small SWR (standing wave ratio). Thermocouple sensors are fabricated using combined thin-film and semiconductor technologies. A simplified circuit diagram of a thermocouple sensor is shown in Fig. 18. The bypass capacitor C_b is a short circuit for RF signal, thus the v_{RF} is applied to the two pairs of resistor-thermocouple connected in parallel. The equivalent resistance of this parallel connection constitutes the 50Ω or 75Ω termination resistance of the RF transmission line. However, from

the standpoint of the dc output voltage V_{OUT} , these thermocouples are connected in series, hence the V_{OUT} equals two times a voltage from a single thermocouple. Typical thermocouples' sensitivities are $250 \mu\text{V}/^\circ\text{C}$ and $160 \mu\text{V}/\text{mW}$. The transfer characteristic of the thermocouple sensor (V_{OUT} as a function of P_D) is slightly nonlinear, because its sensitivity is a little lower for large powers. The thermocouple sensor's V_{OUT} is approximately 160 nV for $1 \mu\text{W}$ of power dissipated in it, consequently the power sensor and power meter are considerably complex. Figure 19 shows a simplified block diagram of the power sensor attached to the power meter. Dc signals at such a low level as the V_{OUT} cannot be transmitted over a cable. Also, in the power meter they need to be amplified significantly. A chopper amplifier is used to obtain the required large amplification of the V_{OUT} . One-half of the gain is provided by the ac amplifier located in the power sensor enclosure, and the other one half of the gain is furnished by an ac amplifier located in the power meter enclosure. The ac output voltage from the amplifier is demodulated synchronously with the chopping signal. An auto-zero feature is implemented in the measurement scheme too. Thermocouple sensors do not have the dc substitution feature; hence it is necessary to provide a reference RF power source for calibration purposes. The output power of this reference power source is controlled with total uncertainty better than $\pm 1\%$.

Diode Sensors and Power Meters

Schottky diodes are used to build diode sensors. $P-N$ junction diodes have much smaller bandwidth than Schottky diodes, without bias have extremely high impedance, and when biased they produce a large amount of noise and thermal drift. Schottky diodes specially designed for applications in power sensors have a low potential barrier of the metal-semiconductor junction, which results in a forward voltage of about 0.3 V . They have a very large bandwidth that is within the microwave frequency range. Current-voltage ($I-V$) characteristic of these diodes may be divided into three regions: (1) the square-law region for very small voltages across the junction, (2) a linear region for large voltages, and (3) the transition region from the square-law region to the linear region. The $I-V$ characteristic is described by the approximate equation

$$I = I_S \left[\exp \left(\frac{V}{nV_T} \right) - 1 \right] \quad (31)$$

where I_S is the saturation current, n is a coefficient specific for a given type of a Schottky diode that compensates the weak

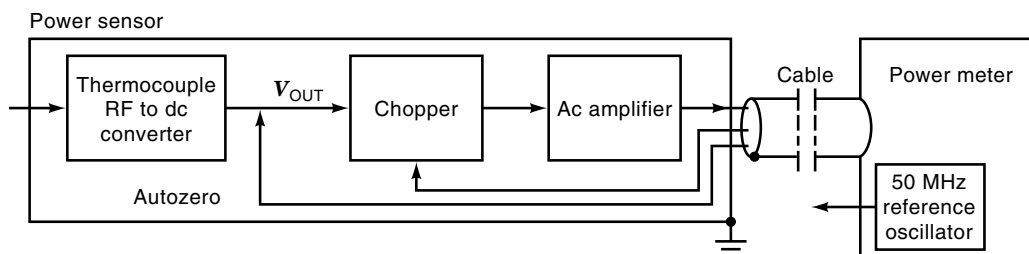


Figure 19. Thermocouple HF power meter. This simplified block diagram illustrates functions of the power meter that together with a HF thermocouple power sensor forms a power measurement system.

dependence of the saturation current on voltage V in the accurate equation, and V_T is the thermal voltage

$$V_T = \frac{kT}{q} \quad (32)$$

In Eq. (32), k is Boltzmann's constant, T is the temperature in kelvins, and q is the charge of an electron. Derivation of the Schottky diode equation may be found in the book by Muller and Kamins (17). Figure 20(a) shows the I - V characteristic of a Schottky diode designed for power sensor applications. The constant n is slightly greater than 1 and for making the graph $n = 1.07$ has been used. Schottky diodes fabricated for applications in power sensors have much larger I_S than typical Schottky diodes. The exponential term in Eq. (31) may be expanded into the infinite series form

$$\exp\left(\frac{V}{nV_T}\right) - 1 = \frac{1}{1!} \frac{V}{nV_T} + \frac{1}{2!} \left(\frac{V}{nV_T}\right)^2 + \frac{1}{3!} \left(\frac{V}{nV_T}\right)^3 + \dots \quad (33)$$

For $V \ll nV_T$ Eq. (31) may be approximated as

$$I_q = I_S \left[\frac{V}{nV_T} + \frac{1}{2} \left(\frac{V}{nV_T}\right)^2 \right] \quad (34)$$

where I_q is the current I approximated by the quadratic Eq. (34). Figure 20(b) shows graphically the error of this approximation (δ) for $-20 \text{ mV} < V < +20 \text{ mV}$, defined as

$$\delta = \frac{I_q - I}{I} 100\% \quad (35)$$

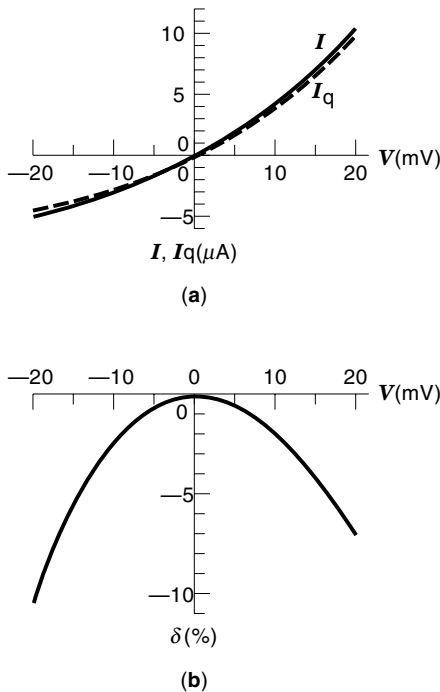


Figure 20. I - V characteristic of a diode power sensor, (a) I - V characteristic of a Schottky diode, (b) error of the square-law relationship.

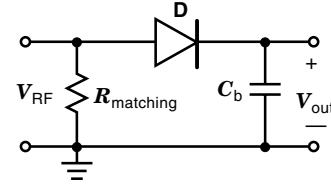


Figure 21. HF diode power sensor.

Figure 21 shows the simplified circuit of an unbiased diode for detecting low level RF and microwave signals. A matching resistor is used to obtain the right termination impedance for a transmission line. In practical circuits two diodes are used in the circuit of a power sensor. When the diode power sensor detects the smallest power level (limited by noise of the diode) of 100 pW the dc output voltage $V_{OUT} = 50 \text{ nV}$. To handle this extremely small signal, a high-gain chopper amplifier is used in the power meter. Part of this amplifier is located in the enclosure of the power sensor, as well as circuitry for temperature compensation and the EEPROM that stores calibration constants of the power sensor. The diode power sensors do not have the dc substitution feature, hence it is necessary to provide a reference RF power source for calibration purposes. More information on RF and microwave power sensors and power meters may be found in Hewlett-Packard Journal papers (18–20).

Uncertainty of a HF power measurement is associated with circuit imperfections of the power sensor and the power meter. Because of the complexity of these circuits, usually main partial uncertainties are specified for a power sensor and for a power meter. It is up to the user of this equipment to decide which partial uncertainties need to be considered for a specific measurement and how the total uncertainty needs to be computed. The total uncertainty may be computed as the worst case uncertainty, root sum of the squares (RSS), or according to the ANSI/NCSL 540Z-2 guide. Typical range of combined uncertainty of power sensors is from 3% to 8%, and the range of combined instrumentation uncertainty is from 1% to 5%.

Optical Power Sensors and Optical Power Meters

A simplified block diagram of an optical power sensor and its connection to an optical power meter is shown in Fig. 22. In the optical power sensor, the optical power, P_O , of the measured lightwave is converted to an electrical current, I_D , following to the relationship

$$I_D = r_D(\lambda) P_O \quad (36)$$

where $r_D(\lambda)$ is the conversion factor or responsivity, that is, a function of wavelength, λ . The output current of the photodetector is very small and needs to be significantly amplified and conditioned before the result of the measurement may be displayed. It is necessary to place a suitable preamplifier (usually two preamplifiers must be used in order to cover the entire dynamic range) in the enclosure of the optical power sensor. One of these preamplifiers has smaller gain but wide bandwidth, while the other has large gain and therefore must have small bandwidth. Depending on measured optical signal level and specific kind of power (average optical power, peak

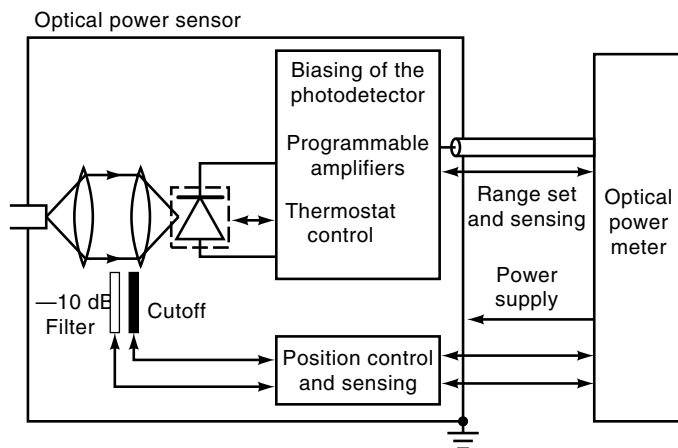


Figure 22. Optical power sensor and optical power meter. An optical power meter and optical power sensor create an optical power measurement system. The optical power meter works with a set of dedicated optical power sensors.

optical modulation power), the control system of the optical power meter chooses the proper preamplifier, its gain and bandwidth. In order to keep the photodetector at a constant temperature, the detector is placed in a thermostat (both heating and cooling techniques are used). If a chopper amplifier is used in an optical power meter, then an optical chopper is used. Disadvantages of optical choppers include reduced bandwidth, slow sampling rate, and possible reflection effects from the chopper. If a dc amplifier is used, constant temperature of the amplifier needs to be secured too. The lightwave may be completely cut off from the optical-to-electrical transducer, which is necessary for zeroing the measurement system. Dynamic range of the optical power meter increases with the use of a reference optical attenuator that may be inserted across the light ray. Typical attenuation of the optical attenuator is -10 dB. Optical power meters cover power ranges from $+27$ to -110 dBm, with total uncertainty range from 2.5% to 5%, and operate in the wavelength range of 400 nm to 1750 nm (contingent to use of a set of optical power sensors). More characteristics of optical power meters may be found in catalogs of Hewlett-Packard company (14,21), and Tektronix company (15). There are ordinary optical power meters, typically in the form of hand-held instruments, or lightwave multimeters that include a power measurement mode of operation. Measurements may be made in lightwave systems that implement lasers or light-emitting diodes (LED). Optical power meters are used to make measurements of absolute power of active components such as lasers or LEDs, insertion losses, and reflections in optical systems. Relative power measurements (with two-channel optical power meters or in single-channel optical power meters related to a specific measurement result) are also possible. Typically the average value of optical signals is measured, but other measurements, like peak power (high and low levels) over a large bandwidth from dc to 250 MHz, are also offered. Optical power meters are broadband receivers, integrating power levels over the wavelength specific for the optical power sensor that is applied. For optical component characterization at specific wavelength, a selective wavelength source must be used. Typical photodetectors used in optical power meters include large-area photodiodes, thermopiles, and pyroelectric crystals. To cover the entire fiber

optic wavelength range, three semiconductor materials are used to make photodetectors. Silicon detectors operate in the range from 400 nm to 1020 nm, germanium detectors cover the range from 900 nm to 1650 nm, and indium gallium arsenide (InGaAs) detectors handle the range from 800 nm to 1700 nm. Detectors made of these materials have some characteristics superior over the other two materials, hence selection of the best detector for a given application is essential. A broad description of topics related to optical power meters may be found in Hewlett-Packard publications (22–27) and in Derickson's book (28).

BIBLIOGRAPHY

1. S. F. Legowski, A. H. M. S. Ula, and A. M. Trzynadlowski, Instantaneous power as a medium for the signature analysis of induction motors, *IEEE Trans. Ind. Appl.*, **32**: 904–909, 1996.
2. D. R. Cunningham and J. A. Stuller, *Basic Circuit Analysis*, Boston, MA: Houghton Mifflin, 1991.
3. W. Kidwell, *Electrical Instruments and Measurements*, New York: McGraw-Hill, 1969.
4. I. F. Kinnard, *Applied Electrical Measurements*, New York: Wiley, and London: Chapman & Hall, 1956.
5. *Valhalla 2100/2101 Digital Power Analyzer, Operation Manual*, San Diego: Valhalla Scientific.
6. *Model 40 Power Harmonics Meter and 41 Power Harmonics Analyzer, Users Manual*, Fluke Corporation, 1994.
7. *1997/8 Catalog*, Fluke Corporation, Everett, WA 98206.
8. B. Drafts, Current sensing solutions for power conversion and intelligent motion applications, *Power Conversion & Intell. Motion*, **21** (9): 53–58, 1995.
9. *PM300A Universal Power Analyzer, User Manual*, Voltech Instruments Ltd, Abingdon, Oxon, UK, 1996.
10. *Linear Product Databook*, Analog Devices 1990/91, pp. 6-39–6-46.
11. *MCRT Non-Contact Strain Gage Torquemeters and Model 66032 Mechanical Power Instrument*, S. Himmelstein and Company, Hoffman Estates, IL.
12. Application Note 64-1A, *Fundamentals of RF and microwave power measurements*, Palo Alto, CA: Hewlett-Packard.
13. Application Note 64-4A, *4 steps making better power measurements*, Palo Alto, CA: Hewlett-Packard.
14. D. Barp, Measurements in telecommunication systems, unpublished article, Electr. Eng. Dept., Univ. Wyoming, 1998.
15. *Test and Measurement Catalog*, 1998, Palo Alto, CA: Hewlett-Packard.
16. *Measurement Products Catalog*, 1996, Tektronix Inc., Pittsfield, MA 01202.
17. R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., New York: Wiley, 1986.
18. D. Scherer et al., The peak power analyzer, a new microwave tool, *Hewlett-Packard J.*, **43** (2): 81–89, 1992.
19. M. C. Fischer, M. J. Schoessow, and P. Tong, GaAs technology in sensor and baseband design, *Hewlett-Packard J.*, **43** (2): 90–94, 1992.
20. D. L. Barnard, H. Black, and J. A. Thalmann, Automatic calibration for easy and accurate power measurements, *Hewlett-Packard J.*, **43** (2): 95–100, 1992.
21. *Lightwave Test and Measurement Catalog*, 1998, Palo Alto, CA: Hewlett-Packard.
22. C. Hentschel, *Fiber Optics Handbook*, Hewlett-Packard, GmbH, Boeblingen Instruments Division, Germany, 1983.

23. W. Berkel et al., A versatile, programmable optical pulse power meter, *Hewlett-Packard J.*, **36** (1): 18–27, 1985.
24. J. Becker, Detectors for optical power measurements, *Hewlett-Packard J.*, **38** (2): 16–21, 1987.
25. B. Maisenbacher and W. Reichert, A lightwave multimeter for basic fiber optic measurements, *Hewlett-Packard J.*, **42** (2): 58–63, 1991.
26. J. Rivoir, H. Schweikardt, and E. Muller, Design of a series of high-performance lightwave power sensor modules, *Hewlett-Packard J.*, **42** (2): 63–69, 1991.
27. C. Hentschel, Calibration of fiber optic power meters, *Hewlett-Packard J.*, **42** (2): 70–72, 1991.
28. D. Derickson, *Fiber Optic Test and Measurement*, Upper Saddle River, NJ: Prentice-Hall, 1998.

STANISLAW F. LEGOWSKI
University of Wyoming

WATTMETERS

The wattmeter is an instrument that measures the average power P_0 entering a one-port. Under periodic conditions, P_0 is defined according to the following relationship:

$$P_0 = \frac{1}{T_1} \int_{t_A - T_1}^{t_A} v(t)i(t) dt \quad (1)$$

where $i(t)$ denotes the instantaneous current entering the port, $v(t)$ represents the instantaneous voltage across the port, and T_1 represents the period of the instantaneous power $p(t) = v(t)i(t)$. Under nonperiodic conditions, Eq. (1) can be generalized by introducing the limit for T_1 tending to infinity. Equation (1) suggests a method for measuring the mean power P_0 by applying the product of the input signals $v(t)$ and $i(t)$, to a linear time invariant (LTI) circuit whose impulse response $h(t)$ is shown in Fig. 1(a) (1). Recalling the modulus of the Fourier transform $H(fT_1)$ of $h(t)$ [Fig. 1(b)], when the condition $fT_1 = k$, with k an integer, is satisfied, only the direct component P_0 of the instantaneous power is present at the output, and it can be measured easily. In order to avoid this synchronization requirement, a low pass filter (LPF) that properly attenuates all the alternate spectral components is normally used instead of the circuit with impulse response $h(t)$. The input conditioning circuits are necessary in order to convert the input signals, both in type and amplitude, so that they can be compatible with the multiplier to optimize the multiplier performance and to insulate the operator from the measurement section when needed. These circuits, together with the multiplier and the LPF, characterize the sensitivity, the bandwidth, and the accuracy of the instrument. Accuracy and sensitivity are also influenced by the instrument which measures P_0 .

The electrodynamic wattmeter is an electromechanical implementation of the block diagram of Fig. 2. In fact, the instantaneous operating torque of the moving system is proportional to the product of the instantaneous currents in the fixed and moving coils. The effect of the couples that oppose the motion (the restoring torque, the mechanical and electromagnetic damping torques, and the inertial torque of the moving system) is equivalent to a low-pass filter with a natural frequency on the order of 1 Hz. This instrument is particularly suitable for ac power measurements at line frequency and usually has

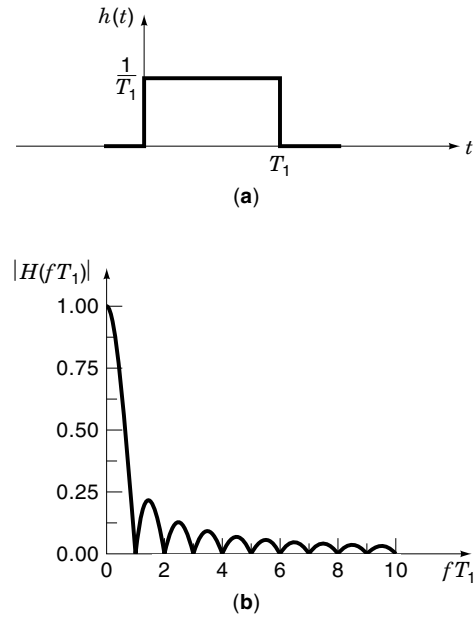


Figure 1. (a) Impulse response and (b) its Fourier transform in modulus of the linear time-invariant (LTI) circuit after the multiplier of a wattmeter.

an uncertainty of 0.5% at full scale. Precision electrodynamic wattmeters, having an uncertainty lower than 0.01% at line frequency, are also available, along with wide-band wattmeters characterized by a frequency range up to a few kilohertz and uncertainty of 0.1%. Other techniques of multiplication were recently introduced (2), and some of them have been implemented in these instruments. One of them is the quarter-square technique. When both the quantities at the output of the conditioning circuits are currents, P_0 can be derived by measuring two root-mean-square (rms) values. In fact, by recalling the relationship

$$i_1 i_2 = \frac{1}{4} [(i_1 + i_2)^2 - (i_1 - i_2)^2] \quad (2)$$

when the currents i_1 and i_2 to be multiplied are a function of the quantities that the instantaneous power depends on, the mean power can be deduced from the difference of the rms values at the output of two thermal converters to which the quantities $(i_1 + i_2)$ and $(i_1 - i_2)$ are applied (3,4,5). However, the most common multiplier technique used in mean power measurement is time division (6–10). Both accuracy and bandwidth featured by

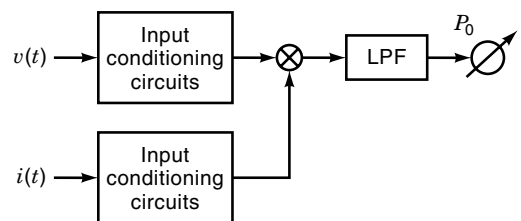


Figure 2. Functional block diagram of a wattmeter, showing that this instrument is basically made of input conditioning circuits, a multiplier, and a low-pass filter (LPF).

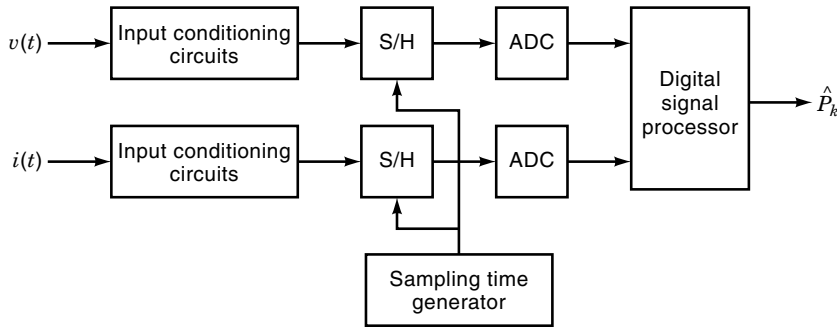


Figure 3. Block diagram of the sampling wattmeter.

commercial instruments using this technique do not substantially differ from those of the best electrodynamic instruments. The dc output is normally applied to a digital voltmeter in order to obtain a digital output.

The integral in Eq. (1) can be approximated by a weighted sum of the product of successive N samples of $v(t)$ and $i(t)$:

$$\hat{P}_k = \sum_{i=0}^{N-1} a_i v(t_{k-i}) i(t_{k-i}) \quad (3)$$

where the integer k identifies a generic discrete output $\hat{P}_k$, t_{k-i} is the i th sampling instant (x), and a_i is the value of the weighting function (11,12). This equation leads to a power meter implemented according to the block diagram of Fig. 3: The outputs of the two input conditioning circuits are simultaneously sampled through a sampling time generator (STG), digitized, and latched. The acquisition process is controlled by a digital signal processor (DSP). The two digital values are then acquired by the DSP to estimate the mean power in real time. In a digital environment, the multiplication does not present particular accuracy problems when a 32 bit floating-point DSP is used; therefore the measurement accuracy is influenced by both the adopted sampling strategy and the filtering algorithms, along with the characteristics of the S/H-ADC (sample/hold analog-to-digital converter) circuits, in particular by the frequency response of the S/Hs when they operate in the sample mode and by the resolution of the ADCs. The adopted sampling strategies are either of an equally spaced type (which can be divided into synchronous, quasi-synchronous, and asynchronous) or a random asynchronous one. The latter represents a convenient solution when a large bandwidth and high resolution are both required and relatively long measurement times are acceptable. The errors arising from the sampling strategy and the filtering procedures will be dealt with in the following sections separately from those associated with the input conditioning circuits.

Clearly, a power instrument based on Eq. (1) or (3) can also be used to implement the measurement of the rms value of each input signal.

The instruments described by the previous block diagrams can be also defined as transmission-type wattmeters, since they are designed to be connected between a source and a load. Another type of wattmeter is the absorption type, which utilizes a power sensor to terminate the transmission line in a prefixed load (usually 50 Ω) and convert the dissipated power into a dc value, which is then measured. Absorption-type wattmeters can be classified according to the type of power sensor

(thermistor, thermocouple, or diode) (13). These instruments can measure power values ranging from less than 1 nW up to some tens of watts in the frequency range from dc up to some tens of gigahertz.

EQUALLY SPACED SAMPLING WATTMETERS

If the N samples of Eq. (3) are taken at equally spaced instants T_s we can write

$$t_{k-1} = \tau_0 + (k-i)T_s \quad (4)$$

where τ_0 is the shift between the initial sampling instant and the time origin of both $v(t)$ and $i(t)$. The minimum value of the time interval T_s depends on which is the greater: the acquisition time of the S/H-ADC devices or the processing time of the digital hardware. Since the uncertainty arising from digital multiplication can be neglected, Eq. (3) can be rewritten as follows:

$$\hat{P}_k = \sum_{i=0}^{N-1} a_i p(t_{k-i}) \quad (5)$$

The periodic instantaneous power $p(t)$ can be represented by the Fourier series:

$$p(t) = \sum_{q=-\infty}^{+\infty} P_q e^{j2\pi q f_1 t} = P_0 + \sum_{\substack{q=-\infty \\ q \neq 0}}^{+\infty} P_q e^{j2\pi q f_1 t} \quad (6)$$

where $f_1 = 1/T_1$, $P_{-q} = (P_q^*)$ being the conjugate of P_q , and P_0 is the mean value of $p(t)$, that is the value of the measurand. By substituting Eqs. (6) and (4) into Eq. (5) we can write

$$\begin{aligned} \hat{P}_k &= \sum_{q=-\infty}^{+\infty} P_q \sum_{i=0}^{N-1} a_i e^{j2\pi q f_1 t_{k-i}} = P_0 \sum_{i=0}^{N-1} a_i \\ &+ \sum_{\substack{q=-\infty \\ q \neq 0}}^{+\infty} P_q e^{j2\pi q f_1 (\tau_0 + kT_s)} H(qf_1 T_s) \end{aligned} \quad (7)$$

where

$$H(fT_s) = \sum_{i=0}^{N-1} a_i e^{-j2\pi i f T_s} \quad (8)$$

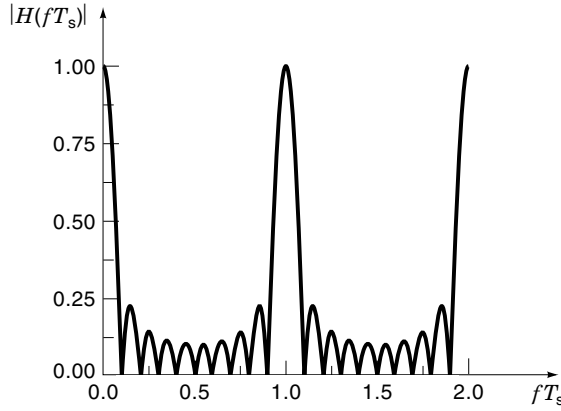


Figure 4. Frequency response in modulus of a rectangular-window finite-impulse response (FIR) filter used in a sampling wattmeter to estimate the average power.

is the frequency response of the finite impulse response (FIR) filter. The filter has a periodic response of period f_s ; moreover, for $fT_s = u$, with u an integer, it results in $H(fT_s) = \sum_{i=0}^{N-1} a_i$. In order to obtain a unity scale factor in Eq. (7), the sum of the N coefficients a_i must be unity. Plainly, the coefficients a_i must be selected to obtain a frequency response which adequately attenuates the contribution to $\hat{P}_k$ [see Eq. (3)] of each spectral component of $p(t)$. By assuming $a_i = 1/N$ (that is a rectangular window) we obtain the well-known Dirichlet–Kernel formula:

$$H(fT_s) = e^{-j\pi fT_s(N-1)} \frac{\text{sinc}(NfT_s)}{\text{sinc}(fT_s)} \quad (9)$$

where the sinc function is defined as follows: $\text{sinc } x = (\sin \pi x)/\pi x$. The frequency response of the FIR filter used is therefore a periodic function that assumes a unity value for fT_s integer and null values for $fT_s = r/N$, r being an integer which is not a multiple of N . Figure 4 shows the shape of $|H(fT_s)|$ when $N = 10$ (this low value of N was chosen only for the sake of simplicity). Therefore, if a synchronous sampling strategy is used in which f_1T_s coincides with one of the zeros of $H(fT_s)$, it follows that

$$f_1T_s = \frac{r}{N} \quad (10)$$

where r is a positive integer with no common submultiple with N , and only the spectral components of $p(t)$ whose harmonic order is an integer multiple of N contribute to the output (14–16). In fact, when $q = zN$ (z integer) in Eq. (7), we obtain

$$\hat{P}_0 = P_0 + \sum_{\substack{z=-\infty \\ z \neq 0}}^{\infty} (-1)^{zr(N-1)} P_{zN} e^{j2\pi zNf_1\tau_0} \quad (11)$$

The output $\hat{P}_0$ is a constant and therefore becomes independent of k ; also, the instrument bandwidth, referred to $p(t)$, becomes equal to $(N-1)f_1$. In fact, the first spectral component of $p(t)$ that gives a contribution to $\hat{P}_0$ is the Nf_1 one ($z = 1$). The quantities T_s and N in Eq. (10) can be independently selected, provided that both signals are under steady-state conditions in the observation (or summation) interval NT_s . This interval, in which the N samples are taken, is also called the response time

of the instrument. According to Eq. (10), this interval must be equal to an integer number of periods T_1 ($NT_s = rT_1$); therefore one of the input signals must drive the sampling time generator in order to synchronize the sampling interval T_s with the period T_1 . This is the most common solution adopted in commercial digital wattmeters. The frequency bandwidth (from dc to some hundreds of kilohertz) is obviously much greater than that of the traditional instruments based on the analog multiplication, while the accuracy is usually of the same order of magnitude (0.1%). Moreover, the availability of the sampled values of the input signals readily allows further computations to be performed to extend the number of quantities measured (for example apparent power, power factor, and so on) and to make the spectral analysis. In this case the instrument is called a power analyzer.

If, instead of the synchronization condition (Eq. 10), the following condition is imposed:

$$|H(qf_1T_s)| \leq \frac{1}{N} \quad (12)$$

for any $q \neq zN$, the contribution to $\hat{P}_k$ of the each corresponding spectral component qf_1 is not zero, as occurs in the synchronous sampling, but can be neglected if N is adequately high. Since Eq. (11) is still approximately valid, this sampling strategy is called quasi-synchronous (17). It can be shown (18) that, on the hypothesis of $f_1T_s < 0.5$, Eq. (12) is verified when

$$|NT_s - r_1T_1| \leq T_s \quad \text{for } f_1T_s \leq 0.5 \quad (13)$$

where r_1 is the nearest integer to NT_s/T_1 . In the quasi-synchronous case, the value of N must therefore be adjusted so that the observation interval NT_s results close enough to an integer number of periods T_1 to satisfy Eq. (13). A prototype implementing this procedure featured an uncertainty of less than ± 50 ppm in the frequency range 50 Hz to 1 kHz (19).

On the contrary, in the asynchronous case the product fT_s can assume any value; therefore the bandwidth is limited by the periodicity of the function $H(fT_s)$, or f_s , and the FIR filter must be selected in such a way to minimize $|H(fT_s)|$ within this interval (20).

RANDOM SAMPLING WATTMETERS

The need for synchronization to overcome the bandwidth limitations of the equally spaced sampling strategy may restrict the possible applications of wattmeters. In fact, synchronization can be conveniently used only when dealing with strictly periodic waveforms, while many applications of practical interest must deal with almost periodic “multitone” signals, which are characterized by a discrete spectrum with nonharmonically related frequencies of the spectral components. Asynchronous sampling may also be more convenient in strictly periodic cases whenever synchronization is a difficult task, either because the signals have a very broad spectrum or because the period is relatively long. The bandwidth of a wattmeter based on an asynchronous equally spaced sampling strategy is limited by the conversion time of the ADC. On the other hand, high resolution and short conversion time are two opposing requirements of an ADC. To build an asynchronous wattmeter whose bandwidth is not limited by the sampling frequency, two different

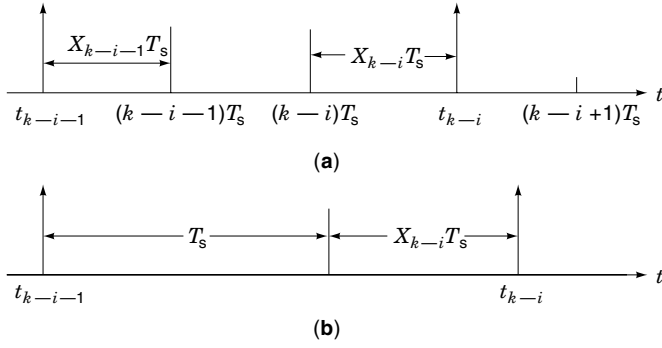


Figure 5. Time allocation of the impulse train in the case of (a) a nonrecursive and (b) a recursive random sampling technique.

random strategies (or more precisely pseudorandom, since the mean sampling time is constant), have been proposed. The first one is defined by the expression (20).

$$t_{k-i} = \tau_0 + (k-i + X_{k-i})T_s \quad (14)$$

where any sampling instant t_{k-i} is given by the sum of a periodic component $(k-i)T_s$ (k and i are integers) and a random component $X_{k-i}T_s$ [(Fig. 5(a)]. The previously introduced quantity T_s [Eq. (4)] is not correlated with any spectral component of the instantaneous power (this sampling strategy is of an asynchronous type), while X_{k-i} is the $(k-i)$ th of a set of random independent variables having a continuous uniform distribution in the interval 0 to $\pm b$. It is noteworthy that Eq. (14) includes, as a special case, the equally spaced sampling strategy [Eq. (4)]. Under the assumption of $b = 0.5$, the successive sampling instants are distributed in consecutive nonoverlapping intervals equal to T_s . Since the distance between two consecutive sampling instants can be small without limitation, while the time interval between the first and the last of a consecutive triplet of samples can never be smaller than T_s , this strategy has the drawback of requiring a pair of S/H-ADC for each input channel.

In the other random strategy, called *recursive random sampling strategy* (21), every sampling instant t_{k-i} is obtained by adding to the preceding t_{k-i-1} a predetermined constant lag T_s plus a random increment $X_{k-i}T_s$ [Fig. 5(b)]

$$t_{k-i} = t_{k-i-1} + (1 + X_{k-i})T_s \quad (15)$$

where X_{k-i} is the $(k-i)$ th of a set of random independent variables having continuous uniform distribution in a prefixed interval 0 to b . The lag time T_s has been devised so that the time interval between two successive samples is never smaller than T_s . Therefore, if its value is not less than the maximum data acquisition and processing time of the digital hardware used, a real-time instrument can be realized by using only one S/H-ADC for each channel.

PERFORMANCE ANALYSIS OF SAMPLING WATTMETERS

By substituting Eqs. (14) or (15) and (6) into Eq. (5), we obtain the output quantity $\hat{P}_k(\tau_0, \overline{CIX}_k)$, which is a function of the index k marking any output state of the instrument, of the shift τ_0 between the initial sampling instant and the time origin of $p(t)$, and of the vector of the successive N sampling random

variables $\overline{CIX}_k = \{X_k, X_{k-1}, \dots, X_{k-N+1}\}$. The incidental quantities k and τ_0 occur as nuisance parameters because they locally condition the evaluation of $\hat{P}_k(\tau_0, \overline{CIX}_k)$ by restricting it to the choice of a particular pair of values of k and τ_0 . To eliminate the influence of these nuisance parameters, we can follow the Bayesian approach (22) by regarding any pair of possible values (k, τ_0) as a realization of a bivariate random variable (k, τ_0) , with a priori probability distribution which must be selected by their intrinsic relationship to the prior information available. As far as k is concerned, it should be considered that every output state of the instrument has an equal chance of being selected. Therefore, the enumerable sets of the labels, in general with $2h+1$ values, which mark the sequence of the output states of the instrument, can be treated as a discrete random variable, uniformly distributed within the whole time interval in which these states are available. As far as τ_0 is concerned, we must distinguish between the synchronous sampling techniques and the asynchronous ones. In the asynchronous cases, the sampling time generator of Fig. 4 is not controlled by any input signal; therefore any realization of τ_0 is independent of the instantaneous power $p(t)$, and since it is strictly related to the turn-on instant of the instrument, it belongs to a continuous set of possible equiprobable values distributed within any time interval. In the synchronous sampling technique, τ_0 depends on the synchronizing circuit that generates the sequence of the sampling instants as a function of one of the input signals. As a consequence, τ_0 is a function not only of the synchronizing circuit used but also of $p(t)$. If we want to take into account all possible states of τ_0 , we should have to foresee all the possible synchronizing operations for the same unknown signal and establish the possible functional relations which connect τ_0 to the signal itself. Obviously, this is practically utopian, and it is convenient to simulate an equivalent configuration in which τ_0 may be assumed as a realization of a continuous set of values distributed within some time interval T . To sum up, the a priori ignorance concerning the nature of the parameter τ_0 can be overcome in both techniques by postulating a uniform distribution in a generic time interval T (23,24).

An appropriate characterization of the output uncertainty of the sampling wattmeters can be obtained by evaluating the statistical parameters of the output $\hat{P}_k$, i.e., the mean value $E\{\hat{P}_k\}$ and the mean square error $E\{(\hat{P}_k - P_0)^2\}$. To incorporate all the possible a priori chances and also to avoid the influence of the conventional time origin on the instrument performance, the number $2h+1$ of the output states, and the excursion T of the initial shift τ_0 must be sufficiently large and must theoretically tend to infinite. Therefore we consider the asymptotic statistic parameters, that is the asymptotic mean

$$\bar{P} = \lim_{\substack{h \rightarrow \infty \\ T \rightarrow \infty}} E\{\hat{P}_k\} \quad (16)$$

and the asymptotic mean square error

$$e^2 = \lim_{\substack{h \rightarrow \infty \\ T \rightarrow \infty}} E\{(\hat{P}_k - P_0)^2\} \quad (17)$$

It can be shown (20,21,23,24) that the output of the instruments set up with one of the previously described sampling strategies is asymptotically unbiased ($\bar{CIP} = P_0$), and consequently, the asymptotic mean square error coincides with the asymptotic variance $e^2 = \sigma^2$, whose final expression can always

be expressed as follows:

$$\sigma^2 = 2 \sum_{q=1}^{\infty} |P_q|^2 W^2(qf_1 T_s) \quad (18)$$

According to the superposition principle, this equation shows that the contribution of the squared rms value $2|P_q|^2$ of each harmonic component of the instantaneous power to the asymptotic variance is weighted by the coefficient $W^2(qf_1 T_s)$. The sequence of the weighting coefficients $W^2(f_1 T_s)$, $W^2(2f_1 T_s)$, ... can be derived from a continuous weighting function $W^2(f T_s)$ by determining its values at the successive occurrence points $f_1 T_s$, $2f_1 T_s$, ..., that is the products between each instantaneous-power spectral frequency and the constant lag T_s . Since Eq. (18) is always valid, we can conclude that a corresponding weighting function can be associated with any sampling strategy, recursive or nonrecursive, random or equally spaced, asynchronous or synchronous. Hence the behavior of the weighting function $W^2(f T_s)$ as a function of $f T_s$ and the possible occurrence points at which the weighting coefficients are evaluated completely describe the performance of any selected sampling strategy with the associated filtering algorithm.

By considering an N -point rectangular window to realize the FIR filter, it can be shown that the weighting function for the sampling strategy defined by Eq. (14) can be expressed as follows:

$$W^2(f T_s) = \frac{1}{N} + \text{sinc}^2(2bf T_s) \left[|H(f T_s)|^2 - \frac{1}{N} \right] \quad (19)$$

where $|H(f T_s)|$ is the modulus of the frequency response of the rectangular window [Eq. (9)]. When $f = 0$ it results in $W^2(0) = 1$. For $b = 0$, an equally spaced sampling, we obviously obtain $W^2(f T_s) = |H(f T_s)|^2$; therefore all the previous considerations of this sampling strategy can be deduced also considering the weighting function. For $b > 0$ and $f T_s \geq 1$, the shape of $W^2(f T_s)$ depends significantly on the given value of b . By selecting $b = 0.5$, when the random sampling is distributed within an interval equal to the time constant T_s , we obtain

$$W^2(f T_s) \cong \frac{1}{N} \quad \text{for } f T_s \geq 1 \quad (20)$$

The shape of the weighting functions for $b = 0$ and $b = 0.5$ are represented in Fig. 6(a) on the hypothesis of $N = 10$. This very low value of N has been chosen only to show the shape of the weighting function more clearly. We can therefore conclude that the bandwidth of this random sampling strategy with $b = 0$ is not limited by the average sampling frequency $f_s = 1/T_s$, but solely by that of the S/H circuit used. It has been shown that these theoretical findings strongly agree with the experimental ones (25).

The weighting function for the sampling strategy defined by Eq. (15) is the following (26):

$$W^2(f T_s) = \frac{1}{N} + \frac{2}{N^2} \sum_{r=1}^{N-1} (N-r) \cos \left[2\pi r \left(1 + \frac{b}{2} \right) f T_s \right] \times \text{sinc}^r(bf T_s) \quad (21)$$

The plot of this function for different values of b ($b = 0.5, 1, 2$) and for $N = 10$ is reported in Fig. 6(b). For $f = 0$ we obtain

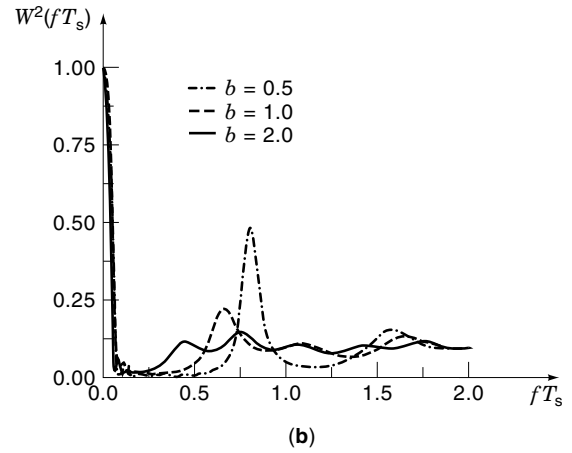
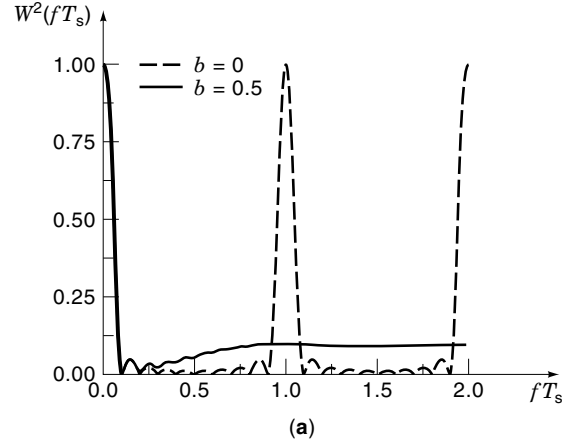


Figure 6. Weighting functions with $N = 10$ of (a) the equally spaced sampling strategy ($b = 0$) and of the random strategy defined by Eq. (14) with $b = 0.5$; (b) the random recursive strategy defined by Eq. (15). Comparison of the three weighting functions shows that neither of the considered random strategies introduces any frequency limitation.

$W^2(0) = 1$, while for $f \rightarrow \infty$ it results in $W^2(f T_s) \rightarrow 1/N$. The shape of this weighting function differs from that of the nonrecursive one only for the presence of a ripple around $1/N$ whose amplitude decreases as b , the range of the random increments X_{k-i} in Eq. (15), increases. Clearly, the overshoot above $1/N$ must not exceed an acceptable threshold in order to contain the contribution of each spectral component to the asymptotic variance through the corresponding weighting coefficients. On the other hand, higher values of b increase the mean response time of the wattmeter, and consequently b must be selected as small as possible. An optimum value is $b = 1.5$, but in any case the value of b is not critical. The theoretical study has been confirmed by the experimental results with a prototype implementing this sampling strategy (27).

It is interesting to note that the standard error associated with the sampling strategy and filtering algorithm of both these asynchronous random strategies results in $1/\sqrt{N}$ while that of the quasisynchronous equally spaced one is $1/N$; therefore a large bandwidth in an asynchronous sampling can be obtained only with a decrease in accuracy. Since the bandwidth of the proposed asynchronous random strategies is not limited by the sampling frequency but only by the bandwidth of the S/H, a

high resolution ADC can be adopted in order to implement a wattmeter characterized both by a large bandwidth and high resolution. A prototype based on the strategy defined by Eq. (4) was used to evaluate the power losses in electronic devices used in switching converters. Modern power devices are characterized by low losses and high switching speed. Therefore, the instrument required to measure their power losses must have a bandwidth from dc to several megahertz and a resolution so high that its performance is still acceptable when the power factor is of a percent order. The accuracy of the experimental results was of some percent in comparison with that of the calorimetric method (28).

TIME-JITTERED EQUISPACED SAMPLING WATTMETERS

The random fluctuations with respect to the nominal sampling instants are commonly called time jitter; since they are present both in the voltage and current channels, the effect of time jitter on the performance of the wattmeter must be considered (29). For instance, the unpredictable temporal variations of the aperture delay of each channel S/H produce different random effects in the sampling instants: The noise associated with the clock generator can instead cause common random independent fluctuations in the sampling instants. Therefore, the effect of time jitter must be analyzed by considering the sampling instants of the two channels separately, and each of them can be determined by adding to the right side of Eq. (4) two quantities: $X_{k-i}T_s$, common to the two channels, and $X'_{k-i}T_s$ for the voltage channel or $X''_{k-i}T_s$ for the current channel respectively:

$$t'_{k-i} = \tau_0 + (k-i + X_{k-i} + X'_{k-i})T_s \quad (22)$$

$$t''_{k-i} = \tau_0 + (k-i + X_{k-i} + X''_{k-i})T_s \quad (23)$$

Obviously, t'_{k-i} and t''_{k-i} are mutually interdependent due to the common component X_{k-i} . Each of the three random variables X_{k-i} , X'_{k-i} , and X''_{k-i} is the i th element of three independent sets of random variables having a continuous distribution with mean value equal to zero. The two variables X'_i and X''_i can be assumed to be characterized by the same distribution.

In the special case of only common time jitter having a uniform distribution in a certain interval $\pm b$, the effect of time jitter can be evaluated by considering the random sampling strategy defined by Eq. (14). Therefore Eq. (19) also represents the weighting function for this case when b has a small value. The shape of the weighting functions $W^2(fT_s)$ of an equally spaced sampling strategy with a jitter uniformly distributed within $\pm b = \pm 0.1$, or $\pm b = \pm 0.05$, and $N = 10$ are shown in Fig. 7 (18). This figure shows that time jitter leads to values of the relative minima increasing with frequency and values of the relative maxima decreasing with frequency. Therefore, in the presence of time jitter, the contribution of the spectral components that meet the condition $q \neq zN$ must also be investigated in order to limit the asymptotic mean square error. By imposing a maximum value on the weighting function, the upper limit of the wattmeter frequency caused by a given time jitter can be deduced from Eq. (19) or Fig. 7.

Separate time jitters for the current and voltage channels generate an asymptotic bias which, under the assumption of periodic input signals and a uniform distribution of the random

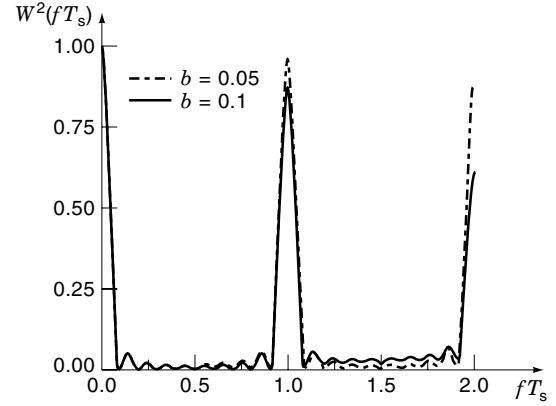


Figure 7. Weighting function with $N=10$ of an equally spaced sampling technique in the case of a common time jitter affecting the two input channels. The function shows that this kind of time jitter increases the relative minima and decreases the relative maxima.

variables in the interval $\pm b_1$, can be expressed as follows (29):

$$\text{Bias} \rightarrow \sum_{\substack{m=-\infty \\ m \neq 0}}^{+\infty} V_m I_m^* (1 - \text{sinc}^2(2b_1 m f_1 T_s)) \quad (24)$$

where V_m and I_m with $V_{-m} = V_m^*$ and $I_{-m} = I_m^*$ are the spectral components in the exponential form of the input signals and $m f_1$ the corresponding frequencies. By Parseval's theorem, the average power P_0 can be expressed as a sum of the average power associated with each individual Fourier component of the input signals:

$$P_0 = \sum_{\substack{m=-\infty \\ m \neq 0}}^{+\infty} V_m I_m^* \quad (25)$$

Therefore the asymptotic bias can be obtained by weighting the contribution of each spectral component with the average power and summing over the whole spectrum. Figure 8(a) shows the plot of the weight $(1 - \text{sinc}^2(2b_1 f T_s))$ in Eq. (24) as a function of fT_s for $\pm b_1 = \pm 0.05$ or $\pm b_1 = \pm 0.1$. Therefore, the bias introduces another bandwidth limitation when it becomes comparable with the instrument's accuracy.

The expression of the asymptotic mean square error in the presence of both common and separate time jitter is rather complex. A simpler result can be deduced when the effect of the common time jitter is negligible and under the assumption of two sinusoidal input signals at frequency $m f_1$. In this case the asymptotic variance can be expressed as follows (28):

$$\begin{aligned} \sigma^2 \rightarrow & 2|V_m I_m^*|^2 \left\{ \frac{2}{N} (1 - \text{sinc}^4(2b_1 m f_1 T_s)) \right. \\ & \left. + \text{sinc}^4(2b_1 m f_1 T_s) |H(2m f_1 T_s)|^2 \right\} \\ & + \frac{P_0^2}{2} \left\{ \frac{1}{N} [\text{sinc}^2(4b_1 m f_1 T_s) - \text{sinc}^4(2b_1 m f_1 T_s)] \right\} \end{aligned} \quad (26)$$

Figure 8(b) shows the shape of the normalized quantity $\sigma_N^2 = \sigma^2 / 2|V_m I_m^*|^2$ for in-phase input signals as a function of fT_s , under the assumption of sampling instants uniformly distributed

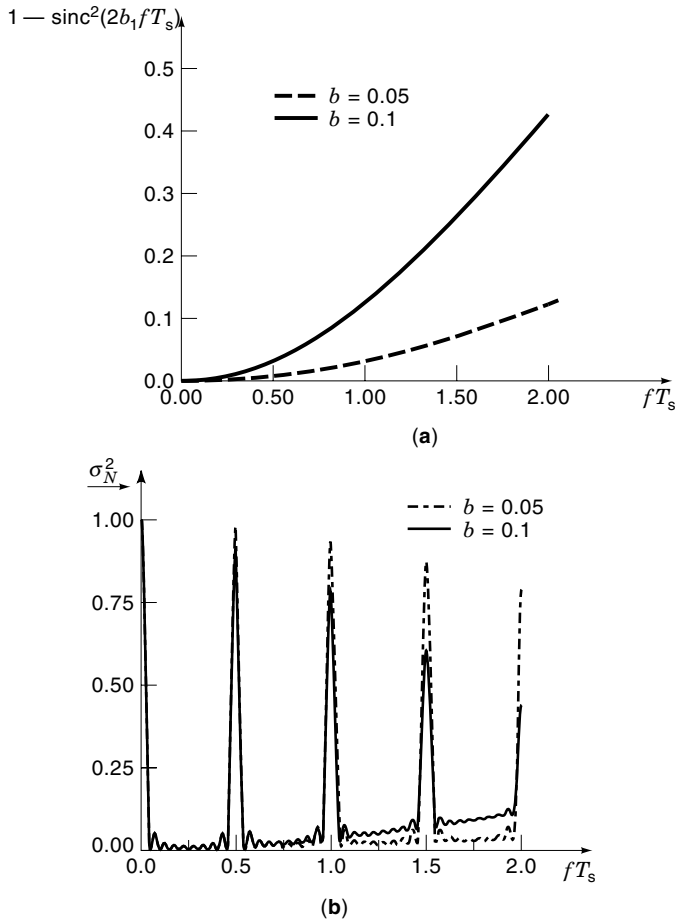


Figure 8. The estimate of the average power in the presence of different time jitters occurring in the voltage and current channels is affected by bias and variance. Plot (a) enables the bias to be deduced according to Eq. (24), while the normalized variance in the case of in-phase input signals can be deduced from plot (b).

over the interval $\pm b_1 T_s$ with $\pm b_1 = \pm 0.05$ or $\pm b_1 = \pm 0.1$ and $N = 10$. The shape of this function is very similar to that of the weighting function in the presence of only common time jitter (Fig. 7); therefore a further bandwidth limitation is introduced.

In conclusion, the effect of time jitter in an equally spaced time sampling must be evaluated, since it introduces an asymptotic variance, and if different time jitters exist in the two channels, it also introduces a bias. The instrument bandwidth must therefore be adequately limited in order to make these effects negligible. When one of the two previously described random sampling strategies is considered instead [Eqs. (14) and (15)], only the different time-jitters introduce a further bandwidth limitation.

INPUT-CONDITIONING CIRCUITS

The conditioning circuits of Fig. 2 must convert the voltage and current, which the power to be measured depends on, into signals acceptable by the multiplier. When the line voltage is hazardous to the operator, these circuits must also ensure insulation between the instrument and the power system.

The input signals of the multiplier must usually be of the same type, either current or voltage, according to the multiplier operating principle. As a consequence, voltage-to-current ($V-I$) and current-to-voltage ($I-V$) transducers are necessary. The multiplier input signals must also have proper amplitude. This calls for current-to-current ($I-I$) and voltage-to-voltage ($V-V$) transducers.

All the above transducers could easily be implemented by means of impedances. In this case, the $V-I$ and $I-V$ transducer transfer functions depend on an admittance or an impedance, respectively; while the transfer function of $I-I$ and $V-V$ transducers depend on the ratio between impedances. It is worth recalling that it is easier to know and keep constant the value of an impedance ratio than the value of an impedance. $I-I$ and $V-V$ transducers can be also implemented by means of instrument transformers, which is the most common solution when electrical safety is the main requirement. Broadband thermal rms/dc converters based on thermocouples are used in wattmeters that implement the quarter-square technique. However, the instrument frequency bandwidth is limited at several orders of magnitude below that of these converters because of the analog circuits that provide the currents $(i_1 + i_2)$ and $(i_1 - i_2)$ shown in Eq. (2).

Power measurement accuracy depends on the characteristics featured by all the signal conditioning chain, hence those of the transducers external to the instrument also must be considered.

The choice of the transducer frequency bandwidth depends on the wattmeter application. For instance, in many industrial processes, where large amounts of power at high currents and voltages are utilized, the active power is substantially transmitted at industrial frequency, even under nonsinusoidal conditions. Indeed, the difference $(P - P_1)$ between the active power P and that at industrial frequency is usually lower than a few parts per thousand of P_1 (30). In particular, the active power associated with the direct voltage and current components can be neglected. In these cases, the use of passive instrument transformers can therefore be quite satisfactory. In other industry applications power measurement may require transducers featuring wider frequency bandwidths (31).

Resistors

Resistors are a simple and cheap solution for the implementation of the above variety of transducers. Their frequency bandwidth includes a direct component and may range up to about 100 MHz. This means that the output of a resistor-based transducer may feature a phase shift of about 0.1 mrad at 10 kHz. The frequency bandwidth is limited by the nominal power of the resistor. Indeed, if the thermal drift is stressed to remain constant, the component sizes must increase as the dissipated power increases, and this increases the component time constant. The power dissipation also sets a limit to the use of resistors in self-contained instruments. For this reason, external series resistors have been used in electrodynamic wattmeters to extend the voltage range.

The stray capacitances may significantly affect the time constants of the resistive voltage divider components. When the two sections of a divider are characterized by different time constants, its transfer function differs from the dc ratio. This is described by means of the ratio and phase-angle errors. The ratio error increases with the product of the

Table 1. Characteristics of Two I - V Wide-Bandwidth Transducers.

Transducer Type	Full-Scale Input (A)	Full-Scale Output (V)	Cut-off Frequencies	Accuracy (%)
Noninductive shunt resistor	100	0.1	1.5 MHz	± 0.1
CT with resistive burden	65	6.5	1 Hz to 20 MHz	$-0/+1$

stray reactances; their difference determines the phase-angle error. The time constants can be compensated by means of capacitors.

Resistor-based transducers do not ensure electrical insulation between the power system and the instrument.

Current and Voltage Transformers

The upper cut-off frequency of instrument transformers is limited by the stray capacitances; hence, a current transformer (CT), even for high-voltage applications, features a wider frequency bandwidth than a voltage transformer (VT). Usually CTs show satisfactory frequency response in the audio frequency range, while ferrite core CTs may feature upper cut-off frequencies of tens of megahertz. Broadband I - V transducers can therefore be implemented by means of CTs with resistive burden. Their output, contrary to that of shunt resistors, does not require amplification. The upper cut-off frequency of the VTs depends on their nominal voltage and is usually lower than 1 kHz.

Table 1 compares some characteristics of two commercial I - V transducers of similar price. Active instrument transformers based on the Hall effect are used when the input signal bandwidth includes the direct component.

Capacitive Dividers

Capacitive voltage dividers are often used to extend the amplitude range of the multiplier for alternating voltages. Capacitors are the passive components that best approach the ideal behavior over a wide frequency bandwidth limited by stray resistances and inductances. The phase shifts exhibited by capacitive dividers depend on the difference between the capacitor loss angles and may be less than 0.1 mrad up to hundreds of kilohertz. At frequencies sufficiently lower than the resonance frequency f_0 depending on the stray inductance, the capacitance of a capacitor is a linear function of the squared ratio of the frequency to the resonant frequency f_0 . Therefore, if the capacitors possess different resonance frequencies, the voltage division accuracy depends on frequency. Moreover, for direct voltages and for alternating voltages at very low frequency, capacitive dividers cannot be used because of stray phenomena. The dc voltage division is independent of the capacitance values of the two sections and is determined only by their equivalent parallel resistances. Moreover, at low frequency the voltage division can also be affected by the occurrence of interfacial polarization.

Like all impedance dividers, capacitive dividers do not ensure the isolation of the instrument from the power system; however, high-voltage transducers based on them are usually built, given that voltage transformers are too expensive for voltage values greater than about 150 kV. Since the divider output voltage commonly ranges between 5 and 20 kV, they are

coupled with a VT that therefore limits the frequency response of the whole conditioning chain.

WATTMETER CALIBRATION

Calibration is the process of deducing or checking the instrument's accuracy. It can be performed by using either a dedicated calibration equipment (direct comparison) or a similar test instrument of known accuracy used as a reference (indirect comparison). The calibration procedure must meet the traceability condition. This means that the individual measurement results must be referred to national standards through an unbroken chain of calibration. Wattmeter calibration is performed at $\cos(\phi) = 1$ for different amplitude and frequency values, in addition, power factor effects must be taken into account. For the latter effect, at least four operating conditions are tested: unity, 0.5 lead and lag, and zero power factors. For the comparison of the results of wattmeter calibration between different Bureaus of Standards, more complex approaches are followed. They provide an estimate of three error coefficients, related to the magnitude and phase errors of the voltage and current circuits, and also of the zero offset (32).

The first method for calibration (direct comparison) requires synchronized sources of ac voltage and current. To this end, voltage and current should be known in amplitude and phase. Moreover, an adequate range of voltages, currents, and phase angle values at different frequencies must be available (33,34). A single instrument may replace the complex combinations of individual voltage and current sources. If the current and voltage circuits of the wattmeter are supplied by separate sources (phantom loading), the source must provide only the power dissipated into the instrument under test. In order to generate accurate reference signals and separate voltage and current channels, synthesized generators are normally used. A frequency range from 20 Hz up to 10 kHz and a best long-term accuracy of less than 100 ppm (referred to the apparent power) are normally provided in commercial calibrators.

The second method (indirect comparison) compares the readings of the unit under test with those of a reference unit. Also in this case it is preferable that the current and voltage circuits of the wattmeter are supplied from separate sources (phantom loading). If only one power source is available, the calibration can be performed by connecting the voltage channels of the two wattmeters in parallel, and connecting the current channels in series with a short-term stable load.

The issues related to electric-power quality have increased the interest in harmonic-power analyzers. Their calibrators offer accuracy and frequency ranges lower than those of the above-mentioned wattmeter calibrators. For instance, a calibration system characterized by an uncertainty of 500 ppm over a frequency bandwidth of about 2 kHz was described in Ref. 35.

BIBLIOGRAPHY

1. W. Mc. Siebert, *Circuits, Signals, and Systems*, New York: McGraw-Hill, 1986.
2. Y. J. Wong and W. E. Ott, *Function Circuits*, New York: McGraw-Hill, 1976.
3. F. K. Harris, *Electrical Measurements*, New York: Wiley, 1952.
4. G. Schuster, Thermal measurement of AC power in comparison with the electrodynamic method, *IEEE Trans. Instrum. Meas.*, **IM-25**: 529–533, 1976.
5. L. G. Cox and N. L. Kusters, A differential thermal wattmeter for the AC/DC transfer of power, *IEEE Trans. Instrum. Meas.*, **IM-25**: 553–557, 1976.
6. R. Bergeest and P. Seyfried, Evaluation of the response of time-division multipliers to AC and DC input signals, *IEEE Trans. Instrum. Meas.*, **IM-24**: 296–299, 1975.
7. M. M. Staborowski, Modern numerical analysis of time-division multipliers, *IEEE Trans. Instrum. Meas.*, **IM-28**: 74–78, 1979.
8. G. J. Johnson, Analysis of the modified tomota-sugiyama-yamaguchi multiplier, *IEEE Trans. Instrum. Meas.*, **IM-33**: 11–16, 1984.
9. P. N. Miljanic, B. Stojanovic, and R. Bergeest, Systematic error of time-division wattmeters when voltage and/or current are distorted. *IEEE Trans. Instrum. Meas.*, **IM-36**: 357–361, 1987.
10. P. S. Filipski, A TDM wattmeter with 0.5-MHz carrier frequency, *IEEE Trans. Instrum. Meas.*, **IM-39**: 15–18, 1990.
11. R. S. Turgel, Digital wattmeter using a sampling method, *IEEE Trans. Instrum. Meas.*, **IM-23**: 337–341, 1974.
12. J. J. Hill and W. E. Alderson, Design of a microprocessor-based digital wattmeter. *IEEE Trans. Ind. Electron. Control Instrum.*, **IECI-28**: 180–184, 1981.
13. C. F. Coombs, *Electronic Instrument Handbook*, 2nd ed., New York: McGraw-Hill, 1995.
14. C. H. Dix, Calculated performance of a digital sampling wattmeter using a systematic sampling. *IEE Proc.* **129** (3): 172–175, 1982.
15. F. J. J. Clarke and J. R. Stockton, Principle and theory of wattmeters operating on the basis of regularly spaced sample pairs, *J. Phys. E: Sci. Instrum.*, 645–652, 1982.
16. F. Filicori, D. Mirri, and M. Rinaldi, Error estimation in sampling digital wattmeters, *IEE Proc.* **132** (3): 122–128, 1985.
17. G. N. Stenbakken, A wideband sampling wattmeter, *IEEE Trans. Power Appar. Syst.*, **PAS-103**: 2919–2926, 1984.
18. D. Mirri et al., Performance function for time-jittered equispaced sampling wattmeters, *IEEE Trans. Instrum. Meas.*, **44**: 671–674, 1995.
19. G. N. Stenbakken and A. Dolev, High-accuracy sampling wattmeter, *IEEE Trans. Instrum. Meas.*, **41**: 974–978, 1992.
20. F. Filicori et al., A random asynchronous sampling strategy for measurement instruments based on non-linear signal conversion, *IEE Proc.*, **136** (3): pt. 141–150, 1989.
21. D. Mirri et al., Recursive random sampling strategy for a digital wattmeter. *IEEE Trans. Instrum. Meas.*, **41**: 979–984, 1992.
22. G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley, 1973.
23. F. Filicori et al., New performance function for the comparison of different sampling strategies in non-linear conversion instruments, *Proc. IEEE Trans. Instrum. Meas. Tech. Conf.*, Washington, April 1989, pp. 307–311.
24. G. Iuculano et al., A criterion for the analysis of synchronous and asynchronous sampling instruments based on non linear processing. *IEE Proc.*, **139** (4): part A, 141–152, 1992.
25. D. Mirri et al., Efficient implementation of digital measurement instruments by random asynchronous sampling, IMEKO TC-4 Symp, “Measurement in Electrical and Electronic Power Systems,” Zurich, September 1989, pp. 231–240.
26. D. Mirri et al., Recursive random sampling strategy for a digital wattmeter, *IEEE Trans. Instrum. Meas.*, **41**: 979–984, 1992.
27. A. Menchetti et al., A wideband random-sampling wattmeter, IMEKO TC-4 Symp, “Intelligent instrumentation for remote and on-site measurements,” Brussels, May 1993, pp. 355–361.
28. A. Menchetti et al., Power loss measurements in electronic switching devices using a new type of digital wattmeter, *Proc. IMEKO XII-World Congr. Meas. Progr.*, Pechino (Cina), September 1991, pp. 1118–1123.
29. D. Mirri et al., Experimental results in a time-jittered equally spaced sampling wattmeter. IMEKO TC-4 Symp., *New Meas. Calibration Meth. Electrical Quantities Instr.*, Budapest, September 1996, pp. 259–262.
30. IEEE Working Group on nonsinusoidal situations, Practical definitions for powers in systems with nonsinusoidal waveforms and unbalanced loads: a discussion, *IEEE Trans. Power Deliv.*, **11**: 79–101, 1996.
31. K. C. Kenneth and D. T. Hess, A 1000 A/20 kV/25 kHz–500 kHz volt-ampere-wattmeter for loads with power factors from 0.001 to 1.00, *IEEE Trans. Instrum. Meas.*, **45**: 142–145, 1996.
32. W. J. M. Moore et al., An international comparison of power meter calibrations conducted in 1987, *IEEE Trans. Instrum. Meas.*, **38**: 395–401, 1989.
33. A. Gubisch et al., Power calibration using sampled feedback for current and voltage, *IEEE Trans. Instrum. Meas.*, **46**: 403–407, 1997.
34. N. Miljanic, Calibrator for alternating voltage, current and power. *IEEE Trans. Instrum. Meas.*, **38**: 384–389, 1989.
35. R. Arseneau and P. Filipski, A calibration system for evaluating the performance of harmonic power analyzers, *IEEE Trans. Power Deliv.*, **10**: 1177–1182, 1995.

DOMENICO MIRRI
GAETANO PASINI
University of Bologna
GAETANO IUCULANO
University of Florence
FABIO FILICORI
RENATO SASDELLI
University of Bologna

VOLT-AMPERE METERS

Volt-ampere meters are multipurpose electric measuring instruments that display the values of active, reactive, and apparent electric power. Some of them also measure other electric quantities, such as current, voltage, and energy.

Before presenting some of the basic ideas that support the operating and implementing principles of volt-ampere meters, it is worth discussing the concepts of active power, reactive power, and apparent power.

Power is the rate at which energy is transferred or transformed into a different form. Active power P , also called true power, is physically related to the energy nonreversibly converted to heat (Joule effect) in the resistive components of electric circuits. Thus, in a resistor R with current i , P is given by

$$P = RI_{\text{rms}}^2 = \frac{U_{\text{rms}}^2}{R} = (ui)_{\text{av}} = \frac{1}{T} \int_T (ui) dt = (p)_{\text{av}} \quad (1)$$

where rms stands for root-mean-square, u is the voltage drop in R , p is the instantaneous power, T is the period of p , u and i represent instantaneous values, and U_{rms} and I_{rms} represent constant values.

Equation (1) is the basic definition of active power in a circuit. Active power is the time average of the product of instantaneous voltage and current (instantaneous power). When, and only when, voltage and current are sinusoidal, Eq. (1) yields the known relationship

$$P = U_{\text{rms}} I_{\text{rms}} \cos \phi \quad (2)$$

where ϕ is the phase angle between u and i . The unit of active power is the watt (W).

Contrary to the active power which has a physical meaning, reactive power (R) and apparent power (S) are electrical quantities that have the dimensions of power but have significance only when voltage and current are sinusoidal or near sinusoidal. These two quantities are related to the behavior of the reactive components of circuits and their meanings result from interpreting the complex Poynting theorem (1). For a circuit under a sinusoidal regime,

$$R = U_{\text{rms}} I_{\text{rms}} \sin \phi \quad (3)$$

$$S = U_{\text{rms}} I_{\text{rms}} = \sqrt{P^2 + R^2} \quad (4)$$

Reactive power is expressed in volt-ampere reactive (VAR) and apparent power in volt-ampere (VA).

TYPES OF VOLT-AMPERE METERS

Nowadays, because of the widespread use of personal computers (PCs) and PC data acquisition boards (DAQs), the design of multipurpose measuring instruments that function in particular as volt-ampere meters is a problem of software as well as hardware. And because the implementation of the hardware to digitize several electric voltages and currents is trivial in many situations, the measurement capacity of a PC with a DAQ depends mainly on the software component of such a measuring instrument, usually designated by a virtual instrument (VI). The use of object-oriented languages and the existence of commercial software packages supported and supporting visual programming has made life easier for VI designers and also for users, because those programming tools facilitate the implementation of instruments with user-friendly interfaces. It is our belief that in the near future PC-based virtual instrumentation will render obsolete many dedicated analog and digital instruments still in use now, particularly in the laboratory. Volt-ampere meters will be no exception. Model TPZ 303 from ZERA Electric GmbH is an example of this type of instrumentation already available. It is a multifunction, high-accuracy (200 ppm for active power) measuring system based on the digitization of the input quantities and signal processing. Designed to be a portable stand-alone instrument, the processing unit is not an external PC but an internal digital signal processor (DSP). However, an overview of some of the other types of volt-ampere meters is justified, because instruments based on different operating principles are still manufactured and used, particularly in industry.

The presentation of volt-ampere meters that follows is organized according to the power measurement principles they use.

According to the measuring method, power meters may be divided into analog and digital types. In analog power meters, the measurement is obtained by an electric quantity that is proportional to the power. This quantity is produced by analog means and is displayed in either analog or digital format. In digital power meters, voltage and current are digitized and the measurement of power is obtained upon adequate reduction of the resulting data. Presentation of the results is usually digital. A VI power meter is an example of a digital power meter.

From Eqs. (1), (3) and (4), one may conclude that any idea that yields the analog product of two voltages or currents supports a power measuring principle and the design of an analog power meter (analog multiplication power meter). Those expressions also clearly identify the operations that the software of digital power meters must implement to produce the measurements.

In the case of active power, some analog meter designs are based on the physical meaning of active power. Examples of methods used are (1) measurement of the voltage or current in a known noninductive resistor by a thermogalvanometer or a thermocouple ammeter; (2) measurement by comparing the brightness of an incandescent lamp driven by the source of power to be measured by using a voltmeter, an ammeter, an incandescent lamp, and a photosensitive meter; (3) measurement of the temperature change when a temperature-sensitive resistor (bolometer) absorbs energy. The corresponding change of resistance is measured by a bridge circuit; (4) measurement of part of the heat produced in a resistor using the bolometer method (calorimetric active power meters). These methods may be used to implement low-frequency active power meters. Nevertheless, the meters commercially available are almost exclusively designed for the radio-frequency domain. For that reason, those types of meters are not considered in this article. Interested readers are directed to Refs. 2 and 3.

Reactive and apparent powers are important in power systems and so the solutions considered in the following paragraphs are used in low-frequency power meters.

ANALOG MULTIPLICATION POWER METER

Electrodynamic Principle (Electrodynamic Volt-Ampere Meters)

Electrodynamic instruments are analog instruments that have two coils, one fixed (current coil or current circuit) and the other able to rotate around an axis inside the first coil (voltage coil). For frequencies higher than 1 Hz (4) the change of the angular position α between the two coils is proportional to the time average of the product of the currents in the two coils:

$$\alpha = k(i_1 i_2)_{\text{av}} = k \frac{1}{T} \int_T i_1(t) i_2(t) dt \quad (5)$$

To measure the active power in a load Z_L , the two coils are connected as shown in Fig. 1. The instrument indicates the active power in the load added to the active power in either the current circuit (connection 1–2) or in the voltage circuit

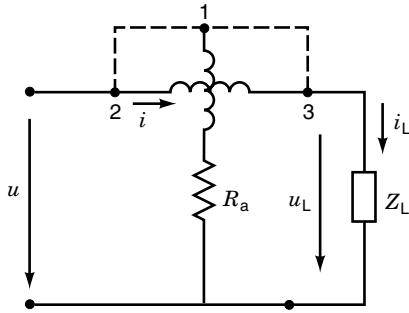


Figure 1. Connections of an electrodynamic wattmeter to a load Z_L . R_a is an additional resistance used in the voltage circuit.

(voltage coil in series with resistor R_a , connection 1–3). Thus, to obtain the correct value of the active power in the load, the indication has to be corrected. For details, see Refs. 4 and 5. Equation (5) reveals that as long as i_1 and i_2 represent the load current and load voltage, the electrodynamic wattmeter measures the active power. The shape of the waveform does not limit the validity of the measurement.

To measure reactive power, the voltage circuit of the wattmeter (fixed coil and resistance in series) must be slightly changed. In fact, taking into account that the measuring situation is relevant only when sinusoidal waveforms are present, for which Eqs. (3) and (4) are valid, and also that

$$\sin \varphi = \cos \left(\varphi - \frac{\pi}{2} \right) \quad (6)$$

the reactive power may be measured if the current in the moving coil is not proportional to the load voltage, but to a voltage lagging the load voltage by $\pi/2$. Figure 2 shows the voltage circuit of an electrodynamic reactive power meter (electrodynamic varmeter) and the vector diagram of the pertinent electrical quantities. The vector diagram presented is valid only for one frequency ω , which means that the electrodynamic varmeter is a single-frequency varmeter. The measurements are correct only when the load current and voltage are sinusoidal at the frequency considered in the design of the instrument. Commercial electrodynamic varmeters for measurements at the frequency of the main power supply are more common.

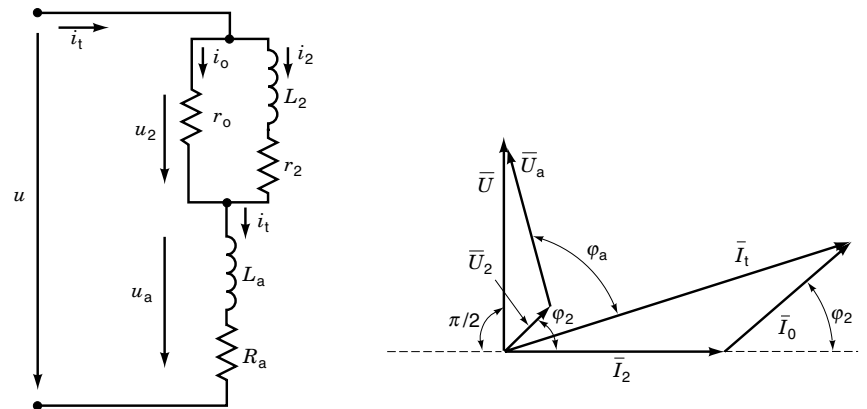


Figure 2. Voltage circuit of an electrodynamic varmeter and the corresponding vector diagram. The resistance r_0 is in parallel with the impedance of the voltage coil ($r_2 L_2$). i_2 lags i_0 and by an appropriate choice of r_0 , R_a and L_a , i_2 lags the load u by $\pi/2$.

The electrodynamic principle also allows constructing an instrument for measuring apparent power directly. For that purpose, the two coils of an electrodynamic instrument must be preceded by circuitry to output dc voltages or currents proportional to the rms values of the voltage and current involved in the circuit under test. This solution, however, is not used in the practice.

The series connection of the two coils of an electrodynamic instrument transforms it into a true rms current meter (electrodynamic ammeter) for frequencies ranging from dc to a few hundred hertz. A resistor in series with the two windings transforms the ammeter into a true rms voltmeter (electrodynamic voltmeter). The upper frequency limit of electrodynamic voltmeters is also a few hundred hertz but slightly lower than for electrodynamic ammeters.

Readers interested in typical values and specifications for electrodynamic instruments, are directed to Refs. 4 and 5.

Thermoelectric Conversion Principle (Thermal Volt-Ampere Meters)

Analog multiplication related to the thermoelectric effect is achieved by using two thermoelectric converters in the configuration depicted in Fig. 3. Currents i_1 and i_2 are proportional to the voltage and current involved in the power to be measured. The two thermoelectric converters are coupled, and their filaments (heaters) have resistances R_f . Because the circuit is symmetrical, the filament currents of the converters i_A and i_B are given by

$$i_A = \frac{1}{2}(i_1 + i_2) \quad (7)$$

$$i_B = \frac{1}{2}(i_1 - i_2) \quad (8)$$

These two currents produce two dc voltages U_A and U_B given by

$$U_A = k_C I_{A_{rms}}^2 = k_C (i_A^2)_{av} \quad (9)$$

$$U_B = k_C I_{B_{rms}}^2 = k_C (i_B^2)_{av} \quad (10)$$

where k_C is the constant of the converters.

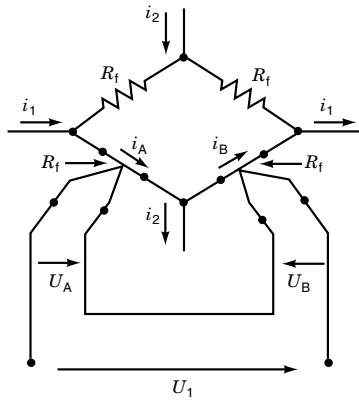


Figure 3. Thermoelectric power meter using two identical thermoelectric converters that have resistance R_t . The currents i_1 and i_2 , whose average product is to be measured, enter the circuit at opposite corners of the bridge.

Replacing Eq. (7) in (9) and Eq. (8) in (10) yields

$$U_A = \frac{k_C}{4}[(i_1 + i_2)^2]_{av} = \frac{k_C}{4}(i_1^2)_{av} + \frac{k_C}{4}(i_2^2)_{av} + \frac{k_C}{2}(i_1 i_2)_{av} \quad (11)$$

$$U_B = \frac{k_C}{4}[(i_1 - i_2)^2]_{av} = \frac{k_C}{4}(i_1^2)_{av} + \frac{k_C}{4}(i_2^2)_{av} - \frac{k_C}{2}(i_1 i_2)_{av} \quad (12)$$

$$U_1 = U_A - U_B = k_C(i_1 i_2)_{av} \quad (13)$$

Equation (13) reveals that the dc voltage U_1 of Fig. 3 is proportional to the active power, provided that i_1 and i_2 themselves are proportional to the voltage and current in the circuit.

To assemble an active power meter, the circuit of Fig. 3 must be complemented with a dc voltmeter to measure U_1 and with a resistor to convert the voltage in the circuit into i_1 or i_2 .

To implement either an apparent power meter or an ammeter or a voltmeter, the considerations before about the corresponding electrodynamic meters are valid. Constructing a reactive power meter based on thermoconverters is also possible. All that is required is to include a circuit before that of Fig. 3 to produce a lag of $\pi/2$ in the current representing the voltage of the circuit. An active integrator can be used as the basic building block of that circuit.

The thermoelectric analog multiplication just presented permits assembling wideband multipower meters because thermoelectric converters can be used at frequencies up to a few hundred megahertz.

Hall Effect Principle (Hall Effect Volt-Ampere Meters)

The Hall effect is a phenomenon that occurs in conductors and semiconductors but is particularly important in the latter. It consists of the appearance of a voltage u_H between two parallel faces of a parallelepiped semiconductor chip when the semiconductor is subjected to a current i and to a magnetic induction field B acting in the two other orthogonal directions. The relationship among the three quantities is given by

$$u_H = \frac{R_H}{d} i B(t) \quad (14)$$

where R_H is the so-called Hall coefficient of the material and d is its thickness. Figure 4 represents the circuit configuration for the analog multiplication of a voltage and a current using the Hall effect principle. The magnetic field in the semiconductor is proportional to the current in the circuit and the current in the semiconductor is proportional to the voltage in the circuit. Thus, the time averaging of u_H , which can be done by mechanical or electrical means, is proportional to the active power. In the first case, a d'Arsonval meter is usually used. In the second, an electronic integrator (active low-pass filter) is the more common solution.

As in the case of thermoelectric conversion, the Hall effect principle can support the design of a multipower meter for measuring active, reactive, and apparent power. However, as far as we know, only active power meters have been assembled and used. The direct field sensor (DFS) from Landis & Gyr is an example of applying the Hall effect to measuring active power and also electrical energy. DFS incorporates a Hall sensor, a custom chip, and additional circuit elements. Its main positive characteristics are: (1) accuracy better than 1%; (2) good stability and the possibility of automatic calibration, which leads to highly reliable maintenance-free meters; (3) very good electromagnetic immunity, in part because of its inherent galvanic isolation; and (4) the possibility of integrating functions, such as radio telemetry.

Feedback Time-Division Multiplier Principle (FTDM Volt-Ampere Meters)

Time-division multiplication is a technique of analog multiplication based on the concept that the area of an electric pulse is equal to the product of the pulse width and pulse height. The circuit represented in Fig. 5(a) is a possible hardware solution

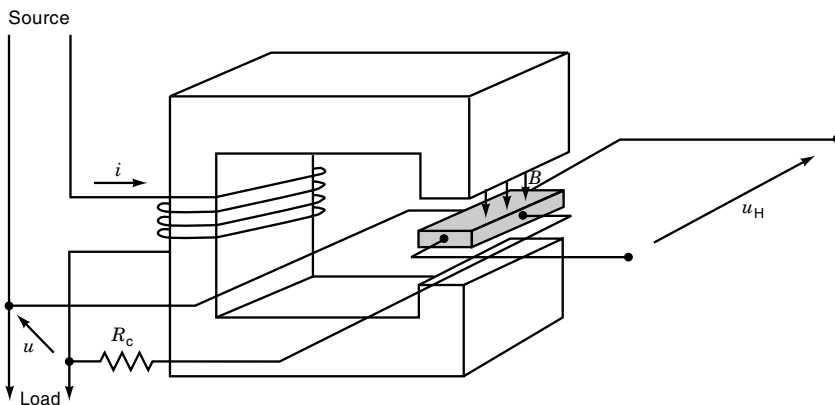


Figure 4. Hall effect analog power meter. The analog multiplication of the current and voltage comprised in the power is obtained by using the Hall effect principle. $B(t)$ is proportional to the current and i to the voltage in the circuit.

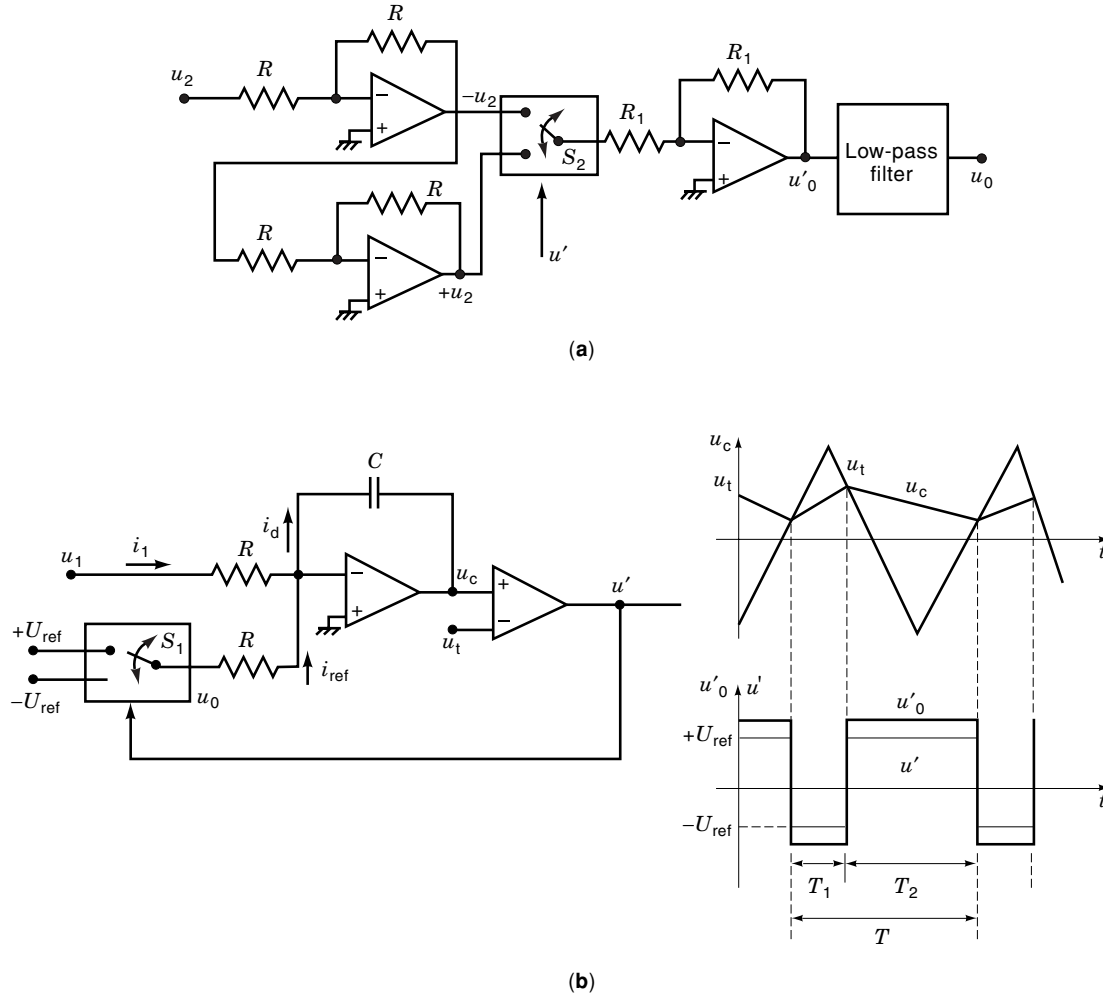


Figure 5. Time division multiplier: (a) amplitude modulator and low-pass filter. u' , the output voltage of the MSA converter, represented in (b), is multiplied by u_2 , and the average of this product is achieved by using a low-pass filter; (b) mark space amplitude converters. The duty cycle of a constant amplitude and period rectangular voltage, the output voltage u' , is modulated by a voltage u_1 .

to implement an analog multiplier according to this concept. Because of its positive characteristics, nowadays the FTDM principle first introduced by Yokogawa Electric Works (YEW) is extremely popular among manufacturers of electronic volt-ampere meters.

The basic component of the multiplier is the mark space amplitude (MSA) converter [Fig. 5(b)]. In this converter, the duty cycle of a constant amplitude and period rectangular voltage is modulated by a voltage u_1 . Thus, and referring to Fig. 5(b),

$$T = T_1 + T_2 = k_1 \quad (15)$$

$$T_1 - T_2 = k_2 u_1 \quad (16)$$

where k_1 and k_2 are two constants. By multiplying the MSA output voltage u' with the other voltage u_2 occurring in the product, one obtains a voltage u'_0 :

$$u'_0 = u_2(T_1 - T_2) = k_2 u_1 u_2 \quad (17)$$

This operation is performed by using an amplitude modulator and a low-pass filter [Fig. 5(a)].

The MSA Converter. The MSA converter represented in Fig. 5(b) was also introduced by YEW and is used, for instance, in its digital wattmeter model 2504.

The operation of the converter is as follows. A comparator compares a voltage u_c with a triangular waveform u_t (this voltage is internally generated in the model previously mentioned). When $u_c > u_t$, u' is positive ($u' > 0$) and S_1 switches so that u_0 becomes equal to $+U_{ref}$. When $u_c < u_t$, the comparator output voltage is negative, and S_1 changes to $-U_{ref}$.

If the frequency of the triangular waveform u_t is much higher than the frequency of u_1 and u_2 , then these voltages can be considered constant compared with u_t . Thus, in the steady state, u_c should be periodic and $(i_d)_{av} = 0$ because u_c is the integral of this current. Thus,

$$T_1 \left(\frac{u_1}{R} - \frac{U_{ref}}{R} \right) + T_2 \left(\frac{u_1}{R} + \frac{U_{ref}}{R} \right) = 0 \quad (18)$$

$$(T_1 + T_2) \frac{u_1}{R} = (T_1 - T_2) \frac{U_{\text{ref}}}{R} \quad (19)$$

$$(T_1 - T_2) = T \frac{u_1}{U_{\text{ref}}} = k_2 u_1 \quad (20)$$

$$T_1 = \frac{T}{2} \left(1 + \frac{u_1}{U_{\text{ref}}} \right) \quad (21)$$

This means that T_1 is continuously adjusted by comparing u_c and u_t . If the average value of i_d is not zero, then the voltage u_c is not periodic but increases or decreases, changing the relationship between T_1 and T_2 . The steady state occurs when $(i_d)_{\text{av}} = 0$. The action of the closed-loop set comparator switch is instrumental for that purpose.

The Amplitude Modulator and Low-Pass Filter. Referring to Fig. 5(a), and taking into consideration that the output voltage u' of the comparator of the MSA converter controls switch S_2 ,

$$u'_0 = \begin{cases} +u_2 & \text{if } u' < 0 \\ -u_2 & \text{if } u' > 0 \end{cases} \quad (22)$$

and thus,

$$\begin{aligned} u_0 &= (u'_0)_{\text{av}} = \frac{1}{T} \int_T u'_0(t) dt = \frac{1}{T} [u_2 T_1 - u_2 (T - T_1)] \\ &= \frac{1}{T} [2u_2 T_1 - u_2 T] = \frac{1}{T} \left[2u_2 \frac{T}{2} \left(1 + \frac{u_1}{U_{\text{ref}}} \right) - u_2 T \right] = \frac{u_1 u_2}{U_{\text{ref}}} \end{aligned} \quad (23)$$

The FTDM principle for analog multiplication can be used for active, reactive, or apparent power measurement. Naturally, the voltage and current of the circuit test must be conditioned before they are applied to the multiplier input. In the case of the current, the conditioner must include a current-to-voltage converter using, for instance, a shunt resistor and an operational amplifier.

For active power measurement, the output voltage u_0 of the circuit of Fig. 5(a), must be averaged by another low-pass filter, usually by electronic means.

For reactive power measurement, the voltage of the circuit test must be lagged by $\pi/2$ before it is input to the FTDM. That lag is easily obtained from electronic circuitry and, thus, it is performed after the voltage input conditioner. The output of the multiplier must also be low-pass filtered to yield the reactive power value.

For apparent power, the easiest solution involves two rms converters placed between the input conditioners and the multiplier. The product of the two dc voltages output by the rms converters indicates the value of the apparent power.

The measurement of voltage and current is also easily implemented by using the FTDM multiplier. For that purpose, it is sufficient to connect the output of the voltage conditioner or of the current conditioner to both inputs of the multiplier.

Several FTDM volt-ampere meter manufacturers, such as YEW, Dranetz, Landis & Gyr, and Schlumberger offer energy counters based on the FTDM multiplying principle. In fact, all that it is required is to add an integrator to the active power meter. In the YEW wattmeter model 2504, the integrator is a stand-alone instrument (YEW 2513).

Logarithmic Conversion Principle

One easy way of performing the analog multiplication of two electric quantities (voltages or currents) is with logarithmic amplifiers (6). In fact, because

$$\log(ab) = \log a + \log b \quad (24)$$

the logarithmic conversion of two quantities transforms its product into a sum, easily achieved by using operational amplifiers. Applying the inverse function to Eq. (24) the product of the two quantities a and b is given by

$$\log^{-1}[\log(ab)] = ab = \log^{-1}(\log a + \log b) \quad (25)$$

Figure 6 presents a circuit based on logarithmic and antilogarithmic converters which multiplies u_1 and u_2 . The first module uses two identical logarithmic converters, whose output voltages are related to the input voltages by

$$\begin{aligned} u_{01} &= -k_1 \log_{10} \frac{u_1}{RI_{\text{ref}}} \\ u_{02} &= -k_1 \log_{10} \frac{u_2}{RI_{\text{ref}}} \end{aligned} \quad (26)$$

where

$$k_1 = \frac{R_T + R_1}{R_T} \cdot \frac{kT}{q} \frac{1}{0.434} \quad (27)$$

In Eq. (27), k is the Boltzmann constant ($1.38 \times 10^{-23} \text{ J}\cdot\text{K}^{-1}$), T is the absolute temperature of the semiconductor junctions, and q is the charge of the electron ($1.6 \times 10^{-19} \text{ C}$).

The operational amplifier of the intermediate stage provides

$$u_{03} = -k_1 \log_{10} \frac{u_1 u_2}{R^2 I_{\text{ref}}^2} \quad (28)$$

The final module performs the antilogarithmic operation. The output voltage is proportional to the antilogarithmic of the input voltage

$$u_0 = RI_{\text{ref}} \log_{10}^{-1} \left[k_1 \times \frac{\log_{10} \left(\frac{u_1 u_2}{R^2 I_{\text{ref}}^2} \right)}{k_1} \right] = \frac{u_1 u_2}{RI_{\text{ref}}} \quad (29)$$

Implementing power meters using this principle is fairly easy but leads to instruments with moderate performance. In fact, the existence of logarithmic and antilogarithmic amplifiers and of log/antilog multipliers as integrated circuits (e.g., Burr-Brown 4127, Burr-Brown 4302) simplifies designing and constructing power meters. However, such components are either inexpensive and moderately accurate or they perform well, but then they are quite expensive. For that reason volt-ampere meters based on this type of component are not popular and are limited to applications demanding only low accuracy (not less than 1%).

As long as the multiplier block operates with positive and negative input voltages (four-quadrant multiplier), its functionality, from the viewpoint of the terminal voltages, is similar to that of the FTDM. This means that the considerations in the

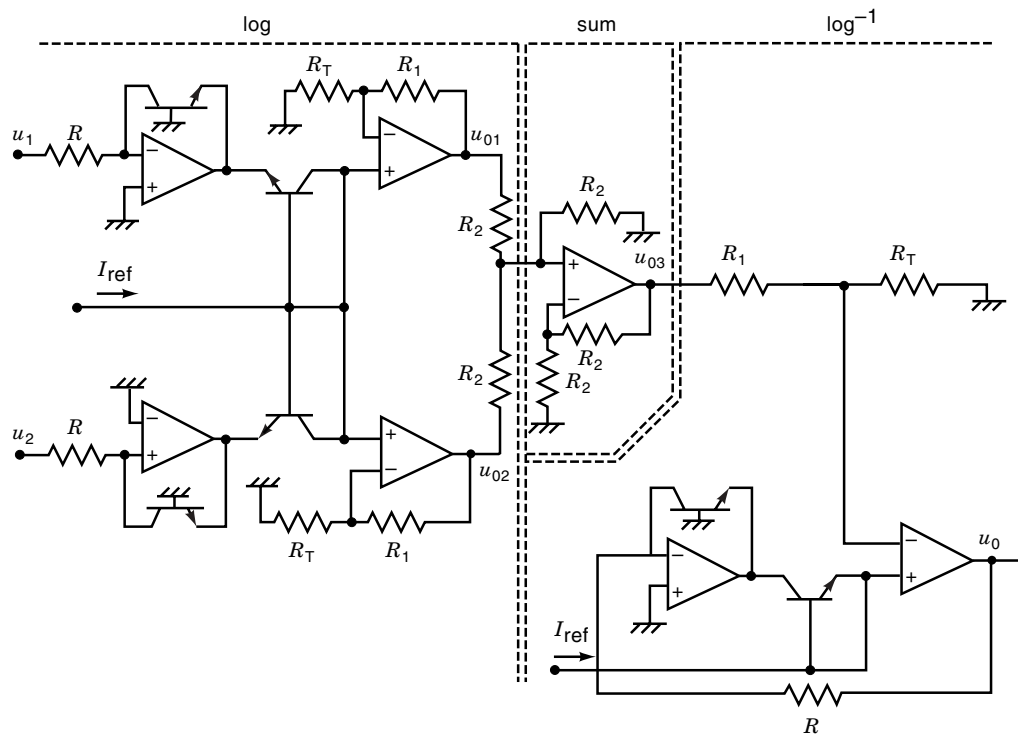


Figure 6. Analog multiplication using logarithmic and anti-logarithmic amplifiers: $u_{01} \propto \log u_1$; $u_{02} \propto \log u_2$; $u_{03} \propto \log u_1 + \log u_2$; $u_0 \propto \log^{-1}(u_{03}) = \log^{-1}(\log u_1 + \log u_2) = u_1 u_2$.

previous section regarding circuitry for measuring the different types of power, voltage, current and even energy are still pertinent.

Other Solutions for Electric Analog Multiplication

Two other types of solutions for analog multiplication of two voltages (or currents) are worth a short discussion.

The first has its origin in the following relationship:

$$(a + b)^2 - (a - b)^2 = 4ab \quad (30)$$

The circuitry used to implement this idea also involves, in general, logarithmic amplifiers, and thus this type of multiplier is considered a variation of the log/antilog multiplier mentioned in the preceding paragraph.

In the second type we include several different multipliers that are based on different principles but have two aspects in common. They are electronic multipliers, and they are available as integrated circuits. The so-called transconductance principle (7,8) is an example of a principle used, for instance, in Burr-Brown MPY 100 and in Harris ICL8013.

DIGITAL POWER METER

The digitization of low-frequency electrical voltages and currents is presently achieved simply and at low cost. It originated from low cost, high-performance, and reliable analog and digital integrated circuits. The basic blocks of a digitizing system are naturally the sample-and-hold and the analog-to-digital converter. The performance of digitizing systems is highly dependent on the performance level of those compo-

nents. Notwithstanding, the advances in digital circuits for control, data processing, data storage, and data transmission led to the replacement, whenever possible, of analog measuring methods by digital measuring methods. This trend is also felt in the domain of low-frequency power and energy measurements. As already mentioned, active power, reactive power, and apparent power are measured digitally in two steps. First, the voltage and current are simultaneously converted to digital words, and then the resultant data is digitally processed by some processing unit. This means that two equally important components are involved in the operation of this type of power meter, the hardware and the software. The routines that constitute the software can be quite different from instrument to instrument. The hardware components of the instruments have, however, some similarities. Figure 7, which represents the implementation used by Yokogawa in models WT110/WT130, gives a good picture of the building blocks of a typical low-frequency multi-power meter.

This instrument consists of various sections: input circuits (voltage and current input circuits), a digital signal processor (DSP), a central processing unit (CPU), a display and interface circuits.

In the voltage input conditioner, the input voltage is attenuated and then sent to an analog-to-digital converter (ADC). In the current input conditioner, the current involved is converted into a voltage by using a shunt resistor. Then the voltage is amplified and formalized by an operational amplifier before entering another ADC converter. The outputs from both ADC converters are sent to the digital signal processor (DSP). The DSP reduces the data to produce the values of voltage, current, and active power for each sampled datum sent from the ADC converter. After processing a certain number of sets of data,

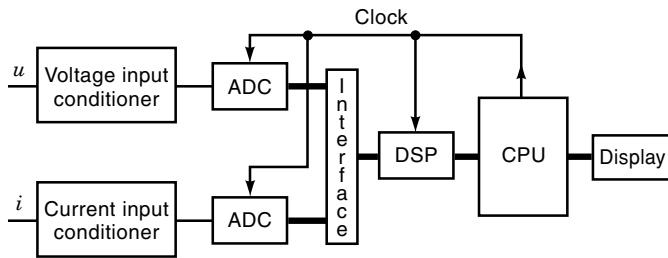


Figure 7. Digital power meter: after digitizing the voltage and current, the CPU reduces the data to compute the active, reactive, or apparent power.

computation of the apparent power, reactive power, power factor, and phase angle starts. Computational results are sent to the CPU where computations, such as range conversion and scaling, are carried out. Display and outputs are also controlled by the CPU. The output goes to a readout driver (display).

GENERAL CONSIDERATIONS

Although the measurement of active power is important in different frequency domains, reactive power and apparent power measurements are required basically in power systems. The measuring principles discussed take that into account.

In power systems, it is often necessary to measure the three-phase power and energy. It is worth mentioning that some measuring methods, such as the Aron method, avoid the use of three volt-ampere meters. Interested readers are directed to Chapter VI of Ref. 4 and Chapter 6 of Ref. 5.

Excluding electrodynamic volt-ampere meters, all of the analog multiplication power meters considered are built around a block that produces a voltage (or current) proportional to the product of the current and voltage of the circuit under test. The meter hardware must be complemented with a voltage (or current) measuring device and with input conditioning circuitry whose composition depends on the quantity to be measured by the power meter and on its accuracy. For measurements in the range 45 to 65 Hz with accuracy better than 0.5%, the input circuits include current and voltage transformers or, less often, voltage dividers and current shunts. These components, that reduce the amplitude of the input current and voltage to levels accepted by the electronic components, are the front end of the meters and much of the meters performance depends on the metrological characteristics of such passive components. The reader should take the solutions considered in the text as guidelines for understanding the internal composition of those types of volt-ampere meters.

The following values are typical for the accuracy of volt-ampere meters: analog type—0.1% to 2% full-scale value (FSV); digital type—0.2% to 2% FSV. Presently, the standards used for calibration are usually of the FTDM type (e.g., YEW 2855) whose accuracy is around 0.02% FSV.

In this article, we elected to separate volt-ampere meters into analog and digital types according to the principle of measurement they use, but this classification is not universal. In fact, most manufacturers use the term digital whenever the meter displays the measurement numerically. The reader must be aware of this fact when consulting different manufacturers' catalogs.

BIBLIOGRAPHY

1. S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and waves I*, Communication Electronics, New York: Wiley, 1984, Chap. 3.
2. H. M. Berlin and F. C. Getz, Jr., *Principles of Electronic Instrumentation and Measurement*, Columbus, OH: Merrill, Chap. 6, 1988.
3. B. M. Oliver and J. M. Cage, *Electronic Measurements and Instrumentation*, New York: McGraw-Hill, 1975, Chap. 16.
4. P. Jacobs and V. Jadin, *Mesures Électriques*, Paris: Dunod, 1968, Chap. II.
5. M. Stockl and K. H. Winterling, *Elektrische Messtechnik*, Stuttgart: Teubner, 1973, Chap. 2.
6. T. T. Lang, *Electronics of Measuring Systems*, New York: Wiley, 1987, Chap. 5.
7. Integrated Circuits Data Book—Linear Products, Burr-Brown Corporation, Tucson, AZ.
8. Linear & Telecom ICs for Analog Signal Processing Applications, Harris Semiconductor, Melbourne, FL.

PEDRO M. B. SILVA GIRÃO

ANTÓNIO M. CRUZ SERRA

HELENA M. GEIRINHAS RAMOS

Instituto Superior Técnico

POWER FACTOR MEASUREMENT

A source of alternating current (ac) delivers electric energy to a receiver at an average rate, over the period of the current, called an active power. The mathematical product of the active power and the time of energy delivery provides the amount of energy delivered. Customers are billed for this energy.

Electric energy receivers operate at a voltage specified by its root mean square (*rms*) value and frequency. The supply source has to provide this voltage. Electrical power systems are built in such a way that this supply voltage has a near to sinusoidal waveform and is almost independent of receivers. While the supply source delivers energy to the receiver, a supply current flows through the source. Similar to the voltage, this current has an rms value.

The mathematical product of the voltage rms value and the supply current rms value is called the apparent power. This power has no physical meaning. The apparent power cannot be lower than the active power, and it is usually higher. When it is, the supply current rms value is greater than that needed for active power delivery.

The ratio of the active power to the apparent power is known as the power factor (PF). Since the active power cannot be higher than the apparent power, the PF cannot be higher than one. A power factor value lower than one indicates that the receiver is loading the supply source with a current of higher rms value than that needed for active power delivery.

The power factor value is a matter of concern for electric power utilities, which provide electric energy to customers. Power system equipment has to have the capability to handle the needed current rms value. At a low power factor, the supply current rms value is higher than that at high power factor. Thus, a low power factor when supplying receivers has several inconveniences for power utilities: (1) To handle currents with higher rms values, distribution system equipment has to be more expensive. (2) When energy is delivered to a customer, a part of it is lost in the distribution system equipment. This

loss increases with the square of the supply current rms value, that is, with the decline of the power factor. (3) The supply current causes a voltage drop at the impedance of the distribution system, which reduced the voltage provided to customers. In a case of inductive loads and a sinusoidal supply voltage, the lower the PF, the higher the voltage drop.

Consequently, a decline of the PF increases the investment and operating cost of electric power utilities. Therefore the power factor affects the financial accounts for energy between utilities and customers, especially the large ones. This is the reason for the measurement of power factor. Also, this is the reason for developing methods of power factor improvement.

The power factor is a feature of energy receivers. The PF of single-phase receivers that, when supplied with sinusoidal voltage, do not cause current distortion, depends on the phase-shift between the supply voltage and the supply current. Such receivers are referred to as linear. This phase-shift is the only cause of the power factor decline for linear receivers. The supply current, which is shifted with respect to the voltage, contains two components. One of them is in phase with the voltage. Another is shifted by a quarter of the current's period. Only the component in phase with the voltage contributes to the active power delivery. The remaining one only increases the rms value of the supply current and the apparent power. The PF of linear receivers is equal to the cosine value of the phase-shift angle between the supply voltage and current.

When supplied with nonsinusoidal voltage, linear receivers may cause supply current distortion. Also, nonlinear receivers cause current distortion. In both such situations the supply current cannot be referred to as shifted with respect to the voltage. The power factor of such receivers does not equal the cosine value of any angle. Harmonics of the supply current contribute to an increase of the current rms value and consequently to the decline of the power factor. Only the ratio of the active and apparent powers provides the value of the power factor, as defined by Eq. (1).

The power factor of linear three-phase receivers supplied with a sinusoidal voltage is specified similarly as for single-phase receivers only if the supply voltage is symmetrical and the receivers are balanced. In such a case, the PF is equal to the cosine value of the phase-shift angle between the line current and the line-to-ground voltage.

When impedances of particular phases of a 3-phase system are mutually different, the system is unbalanced. Line currents of a three-phase system with unbalanced receivers contain components that do not contribute to the active power of the load. These components contribute only to an increase of the supply current rms value. Therefore load imbalance reduces the power factor. In asymmetrical systems, only the ratio of the active and apparent powers provides the value of the power factor. Unfortunately, a few different quantities are considered to be the apparent power in three-phase systems. These apparent powers are mutually equivalent only when a three-phase system is symmetrical. When the system is asymmetrical, these apparent powers differ. Consequently, the PF depends on the chosen definition of the apparent power.

Measurements of the power factor in systems where it is equal to the cosine value of the voltage and current phase-shift angle are a relatively simple task. The first power factor meters for such systems were built in the last years of the nineteenth century. Construction of the power factor meters for systems where the power factor can be expressed only as the ratio of

the active and apparent powers had to wait, however, until sufficient progress in electronics was achieved.

A power factor meter is a direct-reading instrument for measuring power factor. It is provided with a scale graduated in PF or with a digital display. There are two different types of PF meters: (1) electrodynamic meters (iron-vane or crossed-coil), and (2) electronic (analog or digital) meters.

Single-phase, electrodynamic meters of the power factor, known as cross-coil meters, have two moving coils rigidly connected on a shaft located in the center of the third, a fixed coil carrying the load current. The current in one of the moving coils is proportional to the load voltage. The current in the second coil is shifted by 90° with respect to the current in the first moving coil. The currents flowing through the three coils produce magnetic fields, and consequently, torques are exerted on moving coils. One torque increases with the deflection angle of the coils: the remaining torque declines with that angle. They balance each other at the deflection angle equal to the phase-shift angle between the supply voltage and current. To have the scale of the meter graduated in power factor, the values of the cosine function are allocated to particular deflection angles.

Analog meters can measure the power factor when it is equal to the cosine value of the phase-shift angle as well as when it is only the ratio of the active and apparent powers. In the first case, the difference in the zero crossings of the supply voltage and the supply current provide the phase-shift angle as a fraction of the period, 2π . This fraction is next converted electronically to a cosine value. When the power factor is only the ratio of the active and apparent powers, these powers are converted to dc voltage with electronic multipliers, integrators and rms value to dc voltage converters. These two dc voltages proportional to the active and apparent powers are divided electronically. The output voltage of the divider is proportional to the power factor. It can be measured by a voltmeter with a scale graduated in power factor or with a digital display.

Digital meters calculate the PF based on sequences of digital samples of the supply voltage and current. When sufficiently dense sequences of digitized instantaneous values of the voltage and current are available, all powers, thus also the power factor, can be calculated. Sampling circuits and analog-to-digital (ADCs) converters are the main components of such meters. Also, digital hardware capable of performing calculating algorithms is needed. This can be done by a separate microcontroller or a microcomputer. Digital meters can measure PF if it is equal to the cosine value of the phase-shift angle, or if it is only the ratio of the active and reactive powers.

POWER FACTOR AND ITS MEANING

Power factor λ of a receiver of electric energy is the ratio of active power P of the receiver to its apparent power S , that is, the power factor is defined as

$$\lambda = \frac{P}{S} \quad (1)$$

Active power P is the average rate of energy flow from the supply source to the energy receiver, calculated or measured over one period T of the voltage. In single-phase circuits, the apparent power S is the product of rms values of voltage and

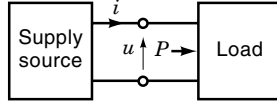


Figure 1. A single-phase load with a supply source and quantities that specify power factor at the supply terminals.

current, U and I , at the terminals of the supply source, as shown in Fig. 1. This is the maximum value of active power P that can be delivered at voltage and current rms values U and I . The PF has similar meaning in 3-phase balanced systems. However, various quantities are considered to be the apparent power in three-phase unbalanced systems, and consequently, there is a controversy regarding the meaning of the power factor.

Various features of energy receivers, such as current phase shift, generation of current harmonics, and current asymmetry of 3-phase equipment cause active power P to be lower than the apparent power S . Therefore, power factor can be interpreted as a measure of the supply source utilization.

POWER FACTOR IN SINGLE-PHASE, TWO-PHASE, AND THREE-PHASE BALANCED CIRCUITS WITH SINUSOIDAL VOLTAGE AND CURRENT

In single-phase circuits with sinusoidal voltage and sinusoidal current, that is, equal to

$$u = \sqrt{2}U \sin \omega t \quad i = \sqrt{2}I \sin(\omega t - \varphi) \quad (2)$$

where U and I denote the rms values of the voltage and current and φ denote their phase-shift angle, the power factor is equal to the cosine value of the voltage and current phase-shift angle φ , since

$$\lambda = \frac{P}{S} = \frac{UI \cos \varphi}{UI} = \cos \varphi \quad (3)$$

The phase-shift between supply voltage and current is the only cause for the power factor's decline in single-phase circuits with sinusoidal voltages and currents. The same is true with balanced 2-phase, 3-wire systems, where

$$\lambda = \frac{P}{S} = \frac{\sqrt{2}UI \cos \varphi}{\sqrt{2}UI} = \cos \varphi \quad (4)$$

where U denotes the rms value of the line-to-line voltage, I is the rms value of the line current, and φ is the phase-shift angle between line-to-ground voltage and the line current. Similarly, in balanced 3-phase systems

$$\lambda = \frac{P}{S} = \frac{\sqrt{3}UI \cos \varphi}{\sqrt{3}UI} = \cos \varphi \quad (5)$$

The power factor in single-phase circuits with sinusoidal voltages and currents can be expressed not only in terms of powers but also in terms of rms values of the active and reactive components of the supply current. These rms values, I_a and I_r , are equal to the real and imaginary parts of the complex rms (crms) value of the supply current $\mathbf{I}$ at the voltage crms value $\mathbf{U} = Ue^{j0}$. At such a voltage, the supply current crms value is

equal to

$$\mathbf{I} = Ie^{-j\varphi} = I \cos \varphi - jI \sin \varphi = I_a - jI_r \quad \text{with} \quad I = \sqrt{I_a^2 + I_r^2} \quad (6)$$

The PF can be expressed as

$$\lambda = \frac{P}{S} = \frac{UI \cos \varphi}{UI} = \frac{I_a}{\sqrt{I_a^2 + I_r^2}} = \frac{1}{\sqrt{1 + (I_r/I_a)^2}} \quad (7)$$

This result means that the presence of the reactive component in the supply current is the cause of the PF decline.

The power factor in single-phase circuits with sinusoidal voltages and currents can also be expressed in terms of the load parameters. If $\mathbf{Y} = Y e^{-j\varphi} = G + jB$ is the load admittance, then the rms values of the active and reactive components of the supply current are equal to $I_a = GU$ and $I_r = |B|U$, respectively, and consequently, power factor can be expressed as

$$\lambda = \frac{1}{\sqrt{1 + (B/G)^2}} \quad (8)$$

If $\mathbf{Z} = Ze^{j\varphi} = R + jX$ is the load impedance, then the load conductance G and susceptance B are equal to

$$G = \frac{R}{R^2 + X^2} \quad B = -\frac{X}{R^2 + X^2} \quad (9)$$

and the power factor can be expressed as

$$\lambda = \frac{1}{\sqrt{1 + (X/R)^2}} \quad (10)$$

Thus, a nonzero reactance (a nonzero susceptance) of the load is the cause of the power factor's decline. The Eqs. (6)–(9) are also valid for balanced 3-phase circuits with sinusoidal voltages and currents.

EFFECTS OF THE POWER FACTOR ON THE SUPPLY SOURCE

Differences in active power delivery at different power factors are illustrated with Fig. 2. Figure 2(a) shows the single-phase equivalent circuit of a supply source and an inductive load. Figures 2(b) and 2(c) show phasor diagrams of the voltage, current, and voltage drop on the inductance and resistance of the supply for two different PF, $\lambda = 0.95$ and $\lambda = 0.5$, but with the same load active power P . Active power P delivery at a low PF requires a higher rms value I of the supply current as compared with active power delivery at a higher PF. Consequently, the active power loss in the supply source, ΔP_s , equal to

$$\Delta P_s = R_s I^2 \quad (11)$$

increases as the PF declines. The same is true for the voltage drop on the internal impedance of the source, ΔU_s . Moreover, as can be seen in Fig. 2(c), the voltage drop on the reactance $\omega_1 L_s$ at a low PF contributes mainly to the magnitude of the difference between the load voltage u and internal voltage e of the supply. The voltage drop on the resistance R_s contributes to a phase-shift between these two voltages. It is almost opposite at a high PF, namely, as shown in Fig. 2(b), the voltage drop on the

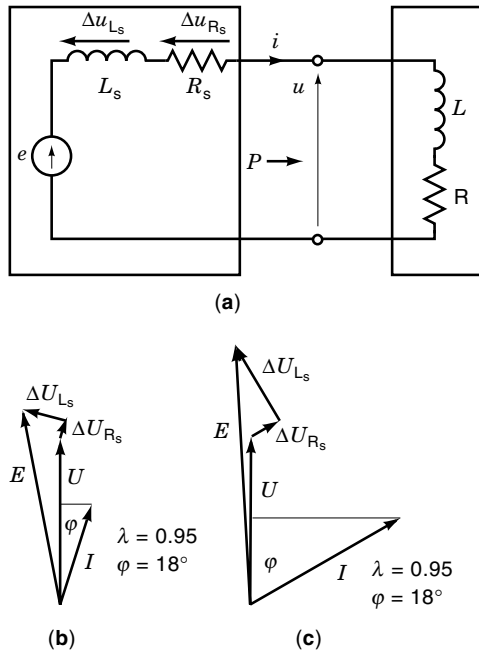


Figure 2. (a) Equivalent circuits of the load and the source, (b) phasor diagram of the circuit with a high power factor, equal to $\lambda = 0.95$, (c) phasor diagram of the circuit with a low power factor, equal to $\lambda = 0.5$. Active power P in both situations is the same but the supply current rms value I and voltage drop ΔU in the supply source are higher at a low power factor.

resistance contributes mainly to the magnitude of the voltage difference, while the voltage drop on the reactance contributes to the phase shift. The reactance of distribution systems is usually a few times higher than the resistance, and therefore the load voltage declines more strongly when the PF is low than when it is high. Consequently, a low PF has a number of harmful effects on distribution systems:

1. It increases the needed power ratings of the distribution equipment, thus increasing the investment cost.
2. It increases the active power loss in the distribution system, thus increasing the fuel cost.
3. It generally reduces the load voltage. This may require the elevation of the distribution voltage, thus increasing the operational cost needed for voltage regulation. Also, at higher values of the distribution voltage e the active power loss in the distribution system is higher.

These detrimental effects of a low PF, λ , are the reason for which it is measured for energy accounts between power utilities and customers. It is important to observe that all extra costs caused by low power factors are on the power utilities' side. Therefore, PF may affect the energy tariff, or customers may be obliged to pay penalties if the power factor of their loads is too low. This is also the cause for developing various methods of power factor improvement.

POWER FACTOR IN THREE-PHASE UNBALANCED SYSTEMS WITH SINUSOIDAL VOLTAGES AND CURRENTS

Power factor is the ratio of active to apparent power. The apparent power in 3-phase systems can be defined, however, in

various ways. When the system is balanced, these definitions result in the same value; differences appear only when the system is unbalanced. Consequently, the power factor in unbalanced systems depends on the chosen definition of apparent power.

The apparent power in 3-phase systems is defined according to the conclusion (1) of the joint committee of AIEE and NELE (presently, Edison Institute) in 1920. Thus, the apparent power can be defined as

$$S_G = \sqrt{P^2 + Q^2} \quad (12)$$

where Q is the reactive power of the load. This quantity is known as geometric apparent power. It can also be defined as

$$S_A = U_R I_R + U_S I_S + U_T I_T \quad (13)$$

where U_R , U_S , and U_T are the rms values of the line-to-ground voltages, and I_R , I_S , and I_T are the rms values of the line currents. This quantity is known as arithmetical apparent power. The controversy and confusion caused by the presence of different definitions of apparent power with respect to PF measurement is discussed in Ref. 2 and remain unsolved. The latest edition of the *Standard Dictionary of Electrical and Electronic Terms* (3) provides both definitions.

The controversy regarding apparent power and power factor is illustrated here with a numerical example. Figure 3 shows a 3-phase symmetrical source that supplies two customers with the same active power, $P = 100$ kW. Thus they are billed for the

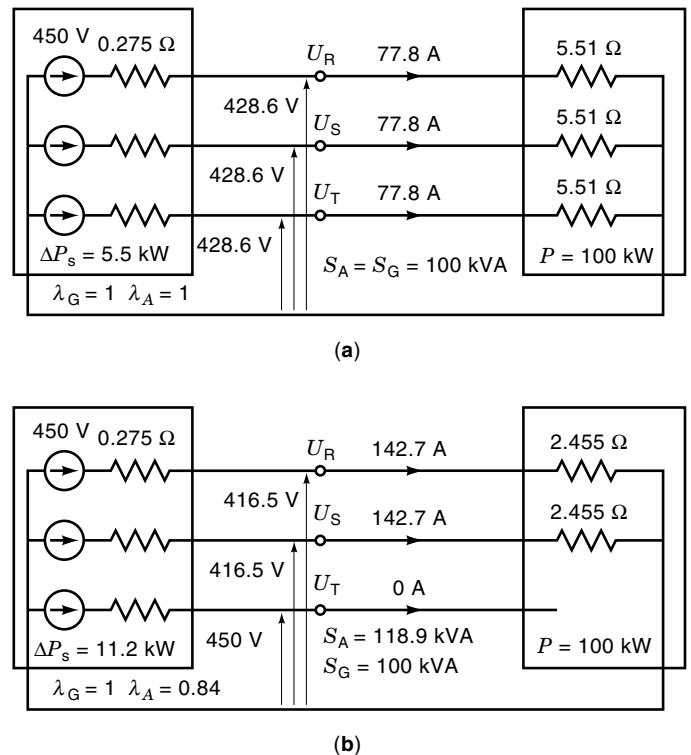


Figure 3. Two three-phase circuits with the same load active power $P = 100$ kW, but circuit (a) is balanced while circuit (b) is unbalanced. The load imbalance causes an increase of the active power loss. Arithmetic and geometric apparent powers and power factors in the balanced circuit are the same, but different in the unbalanced circuit.

same energy. However, one load is balanced, and the second is unbalanced. The unbalanced load requires higher current ratings of the supply source (142.7 A) than the source that supplies the balanced load (77.8 A). Moreover, the active power loss in the supply source is higher when supplying the unbalanced load (11.2 kW) than when supplying the balanced one (5.5 kW). The increase of the active power loss is comparable to the delivery of the same active power $P = 100$ kW in the balanced circuit with the power factor of the order of $\lambda = 0.7$. However, the geometric apparent power S_G is not affected by the load imbalance; $S_G = 100$ kVA and $\lambda_G = P/S_G = 1$. The arithmetic apparent power amounts to $S_A = 118.9$ kVA and $\lambda_A = P/S_A = 0.84$, that is, the arithmetic PF is higher than the PF of an RL load, which causes the same increase of the active power loss. Thus, geometric and arithmetic apparent powers do not have the same properties as the apparent power in single-phase circuits $S = UI$ which is the product of quantities that are responsible for the line and shunt active power loss, $R_s I^2$ and $G_s U^2$.

Three-phase distribution equipment is manufactured as symmetrical devices, thus the line resistance $R_R = R_S = R_T = R_s$. It was shown in Ref. 4 that the active power loss in such a device caused by line currents i_R, i_S , and i_T arranged in a vector

$$\mathbf{i} = \begin{bmatrix} i_R \\ i_S \\ i_T \end{bmatrix}$$

can be expressed as

$$\Delta P_s = R_s \frac{1}{T} \int_0^T (i_R^2 + i_S^2 + i_T^2) dt = R_s \frac{1}{T} \int_0^T \mathbf{i}^t \mathbf{i} dt = R_s \|\mathbf{i}\|^2 \quad (14)$$

where $\mathbf{i}^t$ denotes a transposed vector, and

$$\|\mathbf{i}\| = \sqrt{\frac{1}{T} \int_0^T \mathbf{i}^t \mathbf{i} dt} = \sqrt{\|i_R\|^2 + \|i_S\|^2 + \|i_T\|^2} \quad (15)$$

is the rms value of the 3-phase current vector. A symmetrical 3-phase resistive device with the line currents, i_R, i_S , and i_T , is shown in Fig. 4(a). It is equivalent, with respect to active power loss, to a single-phase device shown in Fig. 4(b) with the current rms value $\|\mathbf{i}\|$. Similarly, the active power loss caused by line-to-ground voltages, arranged in the vector

$$\mathbf{u} = \begin{bmatrix} u_R \\ u_S \\ u_T \end{bmatrix}$$

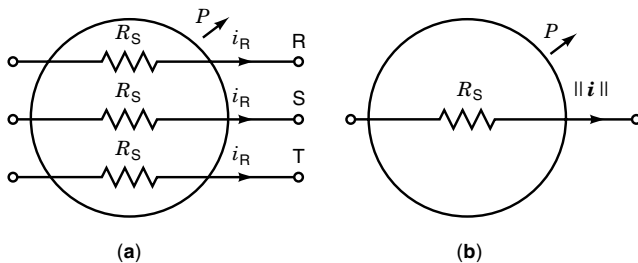


Figure 4. (a) A three-phase symmetrical device with asymmetrical currents i_R, i_S , and i_T and (b) a single-phase device equivalent with respect to the active power P at the current rms value $\|\mathbf{i}\|$.

can be expressed as

$$\Delta P_s = G_s \frac{1}{T} \int_0^T \mathbf{u}^t \mathbf{u} dt = G_s \|\mathbf{u}\|^2 \quad (16)$$

where

$$\|\mathbf{u}\| = \sqrt{\frac{1}{T} \int_0^T \mathbf{u}^t \mathbf{u} dt} = \sqrt{\|u_R\|^2 + \|u_S\|^2 + \|u_T\|^2} \quad (17)$$

is the rms value of a 3-phase voltage vector.

Since the rms values $\|\mathbf{i}\|$ and $\|\mathbf{u}\|$ are related to the active power loss in 3-phase supply sources similarly as the rms values U and I are related to active power loss in single-phase sources, then the apparent power in 3-phase systems with symmetrical distribution equipment should be defined as

$$S = \|\mathbf{u}\| \|\mathbf{i}\| \quad (18)$$

Such a definition was assumed in Ref. 4 for 3-phase systems with nonsinusoidal voltages and currents. For 3-phase systems with sinusoidal waveforms, such a definition was suggested by Buchholz (5) in 1922, but this definition is not referenced by the IEEE Standard (3).

The rms values $\|\mathbf{i}\|$ and $\|\mathbf{u}\|$ in the previous example are equal to $\|\mathbf{i}\| = 201.8$ A, and $\|\mathbf{u}\| = 741.2$ V; and apparent power $S = 149$ kVA. The active power loss in the supply source is equal to $\Delta P_s = R_s \|\mathbf{i}\|^2 = 0.275 \times (201.8)^2 = 11.2$ kW. The power factor using such a definition of apparent power is equal to $\lambda = 0.67$.

The increase of the supply current rms value and decline of the PF in 3-phase systems with sinusoidal voltages and currents is caused by the presence of the reactive current. The example considered above shows that PF may also decline in the absence of the reactive current only because of the load imbalance. This conclusion was reported by Lyon in 1920 (6). Reference 7 shows that this is because the current vector $\mathbf{i}$ contains not only the active and reactive currents, $\mathbf{i}_a$ and $\mathbf{i}_r$, but also an unbalanced current, $\mathbf{i}_u$. Therefore the supply current in three-phase systems can be decomposed into three currents, namely

$$\mathbf{i} = \mathbf{i}_a + \mathbf{i}_r + \mathbf{i}_u \quad (19)$$

Since the rms values of these currents fulfill the relationship

$$\|\mathbf{i}\| = \sqrt{\|\mathbf{i}_a\|^2 + \|\mathbf{i}_r\|^2 + \|\mathbf{i}_u\|^2} \quad (20)$$

then, the power factor can be expressed as

$$\lambda = \frac{P}{S} = \frac{\|\mathbf{u}\| \|\mathbf{i}_a\|}{\|\mathbf{u}\| \|\mathbf{i}\|} = \frac{\|\mathbf{i}_a\|}{\sqrt{\|\mathbf{i}_a\|^2 + \|\mathbf{i}_r\|^2 + \|\mathbf{i}_u\|^2}} \quad (21)$$

where the rms values of these currents depends on the load equivalent conductance G_e , equivalent susceptance B_e , and unbalanced admittance A ; expressed

$$\|\mathbf{i}_a\| = G_e \|\mathbf{u}\|, \quad \|\mathbf{i}_r\| = |B_e| \|\mathbf{u}\|, \quad \|\mathbf{i}_u\| = A \|\mathbf{u}\| \quad (22)$$

Equation (21) shows that the unbalanced current $\mathbf{i}_u$ contributes to the decline of the PF in a manner similar to the reactive current $\mathbf{i}_r$.

POWER FACTOR OF LINEAR LOADS WITH NONSINUSOIDAL VOLTAGE

An increase of the supply current rms value and decline of the PF in linear single-phase circuits with nonsinusoidal voltage is caused by (1) a phase shift of the voltage and current harmonics and (2) by a varying of the load conductance with harmonic frequency. The supply current i in such circuits may contain, according to Ref. 8, not only active and reactive currents i_a and i_r , but also a scattered current i_s , namely

$$i = i_a + i_s + i_r \quad (23)$$

The rms values of these currents fulfill the relationship

$$\|i\| = \sqrt{\|i_a\|^2 + \|i_s\|^2 + \|i_r\|^2} \quad (24)$$

Therefore, the power factor can be expressed as

$$\lambda = \frac{P}{S} = \frac{\|i_a\|}{\sqrt{\|i_a\|^2 + \|i_s\|^2 + \|i_r\|^2}} \quad (25)$$

The scattered current i_s contributes to a PF decline in a manner similar to the reactive current i_r . However, reactive shunt compensators are incapable of compensating scattered current.

POWER FACTOR OF NONLINEAR LOADS WITH SINUSOIDAL SUPPLY VOLTAGE

The conclusion that the PF can decline because of current distortion, without any phase shift, was drawn for the first time by Steinmetz (9) in 1892 and discussed in Ref. 10.

If the supply voltage is sinusoidal, then current distortion occurs when the load is nonlinear, and/or its parameters change periodically because of periodic switching. The supply current i of such a load may contain not only active and reactive currents i_a and i_r , but also a load generated harmonic current i_g , namely

$$i = i_a + i_r + i_g \quad (26)$$

Since the rms values of these currents fulfill the relationship

$$\|i\| = \sqrt{\|i_a\|^2 + \|i_r\|^2 + \|i_g\|^2} \quad (27)$$

the power factor can be expressed as

$$\lambda = \frac{P}{S} = \frac{\|i_a\|}{\sqrt{\|i_a\|^2 + \|i_r\|^2 + \|i_g\|^2}} \quad (28)$$

The load generated harmonic current i_g contributes to the decline of the PF in a manner similar to the reactive current i_r .

Active and reactive currents i_a and i_r are sinusoidal currents of the fundamental frequency, that is, $i_a + i_r = i_1$. Their rms values are commonly denoted by I_a , I_r , and I_1 . Hence, the PF can be expressed as

$$\lambda = \frac{\|i_a\|}{\sqrt{\|i_a\|^2 + \|i_r\|^2}} \frac{1}{\sqrt{1 + \frac{\|i_g\|^2}{\|i_a\|^2 + \|i_r\|^2}}} = \frac{\lambda_1}{\sqrt{1 + \delta_i^2}} \quad (29)$$

where

$$\lambda_1 = \frac{\|i_a\|}{\sqrt{\|i_a\|^2 + \|i_r\|^2}} = \frac{I_a}{\sqrt{I_a^2 + I_r^2}} = \cos \phi_1 \quad (30)$$

is known as a *displacement power factor*. The angle ϕ_1 is the phase-shift angle between the supply voltage and the current fundamental harmonic i_1 . The coefficient

$$\delta_i = \sqrt{\frac{\|i_g\|^2}{\|i_a\|^2 + \|i_r\|^2}} = \sqrt{\sum_{n=2}^{\infty} \left(\frac{I_n}{I_1}\right)^2} \quad (31)$$

where I_n denotes the rms value of the n -th order harmonic, specifies the supply current harmonic distortion.

The power factor is affected, according to Eqs. (29) and , by the phase shift ϕ_1 of the voltage and current fundamental harmonics and the current harmonic distortion δ_i . Therefore, the power factor λ alone does not provide (26) sufficient information relative to the loading conditions and the possibility of the power factor improvement. The displacement power factor λ_1 and the current harmonic distortion δ_i have to be known for this. Methods of improvement of the power factor have to take into account their effect both on the displacement power factor λ_1 and on the current distortion δ_i . In particular, capacitive compensation may be ineffective, since an improvement in the displacement power factor is accompanied usually with an increase in the current distortion.

POWER FACTOR MEASUREMENT

To measure the power factor, occasionally it is enough to measure the active power P , the supply voltage and current rms values U and I , and then calculate the apparent power S . The ratio of powers P and S provides the power factor. Such a measurement in single-phase circuits, shown in Fig. 5, provides the power factor whether or not the voltage and current waveforms are sinusoidal. However, such a measurement may provide an incorrect value of the power factor in 2-phase, 3-wire circuits and in 3-phase circuits when the voltages and currents are non-sinusoidal. When waveforms are sinusoidal and the load is balanced, then the meters connected as shown in Fig. 6 and Fig. 7 provide data needed for calculating the power factor in 2-phase and in 3-phase, 3-wire circuits. The PF in a balanced 2-phase circuit with meters connected as shown in Fig. 6 can be calculated as in a single-phase circuit, that is,

$$\lambda = \frac{P}{UI} \quad (32)$$

In the case of a three-phase balanced circuit with meters

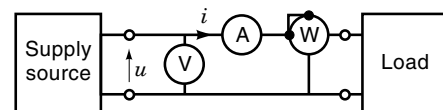


Figure 5. Meter arrangement for power factor calculation in single-phase circuits.

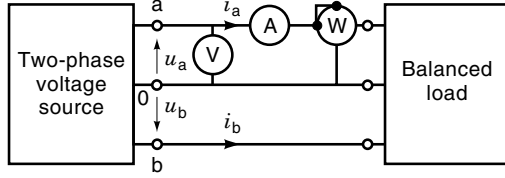


Figure 6. Meter arrangement for power factor calculation in balanced two-phase, three-wire circuits.

connected as shown in Fig. 7, the power factor is equal to

$$\lambda = \frac{W_1 + W_2}{\sqrt{3}UI} \quad (33)$$

where W_1 and W_2 are the values measured by wattmeters. In two-phase, three-wire balanced circuits, the power factor can also be calculated from data provided by two wattmeters connected as shown in Fig. 8. Since the voltage coil of the wattmeter W_2 is supplied with a voltage shifted by $\pi/2$, this wattmeter measures the value

$$W_2 = -UI \cos(\varphi + \pi/2) = UI \sin \varphi \quad (34)$$

the ratio

$$\frac{W_2}{W_1} = \frac{UI \sin \varphi}{UI \cos \varphi} = \tan \varphi \quad (35)$$

Hence

$$\lambda = \cos \left(\text{atan} \left\{ \frac{W_2}{W_1} \right\} \right) \quad (36)$$

Similarly, the power factor can be measured with only two wattmeters in 3-phase balanced systems. As shown in Ref. 11, wattmeters connected for active power measurement also provide data sufficient for the PF calculation. Wattmeters connected as shown in Fig. 9 measure the values

$$W_1 = UI \cos(\varphi - 30^\circ) \text{ and } W_2 = UI \cos(\varphi + 30^\circ) \quad (37)$$

respectively, where U is the rms value of the line-to-line voltage. Hence, the power factor is equal to

$$\lambda = \cos \left(\text{atan} \left\{ \sqrt{3} \frac{W_1 - W_2}{W_1 + W_2} \right\} \right) \quad (38)$$

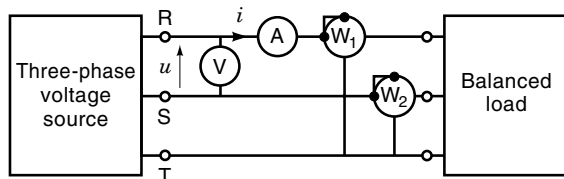


Figure 7. Meter arrangement for power factor calculation in balanced three-phase, three-wire circuits.

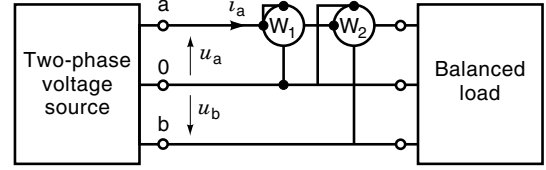


Figure 8. Wattmeter connection that enables PF calculation without measuring voltage and current rms values in two phase, three-wire balanced circuits.

POWER FACTOR METERS

Until a sufficient progress in electronics was made so that the measurement of the power factor according to definition (1) was possible, all PF meters were built only as meters of the phase-shift angle ϕ between voltage and current. Such meters are scaled in the cosine of this angle, that is, in power factor, or this angle is electronically converted to the power factor.

There are two main types of such power factor meters:

1. Electrodynamic PF meters (iron-vane or crossed-coil PF meters),
2. Electronic PF meters, built as analog or digital instruments.

IRON-VANE PF METERS

An iron-vane PF meter was the first PF meter built in the USA, patented in 1899 (12). There are a number of varieties of construction of iron-vane PF meters. Essentially, such meters are 3-phase instruments; one of them is shown in Fig. 10. It is built of three stationary coils supplied with 3-phase voltage, a stationary coil with the line current, and a movable soft-iron vane with a pointer to a PF-scale. The resistors connected in series with the 3-phase coils are sufficiently large so that the coil currents are approximately in phase with the line-to-ground voltages. The conductors in the coil are distributed such that the radial component of the magnetic field intensity in Θ direction changes as the sine of direction angle Θ , that is, for the coil supplied with the voltage $u_R = U_m \cos(\omega t + \alpha)$, the radial component of the magnetic field intensity is equal to

$$H_R(t, \Theta) = k u_R \sin \Theta = H_{vm} \cos(\omega t + \alpha) \sin \Theta \quad (39)$$

where k is a dimensional coefficient. The voltage coils are distributed around axes rotated by 120° and are supplied with the

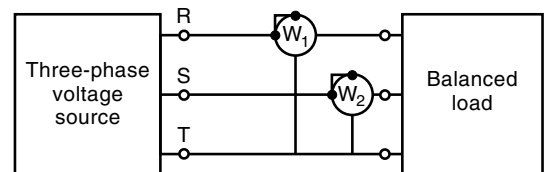


Figure 9. Wattmeter connection that enables PF calculation without measuring voltage and current rms values in three-phase, three-wire balanced circuits.

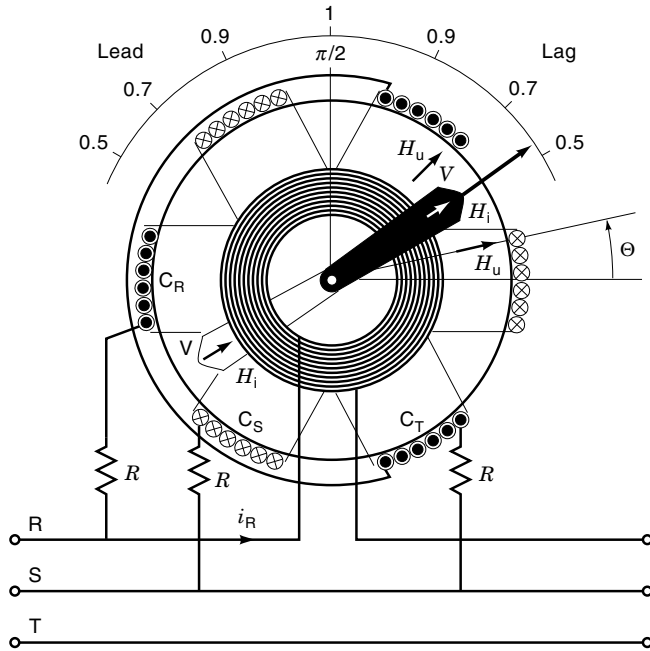


Figure 10. Iron-vane PF meter and its connection to a three-phase system. Coils C_R , C_S , and C_T create a rotating magnetic field H_u , which deflects the iron vane V with magnetic field H_i , which is created by stationary coil by the angle Θ equal to the phase-shift angle ϕ .

voltage shifted by 120° . Hence

$$H_S(t, \Theta) = H_m \cos(\omega t + \alpha - 120^\circ) \sin(\Theta + 120^\circ) \quad (40)$$

$$H_T(t, \Theta) = H_m \cos(\omega t + \alpha - 120^\circ) \sin(\Theta + 120^\circ) \quad (41)$$

The resultant radial component in Θ direction, $H_v(t, \Theta)$, of the magnetic field intensity created by the voltage coils is equal to

$$\begin{aligned} H_v(t, \Theta) &= H_R(t, \Theta) + H_S(t, \Theta) + H_T(t, \Theta) \\ &= \frac{3}{2} H_{vm} \sin(\omega t + \alpha + \Theta) \end{aligned} \quad (42)$$

The maximum of this radial component of the magnetic field intensity occurs at direction

$$\Theta = \pi/2 - \alpha - \omega t \quad (43)$$

that is, it rotates with the radial frequency ω . Because of this, these meters are often referred to as PF meters with a rotating field. This term does not distinguish them from crossed-coil PF meters, however, since the crossed coils in those meters also induce a rotating magnetic field.

The current coil induces a magnetic field in the soft-iron vane. Since the vane forms an asymmetrical magnetic path, the maximum field intensity is along the vane. If the current in line R is equal to $i_R = I_m \cos(\omega t + \alpha - \phi)$, the magnetic field intensity along the vane changes as

$$H_c(t) = H_{cm} \cos(\omega t + \alpha - \phi) \quad (44)$$

The vane positions in the rotating field in such a direction Θ that the maximum of its magnetic field H_c , which occurs at

$\omega t = -\alpha + \phi$, coincides with the maximum of the rotating field H_v , in that direction, hence,

$$\Theta = \pi/2 - \alpha - (-\alpha + \phi) = \pi/2 - \phi \quad (45)$$

Thus, the pointer deflects from the $\pi/2$ angle by the phase-shift angle ϕ . Cosine values are allocated on the scale to the phase-shift angles ϕ , so that the device measures the power factor of the load, both for lagging and leading current. Due to properties of the cosine function, the scale of such meters cannot be uniform. It is denser for lower power factors than for higher ones.

Iron-vane PF meters are essentially 3-phase devices, though they also can be built for 2-phase systems, and, equipped with additional phase-shift circuits, even for single-phase systems.

However, accuracy of iron-vane PF meters is low because the space distribution of magnetic fields created by voltage coils may differ substantially from the required sinusoidal distribution. Also, the sine function varies slowly around its maximum. Consequently, because of the friction of the movable vanes, fields H_v and H_c may not align accurately. Iron-vane PF meters for 2-phase and for single-phase systems are even less accurate than 3-phase PF meters, since the magnitude of the magnetic field intensity H_v cannot be kept constant during the field rotation (because it is induced with only two coils). In the case of PF meters for single-phase circuits, a capacitor or an inductor is connected instead of the series resistor to obtain a phase-shift between the magnetic fields produced by these two coils.

CROSSED-COIL PF METERS

Crossed-coil power factor meters are developments from the phase-shift meter developed by Tuma (13) in 1897. Detailed analyses can be found in Refs. 14 and 15.

A single-phase crossed-coil PF meter is shown in Fig. 11. It is built of two movable coils, C_a and C_b , pivoted to rotate freely with a pointer inside of a stationary coil, C_c . Movable coil planes are deflected by a fixed angle Δ , hence these devices are known as *crossed-coil* meters. Movable coils are supplied with the load voltage in such a way that no mechanical torque is exerted on the movable part. Their currents i_a and i_b are mutually shifted, since one coil is supplied through a series resistor R and the other through a series inductor L . The stationary coil induces a magnetic field proportional to the load current. The stationary coil is constructed in such a way that the magnetic field around the movable coils is uniform.

At the supply voltage $u = U_m \sin \omega t$, the currents in coils C_a and C_b are equal to

$$i_a = I_{am} \sin(\omega t - \alpha) \quad (46)$$

$$i_b = I_{bm} \sin(\omega t - \beta) \quad (47)$$

The instantaneous torque exerted on the coil C_a , deflected from the coil C_c plane by angle Θ , at the load current $i = I_m \sin(\omega t - \phi)$, is equal to

$$\tau_a = k_a i_a i \sin \Theta = k_a I_{am} I_m \sin(\omega t - \alpha) \sin(\omega t - \phi) \sin \Theta \quad (48)$$

where k_a is a coefficient dependent on the windings and geometry of coils C_a and C_c . The average torque in one cycle T of the

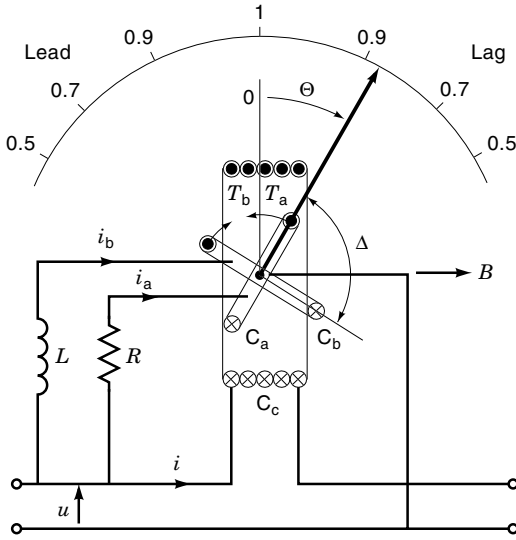


Figure 11. Crossed-coil PF meter and its connection to a single-phase circuit. Movable coils with currents i_a and i_b are deflected in the magnetic field of the stationary coil. Torques T_a and T_b are in equilibrium at deflection angle Θ equal to the phase-shift angle ϕ .

supply voltage is equal to

$$T_a = \frac{1}{T} \int_0^T \tau_a dt = T_{am} \sin \Theta \cos(\varphi - \alpha), \quad T_{am} = \frac{1}{2} k_a I_{am} I_m \quad (49)$$

Similarly, the average torque exerted on the coil C_b deflected with respect to coil C_a by angle Δ is equal to

$$T_b = T_{bm} \sin(\Delta + \Theta) \cos(\varphi - \beta), \quad T_{bm} = \frac{1}{2} k_b I_{bm} I_m \quad (50)$$

If the currents in coils C_a and C_b are such that maximum torques T_{am} and T_{bm} are mutually equal, which requires that $k_a I_{am} = k_b I_{bm}$, then the movable part with the pointer is in equilibrium when

$$T_a = T_b \quad (51)$$

at such a deflection angle Θ that

$$\sin \Theta \cos(\varphi - \alpha) - \sin(\Delta + \Theta) \cos(\varphi - \beta) = 0 \quad (52)$$

In an ideal case, when $\alpha = 0$ and $\beta = 90^\circ$, coil C_b should be deflected with respect to coil C_a by the angle $\Delta = 90^\circ$. Then, Eq. (52) simplifies to the form

$$\sin \Theta \cos \varphi - \cos \Theta \sin \varphi = 0 \quad (53)$$

and results in the deflection angle

$$\Theta = \varphi \quad (54)$$

equal to the phase-shift angle between the supply voltage and current.

Because of some inductance in the coil C_a circuit there is, however, a phase shift between the supply voltage and the coil

current i_a , thus $\alpha > 0$. Also, because of a resistance in the coil C_b circuit, the phase-shift angle β between the coil current i_b and the supply voltage is lower than 90° . In such a case, even at $\phi = 0$, the movable part deflects by an angle Θ_0 which satisfies the equation

$$\sin \Theta_0 \cos \alpha - \sin(\Delta + \Theta_0) \cos \beta = 0 \quad (55)$$

Hence, it is equal to

$$\Theta_0 = \tan^{-1} \left(\frac{\sin \Delta \cos \beta}{\cos \alpha - \cos \Delta \cos \beta} \right) \quad (56)$$

When the angle between coils C_a and C_b is chosen such that

$$\Delta = \beta - \alpha \quad (57)$$

thus the initial deflection Θ_0 becomes independent of Δ and it is equal to

$$\Theta_0 = 90^\circ - \beta \quad (58)$$

If condition (57) is fulfilled, the equilibrium Eq. (53) has the solution

$$\Theta = \Theta_0 + \varphi \quad (59)$$

The meter can be scaled in power factor with respect to the initial deflection angle Θ_0 , which can be used for calibrating the meter rotating the stationary coil by that angle, that is, to $\lambda = 1$ at a resistive load.

Figure 12 shows the structure of a three-phase crossed-coil PF meter. Since there are voltages shifted mutually in three-phase circuits, the inductor is not needed for shifting the current in the coil C_b . Such a device measures the power factor in balanced circuits with sinusoidal and symmetrical supply

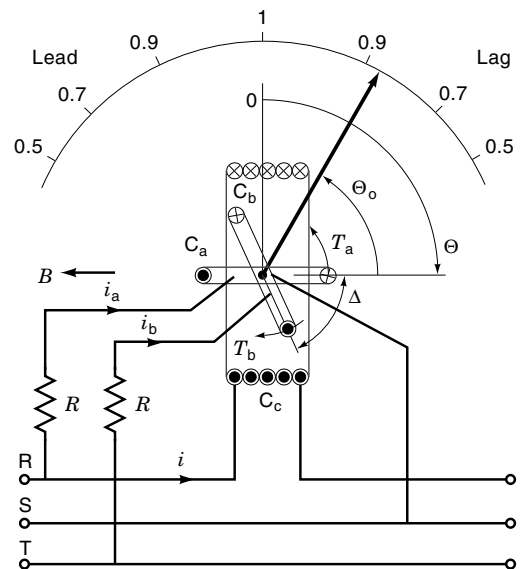


Figure 12. Three-phase crossed-coil PF meter and its connection to the circuit. The pointer is permanently deflected from the coil C_a plane by 60° .

voltage. In such a case, the phase-shift angles of the coil currents are $\alpha = -30^\circ$, $\beta = 30^\circ$, respectively. Consequently, coil C_b should be deflected with respect to coil C_a by the angle $\Delta = \beta - \alpha = 60^\circ$. The initial deflection angle, Θ_0 , according to Eq. (58), amounts to $\Theta_0 = 60^\circ$. To simplify the scale and the meter design, the pointer is deflected by this angle from the coil C_a plane, as shown in Fig. 12.

ELECTRONIC POWER FACTOR METERS BASED ON PHASE-SHIFT ANGLE MEASUREMENT

Electronic PF meters are built as analog, digital, or hybrid devices. Analog devices measure the phase-shift angle as an interval of time by detecting the zero crossings of the supply voltage and current. This interval of time, referenced to the voltage period T , is converted next to the cosine value, hence, to the power factor. It can be done with continuous or with digital signals, consequently, the PF meter is referred to as an analog or as a hybrid device. Digital PF meters provide the power factor as a result of a digital algorithm on sequences of digital samples of the supply voltage and current.

The structure of one of the first analog electronic PF meter, described in Ref. 16, is shown in Fig. 13. A synchronous switch, built of four diodes and resistors, is the main component of the device. When current i_b is positive, diodes D1 and D2 are conducting and the voltage $u_x = 0$. The voltage at the voltmeter is equal to

$$u_v = u_x + u_a = u_a \quad (60)$$

When current i_b is positive, diodes D3 and D4 are conducting and the voltage $u_y = 0$. The voltage at the voltmeter is equal to

$$u_v = u_y - u_a = -u_a \quad (61)$$

and has the waveform shown in Fig. 14. The voltmeter measures the average value of this voltage. It is equal to

$$\begin{aligned} \bar{u}_v &= \frac{1}{\pi} \int_{\varphi}^{\pi+\varphi} u_a d(\omega t) = \frac{1}{\pi} \int_{\varphi}^{\pi+\varphi} U_{am} \sin \omega t d(\omega t) \\ &= \frac{2U_{am}}{\pi} \cos \varphi = k\lambda \end{aligned} \quad (62)$$

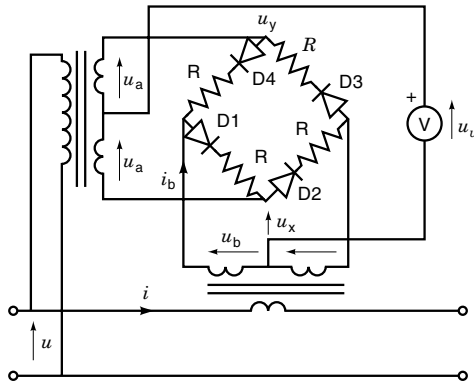


Figure 13. Electronic PF meter with synchronous switch. Diodes D1 and D2 are conducting at positive current i_b . Diodes D3 and D4 are conducting at negative current i_b .

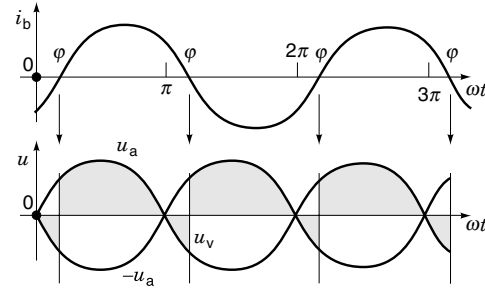


Figure 14. Voltage at the voltmeter provided by the synchronous switch, (shown in Fig. 13) that is controlled by the load current. The average value of voltage u_a is proportional to the power factor of the load.

Thus, the voltmeter can be calibrated directly in power factor. It has a linear scale. Unfortunately, changes in the voltage amplitude, U_{am} , affect the measurement result. Also, lagging and leading PF cannot be distinguished. Frequency division is suggested in (Refs. 17–19) to eliminate the dependence of the measurement result on the voltage. A voltage to frequency converter (VFC) converts the mean value of the voltage u_v into sequence of binary pulses of frequency

$$f_1 = k_1 U_{am} \cos \varphi \quad (63)$$

At the same time, voltage u_a is rectified and its mean value converted to binary pulses of frequency

$$f_2 = k_2 U_{am} \quad (64)$$

by another VFC. The device has two binary counters, C1 and C2. Counter C2 specifies the time interval T_2 needed for counting N_2 pulses of frequency f_2 , which is expressed

$$T_2 = \frac{N_2}{f_2} \quad (65)$$

The number of pulses of frequency f_1 counted in the interval T_2 is equal to

$$N_1 = T_2 f_1 = N_2 \frac{f_1}{f_2} = \frac{k_1}{k_2} \cos \varphi = k\lambda \quad (66)$$

Such a device can be directly equipped with a digital display for reading the PF value.

Figure 15 shows the structure of another analog PF meter, described in Ref. 20 with the measuring results independent of the supply voltage. A reference dc voltage U_{ref} is applied to a voltmeter through two electronically controlled switches. The switch S_A turns dc voltage U_{ref} ON when the supply voltage u changes to a positive value, and turns it OFF when the supply current i changes to a negative value, as shown in Fig. 16. The switch S_B turns the voltage U_{ref} ON when the supply current i changes to a positive value, and turns it OFF when the supply voltage u changes to a negative value. The average value of the voltmeter voltage, $u_v = u_A - u_B$, is equal to

$$\bar{u}_v = U_{ref} \frac{\varphi}{90} = k\varphi; -90 \text{ deg} < \varphi < 90 \text{ deg} \quad (67)$$

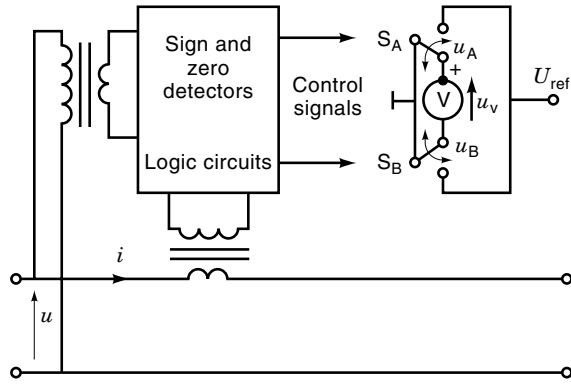


Figure 15. Electronic PF meter with sign and zero detectors of the supply voltage and current. A dc voltmeter measures the value proportional to the phase-shift angle ϕ both for lagging and leading power factor.

and a voltmeter, scaled in cosine value of the angle ϕ , with zero in the middle, provides both the lagging and leading power factor. However, its scale is not uniform. It becomes denser as the phase-shift angle approaches $\pm 90^\circ$. It is inconvenient in the case of an analog meter, that is, with a voltmeter scaled in PF. However, the power factor cannot be displayed digitally. A non-linear converter of the voltage to its cosine value is needed in such a case. Reference 21 describes such a converter for power factor measurement applications.

ELECTRONIC POWER FACTOR METERS BASED ON THE MEASUREMENT OF THE RATIO OF ACTIVE TO APPARENT POWER

When the voltage and current waveforms are nonsinusoidal and/or the 2-phase or 3-phase system is asymmetrical, then data provided by PF meters, that are based on phase-shift angle measurement are meaningless. The PF meter has to operate

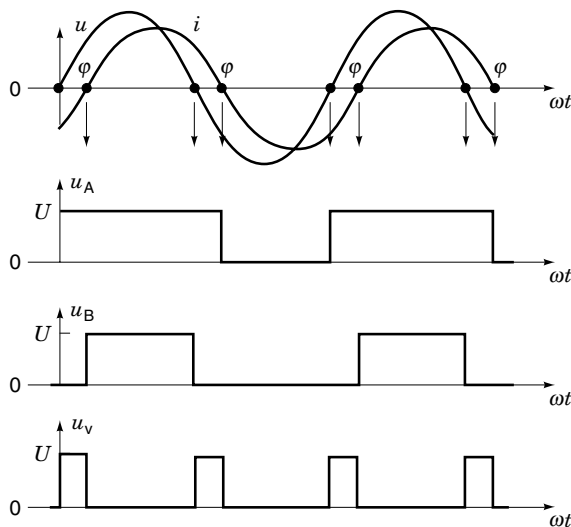


Figure 16. Generating switching pulses by the sign and zero detectors and logic circuits in the PF meter shown in Fig. 15.

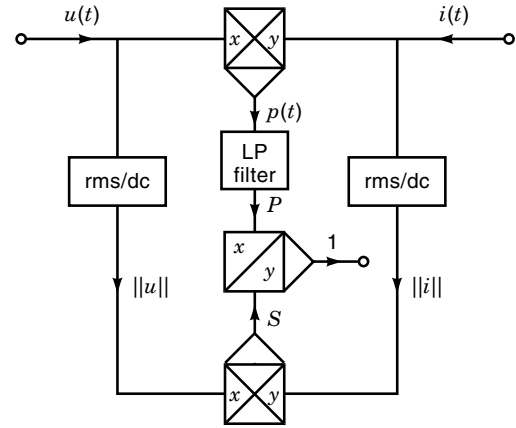


Figure 17. Structure of PF meter built of rms/dc converters, analog multipliers, divider, and a low-pass filter.

in such a case according to power factor definition (1). It has to measure the ratio of the active and apparent power. Analog, binary, and digital signal technology can be employed for this. There is, however, the problem of selection of the apparent power definition in asymmetrical 3-phase systems. When the system is symmetrical, PF can be measured with a single-phase power factor meter.

Single-phase analog PF meters can be built of two integrated analog multipliers, a divider, two rms-to-dc converters, and a low-pass filter. The supply voltage and current have to be conditioned to voltage signals, usually bounded by ± 10 V with voltage and current converters. Figure 17 shows the basic structure of an analog PF meter. Conditioning and supply circuits are not shown. Arrows denote voltage signals proportional to various quantities in the device. Dimensional coefficients are omitted. The output voltage can be used for PF control or measured by a voltmeter scaled in PF. Such a meter cannot have PF value $\lambda = 1$ in the middle of the scale, since definition (1) does not specify whether the PF is lagging or leading. Also, in circuits with nonsinusoidal waveforms these terms are meaningless. Analog PF meters can also be built with a digital display, as discussed in Refs. (22,23).

An operation principle of a hybrid, that is, an analog and digital PF meter, is presented in Ref. 24. Two converters of the supply voltage and current provide voltages $k_u u$ and $k_i i$, where k_u and k_i are conversion coefficients, to a rms value meter. It measures the rms values $k_u |u|$ and $k_i |i|$, as well as, the rms value of the sum $u_s = k_u u + k_i i$ and the difference $u_d = k_u u - k_i i$. These four rms values are transferred to a computer. The difference of the squares of the rms values of voltages u_s and u_d , calculated by the computer gives

$$\|u_s\|^2 - \|u_d\|^2 = 4k_u k_i P \quad (68)$$

This value divided by $4k_u |u| k_i |i|$ is the PF of the load.

Figures 18 and 19 show structures of an analog-to-binary (A/B) signal converter and a PF meter based on binary signal technology, described in Ref. 25. A binary voltage signal x_B is a signal that has only two values, U and $-U$. Information on the magnitude of the analog input quantity $x(t)$ is coded in the duty factor $d(t)$, actually the ratio of time interval τ_+ (when $x_B = U$) to the switching cycle time $\tau = \tau_+ + \tau_-$, where τ_- is the

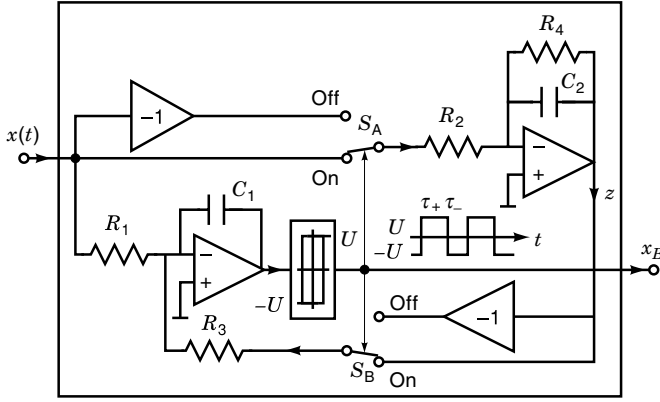


Figure 18. Converter of analog signal $x(t)$ to its binary representation x_B . Information on the analog signal's instantaneous value is coded in the duty factor of the binary signal.

time interval when $x_B = -U$, namely

$$d(t) = \frac{\tau_+}{\tau_+ + \tau_-} \quad (69)$$

Electronically controlled switches S_A and S_B are turned ON when $x_B = U$ and turned OFF when $x_B = -U$, consequently, capacitors C_1 and C_2 are charged with positive or negative voltages x and z . The charges on these capacitors must remain unchanged during each operation cycle, that is, time interval τ . The charging and discharging equilibrium occurs at the duty factor

$$d(t) = \frac{1}{2} - \frac{1}{2}k \frac{x(t)}{\|x\|} \quad k = \sqrt{\frac{R_2 R_3}{R_1 R_4}} \quad (70)$$

and is dependent on the instantaneous value $x(t)$ of analog signal normalized with respect to its rms value $|x|$.

Binary representations u_B and i_B of the supply voltage and current are applied to a logical EXCLUSIVE OR gate as shown in Fig. 19. The duty factor $d_E(t)$ of the output signal of this gate is equal to

$$d_E(t) = d_u(t)\bar{d}_i(t) + d_i(t)\bar{d}_u(t) \quad (71)$$

where $\bar{d}$ denotes the duty factor of the negative binary signal x_B . Since $d(t) + \bar{d}(t) = 1$, then

$$\bar{d}(t) = \frac{1}{2} + \frac{1}{2}k \frac{x(t)}{\|x\|} \quad (72)$$

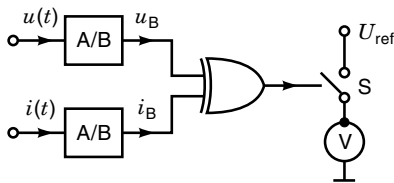


Figure 19. Electronic PF meter built of two analog-to-binary (A/B) converters, an EXCLUSIVE OR gate with an electronically controlled switch, and a dc voltmeter supplied from a source of reference voltage U_{ref} .

When the duty factors $d_u(t)$ and $d_i(t)$ of binary representations of the supply voltage and current are expressed in the form of Eq. (69), then Eq. (70) results in

$$d_E(t) = 1 - k_u k_i \frac{u(t)i(t)}{\|u\| \|i\|} = 1 - \frac{p(t)}{S} \quad (73)$$

since the coefficients k_u and k_i should be chosen such that $k_u k_i = 1$. Thus, a dc voltmeter supplied from a reference voltage U_{ref} through a switch S can be scaled in power factor. If the EXCLUSIVE OR gate ON signal opens the switch, the value $\lambda = 1$ is at the right end of the scale; if the gate ON signal closes the switch, the value $\lambda = 1$ is at the left end.

A single-phase, digital PF meter provides the power factor calculated from sequences of uniformly distributed digital samples u_n and i_n of the supply voltage and current. The meter is built of voltage and current converters that provide conditioned voltages in the range of, usually, ± 10 V; sample and hold (S&H) circuits, analog to digital converter (ADC), and an arithmetic unit, usually in the form of a microcontroller or microcomputer with a digital display or a printer.

The sampling frequency should be more than twice the frequency of the highest harmonic order of the supply voltage and current. If N denotes the number of samples in one cycle of the voltage, then the power factor in a single-phase or symmetrical 3-phase circuit is calculated from the formula

$$\lambda = \frac{P}{S} = \frac{\frac{1}{N} \sum_{n=0}^{N-1} u_n i_n}{\sqrt{\frac{1}{N} \sum_{n=0}^{N-1} u_n^2} \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} i_n^2}} = \frac{\sum_{n=0}^{N-1} u_n i_n}{\sqrt{\sum_{n=0}^{N-1} u_n^2} \sqrt{\sum_{n=0}^{N-1} i_n^2}} \quad (74)$$

Such a digital PF meter can be very accurate but expensive. Such a measurement is inexpensive, however, taking into account that the PF measurement is usually accompanied by a variety of other measurements, recordings, and analysis.

BIBLIOGRAPHY

1. Joint Committee of AIEE and NELE, Apparent power in three-phase systems, *Trans. AIEE*, **39**: 1450, 1920.
2. P. A. Borden and M. F. Behar, Measurements on unbalanced electrical systems, *Instruments*, 185–191, 1934.
3. *The New IEEE Standard Dictionary of Electrical and Electronics Terms*, New York: IEEE, 1992.
4. L. S. Czarnecki, Orthogonal decomposition of the current in a three-phase nonlinear asymmetrical circuit with nonsinusoidal voltage, *IEEE Trans. Instrum. Measur.*, **IM-37**: 30–34, March 1988.
5. F. Buchholz, Die drehstrom-scheinleistung bei ungleichmabiger belastung der drei zweige, *Licht und Kraft*, **2**: 1922.
6. W. V. Lyon, Reactive power and unbalanced circuits, *Electrical World*, **75** (25): 1417–1420, June 1920.
7. L. S. Czarnecki, Power related phenomena in three-phase unbalanced systems, *IEEE Trans. Power Deliv.*, **10**: 1168–1176, 1995.
8. L. S. Czarnecki, Considerzations on the reactive power in nonsinusoidal situations, *IEEE Trans. Instrum. Measur.*, **IM-34**: 399–404, 1984.
9. Ch. P. Steinmetz, Findend eine Phasenverschiebung im Wechselstrom Lichtbogen statt? *Elektrische Zeitschrift*, **42**: 567–568, Berlin, 1892.

10. A. F. Ganz, The physical meaning of power factor, *J. Franklin Inst.*, **429**: 1906.
11. D. G. Fink and J. M. Carroll, *Standard Handbook for Electrical Engineers*, New York: McGraw-Hill, 1968, 35–45.
12. P. MacGahan, Power factor meter, *USA Patent No.* 695 913, 1898.
13. J. Tuma, Ein phasenmessapparat fur wechselstrom, *Sitzb. Akad. Wiss., Wien*, **106**: 521, 1897.
14. F. A. Laws, *Electrical Measurements*, New York: McGraw-Hill, 1938, 545–576.
15. N. P. Millar, Crossed-coil power-factor meter, *AIEE Trans. Electr. Eng.*, **63**: 294–301, 1944.
16. B. M. Oliver and J. M. Cage, *Electronic measurements and instrumentation*, New York: McGraw-Hill, 1971.
17. M. A. H. Abdul-Karim and D. H. Boghosian, A digital power-factor meter design based on binary rate multiplication techniques, *IEEE Trans. Instrum. Meas.*, **IM-29**: 435–438, 1980.
18. M. S. M. Al-Ani and M. A. H. Abdul-Karim, Digital power factor meter based on voltage-to-frequency conversion, *Int. J. Electron.*, **52** (5): 463–470, 1982.
19. S. M. R. Taha and M. A. H. Abdul-Karim, A novel digital power-factor meter design, *IEEE Trans. Instrum. Meas.*, **IM-35**: 638–640, 1986.
20. T. K. M. Babu and Tak Wing Wong, Novel low-cost power-factor meter, *Int. J. Electron.*, **67** (1): 147–151, 1989.
21. B. A. Hafeth and M. A. H. Abdul-Karim, Digital power factor meter based on nonlinear analogue-to-digital conversion, *Int. J. Electron.*, **58** (3): 513–519, 1985.
22. R. M. Inigo, An electronic energy and average power-factor meter with controllable nonuniform rate, *IEEE Trans. Ind. Electron. Control Instrum.*, **IECI-27**: 271–278, November 1980.
23. J.-C. Wu and H.-L. Jou, Fast response power factor detector, *IEEE Trans. Instrum. Meas.*, **44**: 919–922, August 1994.
24. P. Braga and D. Slomovitz, RMS voltmeter based power and power-factor measuring system, *Int. J. Electron.*, **75** (3): 561–565, 1993.
25. P. N. Miljanic, B. Stojanovic, and V. Petrovic, A novel method for the measurement of power, power factor, rms, average values of voltage and current of distorted power frequency waves, *IEEE Trans. Instrum. Meas.*, **IM-29**: 432–434, December 1980.
26. D. Sharon, Power factor definitions and power transfer quality in nonsinusoidal situations, *IEEE Trans. Instrum. Meas.*, **45**: 728–733, 1996.

LESZEK S. CZARNECKI
Louisiana State University

ELECTRICAL ENERGY MEASUREMENT

Energy can be defined as the capacity to do work, while power can be defined as the rate of doing work. There are different types of energy, including electrical energy, heat energy, nuclear energy, and solar energy. Energy may be converted from one form to another; for instance, electricity is often generated in power plants by converting hydraulic or fuel energy into electrical energy. Most vehicles are driven using energy derived from hydrocarbon fuels, and cooking is often done using electrical energy. Figure 1 shows the rapid increase of energy consumption in recent decades which helps us to get an idea of our ever increasing dependence on energy. Therefore, energy measurement becomes an extremely important topic for scientists and engineers. In this article, we will deal with different aspects of electrical energy measurement. Generally

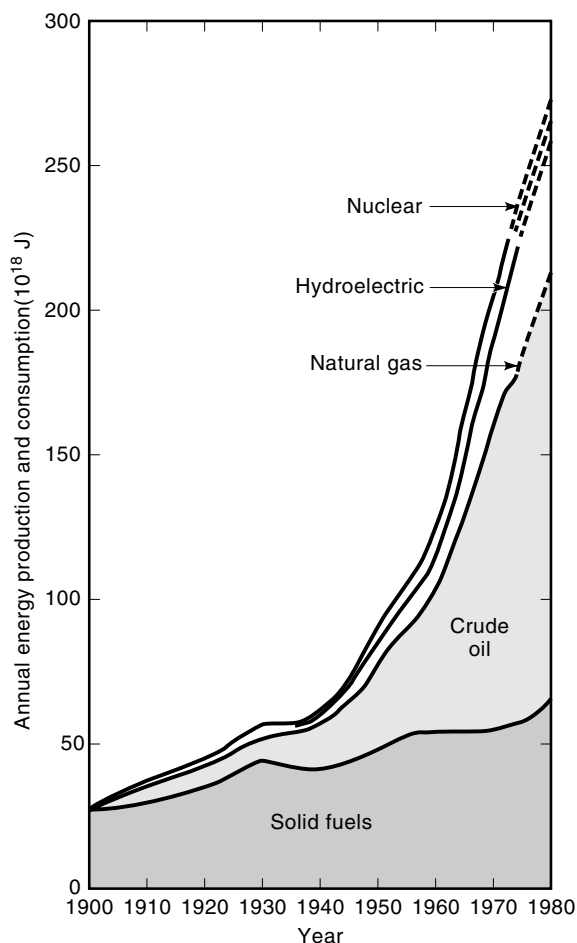


Figure 1. Graph depicting our increasing dependence on energy.

speaking, the measurement of energy is essentially the same as the measurement of power, except that the instrument must not merely indicate the power, but must take into account the length of time for which this power supply is maintained (1). Energy meters can be broadly classified into (1) electrolytic meters, (2) clock meters, (3) motor meters, and (4) electronic or microprocessor-based meters.

THEORY

Before discussing energy measurement itself, let us mathematically define electrical energy with respect to power. This will help us to understand energy measurement in greater detail. In this context it is desirable to consider alternating current (ac) and direct current (dc). Alternating current varies in magnitude and direction with time. However, dc current remains constant with time. Now, dc power (P) is defined as the product of voltage and current, that is,

$$P = VI \quad (1)$$

where V is the dc voltage in volts and I is the dc current in amperes. The direction of positive current flow is defined to be the direction a positive charge flows (or would flow if possible), and positive power is delivered when positive current flows from a

point of higher voltage or potential to a point of lower voltage. Since V or I may be either positive or negative, power may itself be positive or negative. Positive power implies power delivered to a device or load, while negative power implies that power has been generated, that is, converted from some other form to electrical. Electrical power is typically expressed in watts.

In many systems the voltage and current are proportional, and are related by Ohm's law, $V = IR$, where R is the load resistance in ohms. Using Ohm's law, Eq. (1) may be rewritten as

$$P = I^2 R \quad (2)$$

If the voltage and current vary with time, the instantaneous power (p) is defined as

$$p = vi \quad (3)$$

where v is the instantaneous voltage and i is the instantaneous current. Typically power is distributed to consumers in sinusoidal form, so that both the voltage and current are assumed to be sinusoidal and have the same frequency. An important measure of power, especially for periodic voltages and currents, is called average power. The average power is equal to the average rate at which energy is absorbed by an element, and it is independent of time (2). This is the power monitored by the electric utility company in determining monthly electricity bills. The average power (P_a) associated with a periodic instantaneous power signal is given by

$$P_a = \frac{1}{T_p} \int_{t_0}^{t_0+T_p} p \, dt = \frac{1}{T_p} \int_{t_0}^{t_0+T_p} vi \, dt \quad (4)$$

where T_p is the period of both v and i , and t_0 is arbitrary. If the effective or root mean square (rms) values of the instantaneous voltage v and instantaneous current i are denoted by $|V|$ and $|I|$, and the phase angles of the voltage and current are θ_v and θ_i , then the average power delivered to the load is

$$P_a = |V| |I| \cos \theta \quad (5)$$

where $\theta = \theta_v - \theta_i$ is the phase difference between v and i . The factor $\cos \theta$ is called the power factor, and θ is the power factor angle. P_a is called real power, which implies that the power has transformed from electrical to nonelectrical form such as radiant, chemical, or mechanical. If the load is purely resistive, the voltage and current signals are in phase, and the average power becomes $P_a = |V| |I|$. The average power associated with purely inductive or capacitive circuits is zero, since power delivered to the inductor or capacitor during part of the cycle is returned during another part of the cycle. The power factor is zero, since the phase angle is 90° . In this case, the power is purely reactive. The reactive power (Q) is defined as

$$Q = |V| |I| \sin \theta \quad (6)$$

where the factor $\sin \theta$ is called the reactive factor. Reactive power is expressed in vars (volt-amperes reactive).

Energy is defined as the total work done over an interval of time (T), and the basic unit of energy is called the joule or watt-second. In many instances, this unit is inconvenient to use and

larger units such as kilowatt-hour are used. Using Eq. (2), the dc energy can be expressed as

$$E_d = I^2 RT \quad (7)$$

Using Eq. (5), the total energy delivered during time interval T can be expressed as

$$E_a = |V| |I| T \cos \theta \quad (8)$$

Similarly, using Eq. (6), the reactive energy can be expressed as

$$E_r = |V| |I| T \sin \theta \quad (9)$$

The generation, transmission, distribution, and utilization of large amount of electric energy are accomplished by means of three-phase circuits. A comprehensive analysis of three-phase circuits is beyond the scope of this article; a general understanding of balanced three-phase circuits is sufficient for our purpose. The basic structure of a three-phase system consists of voltage sources connected to the loads by means of transformers and transmission lines. Figure 2 shows the simplified block diagram of a basic three-phase circuit. A set of balanced three-phase voltages consists of three sinusoidal voltages that have identical amplitudes and frequency but are out of phase with each other by exactly 120° . In three-phase circuits the standard practice is to refer to the three phases as A, B, and C. Furthermore, phase A is almost always used as the reference phase. Because the phase voltages are out of phase by 120° the following two possible relationships can exist between the voltages:

First Possibility. $V_A = V_m \angle 0^\circ$, $V_B = V_m \angle -120^\circ$, $V_C = V_m \angle +120^\circ$

Second Possibility. $V_A = V_m \angle 0^\circ$, $V_B = V_m \angle +120^\circ$, $V_C = V_m \angle -120^\circ$

Here V_m represents the peak amplitude of the sinusoidal voltage. The energy delivered by a three-phase source and consumed by a three-phase load is found simply by adding the energies in the three phases. In a balanced three-phase circuit, however, this is the same as multiplying the average energy in any one phase by 3, since the average energy is same in all phases. Thus for a balanced three-phase circuit the active energy or energy may be expressed as

$$E_a = 3V_p I_p T \cos \theta_p \quad (10)$$

where V_p represents the magnitude of the phase voltage, I_p represents the magnitude of the phase current, and θ_p represents

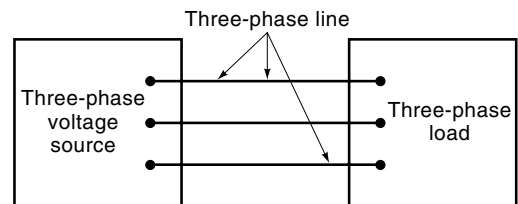


Figure 2. A basic three-phase circuit.

the phase difference between V_p and I_p . Similarly, the total reactive energy can be expressed as

$$E_r = 3V_p I_p T \sin \theta_p \quad (11)$$

In the case of energy the measuring instrument is called a watt-hour meter, while in the case of reactive energy the measuring instrument is called a reactive volt-ampere-hour meter.

WATT-HOUR METERS

A meter that records energy in watt-hours or kilowatt-hours is called a watt-hour meter. An important requirement of an energy meter is that it should indicate a given amount of energy proportional to power and time. For example, it should

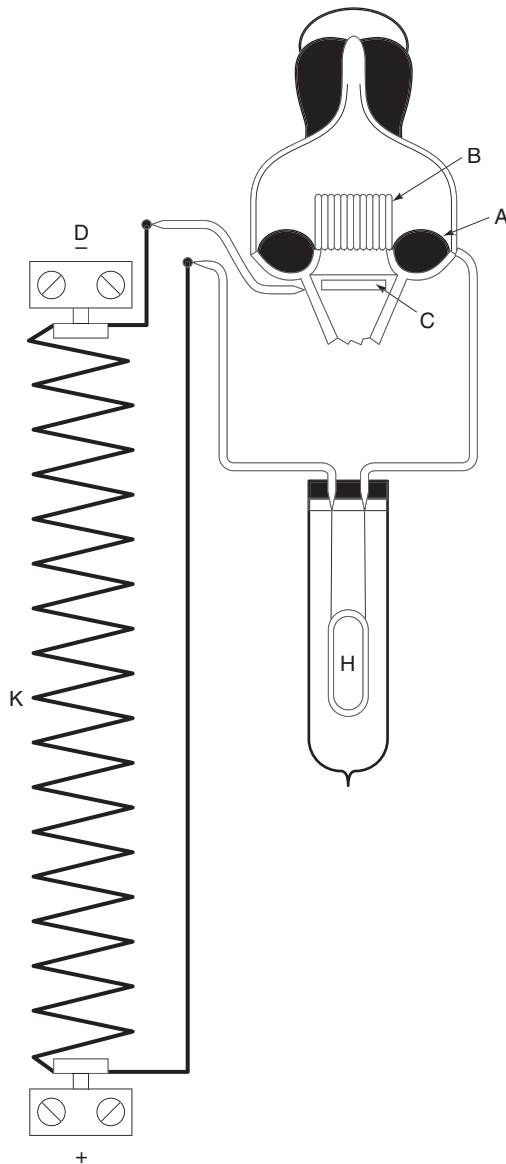


Figure 3. Electrolytic watt-hour meter. A, anode mercury; B, glass fence; C, cathode; D, negative terminal; E, positive terminal; K, shunt; H, compensating resistance in series with tube.

record 1 kW · h, whether this consists of 1 W flowing for 1000 h or 1000 W flowing for 1 h. Different types of watt-hour meters have been developed in the last few decades for the measurement of energy. We shall briefly discuss only those that are relevant today.

Electrolytic Watt-Hour Meters

Figure 3 shows an electrolytic watt-hour meter. This type of meter is mainly used for dc energy measurement, although it can be adapted, by using a metal rectifier circuit and a current transformer, to function as an ac circuit for measuring kilovolt-ampere-hours. The operating current is passed through a solution, causing electrolytic action. Depending on the meter type, this gives a deposit of mercury or liberates gas proportional to the number of coulombs or ampere-hours passed through the meter. Assuming the voltage supply to the meter remains constant, the meter can be calibrated in kilowatt-hours; otherwise it is calibrated in ampere-hours. These meters are inexpensive to manufacture, but tend to require fairly frequent inspection, as they include a large amount of glass in their construction.

Clock Watt-Hour Meters

Figure 4 shows a clock watt-hour meter. There are two identical circular coils (C1 and C2), placed at the bottom ends of two pendulums, which are continuously driven by clockwork. These coils are connected in series with one another and with a high resistance, and they carry a current proportional to the line voltage. There are two current coils (C3 and C4), placed beneath the pendulums, which are connected in series with the line and are wound in such a way that their magnetic fields are in the opposite directions. When no current is flowing, the pendulums swing at the same rate. However, when current flows through C3 and C4, one of these coils exerts an accelerating force on one pendulum and the other coil exerts a retarding force on the other pendulum. The resulting difference in the time period of oscillation of two pendulums is arranged to give an indication on the dial register proportional to the energy passing through the meter. This meter is comparatively free from temperature errors and stray fields and is suitable for both ac and dc energy measurements.

Motor Watt-Hour Meters

Motor watt-hour meters can be broadly divided into two categories: those for dc and for ac energy measurement. The latter—

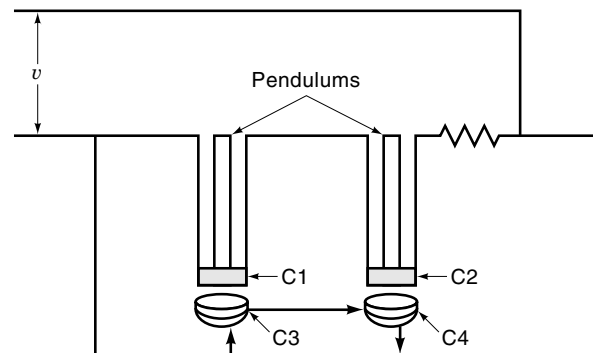


Figure 4. Clock watt-hour meter.

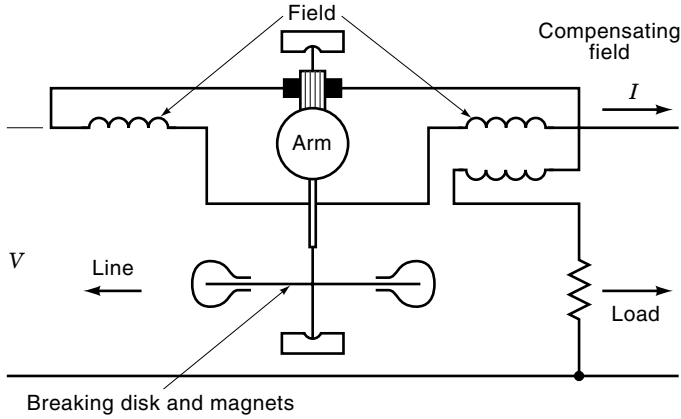


Figure 5. Motor watt-hour meter for dc energy measurement.

also called induction watt-hour meters—can be further classified into single-phase and polyphase watt-hour meters.

Motor Watt-Hour Meter for dc Energy Measurement. A motor meter for dc energy measurement essentially consists of a small motor that is provided with a magnetic braking mechanism as shown in Fig. 5. The field coils of this meter consist of a few turns of heavy copper wire carrying the current under measurement, so that the field strength is directly proportional to the load current. Motor meters for dc energy measurement consist of three main parts—a rotating element, a braking system, and a clock or dial register. The rotating element is driven at a speed proportional to the energy or in some cases the quantity of electricity passing through the driving system. Proportionality between the energy and speed is ensured by the braking system, which supplies a controlling action proportional to the speed of the rotor element.

Single-Phase Induction Watt-Hour Meter. Induction meters are almost universally used for ac energy measurement, since they are simple in construction, provide high torque-to-weight ratio, and are relatively inexpensive. An induction watt-hour meter consists of an induction motor whose output is largely absorbed by the braking disk and dissipated as heat as shown in Fig. 6. In this watt-hour meter there are two current poles (2 and 4), which are displaced from the voltage pole (3) as shown in Fig. 6(a). At unity power factor, the flux ϕ_i from the current coils is in phase both with the voltage v and current i . The flux from the voltage coil, ϕ_v , is in quadrature (90° phase

displacement) with respect to ϕ_i , as shown in Fig. 6(b). At time instant a , both the current i and current flux ϕ_i are maximum, the voltage v is maximum, and the voltage flux ϕ_v is minimum. The flux paths through the disk are from 2 to 1, from 2 to 3, from 3 to 4, and from 5 to 4. At time instant b , both the current i and current flux ϕ_i are minimum, but the voltage v is minimum and the voltage flux ϕ_v is maximum. At time instant c , both the current i and current flux ϕ_i are maximum, but the voltage v is maximum and the voltage flux ϕ_v is minimum. The flux paths through the disk are from 1 to 2, from 3 to 2, from 4 to 3, and from 4 to 5. Thus the flux has effectively moved across the disk from left to right. This change causes eddy currents to be set up in the disk. The reaction between the eddy currents and the field tends to move the disk in the direction of the field.

Since the disk revolves continuously when on load, electromagnetic forces (emfs) will be induced in it dynamically, as it cuts through the flux between the poles, in addition to the statically induced emfs due to the alternating flux in these poles. The torque due to the dynamically induced eddy currents in the disk will be negligible compared to the operating torque produced by the statically induced currents. Neglecting the effect of friction in the meter, and assuming that the active flux from the voltage pole lags 90° behind the impressed voltage, the operating or driving torque T_d becomes proportional to the power in the circuit, i.e.,

$$T_d \propto |V| |I| \cos \theta \quad (12)$$

Now the retarding torque T_r due to eddy currents is proportional to the speed of revolution, N , of the disk, i.e., $T_r \propto N$. Since for a steady speed of the disk, T_d must be equal to T_r , we can write

$$N \propto |V| |I| \cos \theta \quad (13)$$

i.e., the speed of revolution of the disk is proportional to the power. The total number of revolutions, N_T , over a time interval T may be expressed as

$$N_T = NT = |V| |I| T \cos \theta \quad (14)$$

From Eq. (14), it is evident that the total number of revolutions is proportional to the total energy supplied.

Polyphase Induction Watt-Hour Meter. Polyphase energy can be measured by several single-phase instruments, since the

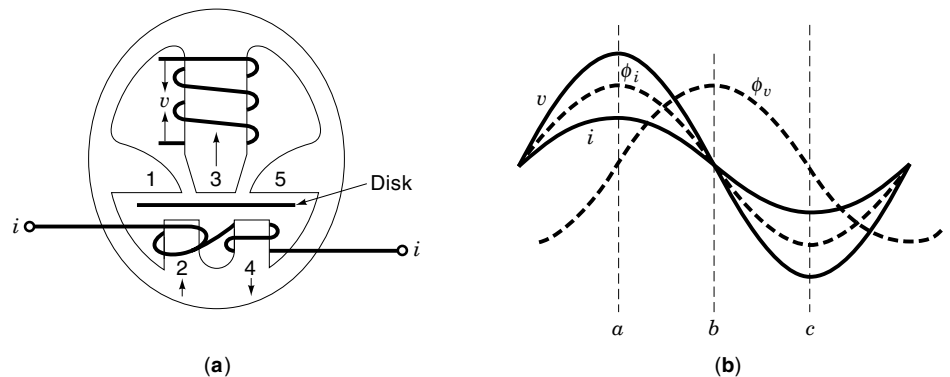


Figure 6. Induction watt-hour meter for ac energy measurement.

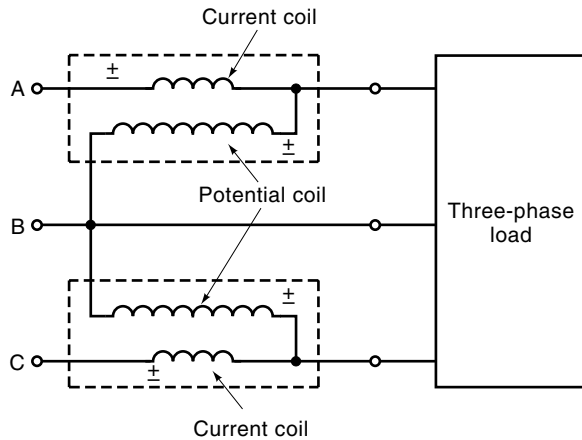


Figure 7. Connection diagram for polyphase watt-hour meter method of energy measurement.

total power or energy is the sum of the readings of all the instruments. The connections to the polyphase meters are the same as those employed in the two wattmeter method of measuring three-phase power, as shown in Fig. 7. Most polyphase instruments, however, utilize two or more single-phase elements mounted in a single shaft, which drives the dial. These elements must be shielded from each other to avoid interaction between the fluxes produced by individual elements. For instance, three-phase energy can be measured by having three watt-hour meters with current coils in each line and potential coils connected across the given line and any common junction. Since the common junction is completely arbitrary, it may be placed in any of the three lines. When this is done, the watt-hour meter connected in that line will indicate zero energy, because its potential coil has no voltage across it. Therefore, the third watt-hour meter can be dispensed with, and the three-phase energy can be measured with only two watt-hour meters. In general, m -phase energy can be measured by $m - 1$ watt-hour meters.

Polyphase Watt-Hour Meter Connections. It is crucial to ensure appropriate connection of the various components of a polyphase watt-hour meter for registering accurate results. In a three-phase circuit, when the power factor is 1 corresponding to resistive loads, both watt-hour meters as shown in Fig. 7 will indicate the same reading; that is, the disk rotation will be

in the forward direction at the same speed. If the power factor is above 0.5, the disk rotation will always be forward whenever the current or the potential coil of either of the two watt-hour meters is disconnected. On the other hand, if the power factor is less than 0.5, the disk rotation of the two watt-hour meters will be in the opposite directions. It should be mentioned that if the current coils of the two watt-hour meters of Fig. 7 are interchanged and the power factor is 0.5, the meter reading will be 100% wrong although it will run at normal speed.

Induction Watt-Hour Meter Adjustments. All adjustments of induction watt-hour meter are usually made either at full load or at light load. To adjust the speed at full load, the drag magnet is usually shifted relative to the disk axis. Alternatively, speed adjustments can be made by using a movable soft iron bar for shunting the flux. To compensate for variations in ambient temperature, a small piece of temperature-sensitive iron-nickel alloy is generally used as a shunt in the air gap of the drag magnets.

MICROPROCESSOR-BASED ENERGY METERS

Microprocessor-based systems are increasingly being used in energy measurement. Figure 8 shows the simplified block diagram of a microprocessor-based energy measurement system employing synchronous counting (3). The system consists of an analog unit and a digital unit. The input signals v and i are multiplied in an analog multiplier. An analog integrator, such as a dual slope converter, performs integration on the multiplier output signal over a time interval T_i . The microprocessor sends a start pulse of suitable duration to activate the synchronization circuit. The synchronization circuit, after synchronization with the input signal v , generates a signal of duration T_i , which controls the time counter gating logic. Notice that T_i is the period of integration for the multiplier output signal and is always equal to an integer multiple of the input signal period. Subsequently, the analog integrator integrates a reference signal over a period T_j , which is measured by the time counter. The microprocessor keeps track of the entire duration (T) of measurement using an internal counter. The outputs of the event and time counters are fed into the microprocessor system, which internally calculates the total energy and outputs the calculated or measured energy on the display. With microprocessor-based energy meters, an accuracy of the order of 0.1% can be achieved.

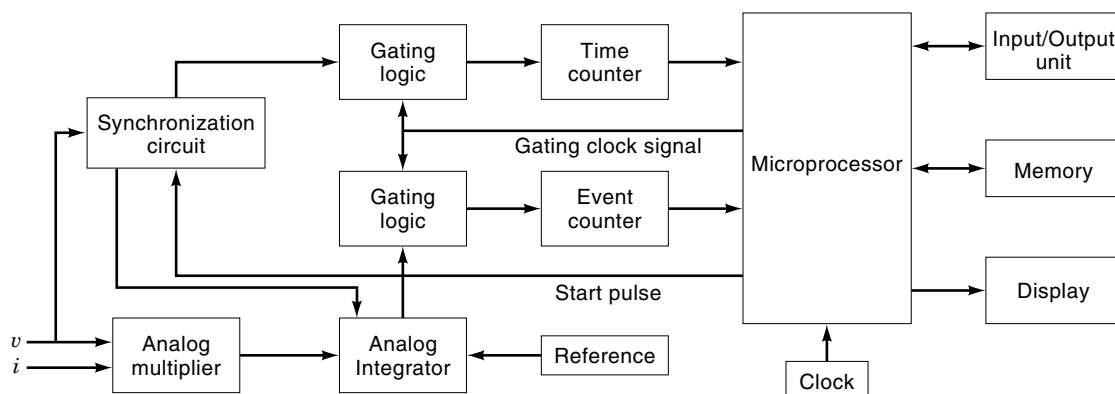


Figure 8. Simplified block diagram of a microprocessor-based energy meter.

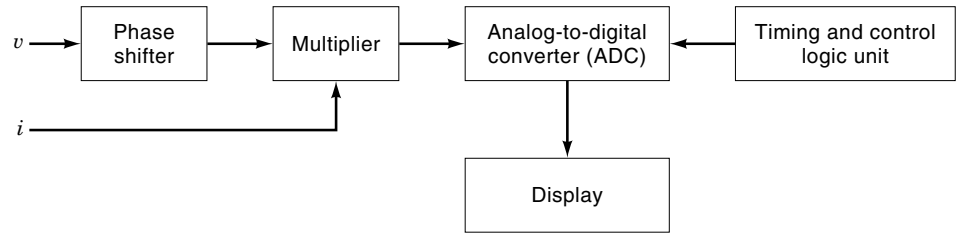


Figure 9. Simplified block diagram of an electronic reactive energy measurement system.

REACTIVE VOLT-AMPERE-HOUR METERS

Reactive energy is defined as the quantity measured by a perfect watt-hour meter that carries the current of a single-phase circuit and a voltage equal in magnitude to the voltage across the single phase circuit but in quadrature therewith (4). Reactive energy is often measured at single-phase and three-phase services. Reactive energy measurement is generally more complex than active energy measurement because additional processing of input signals is necessary. The most common reactive energy meters are watt-hour meters with the currents through the voltage coils displaced by 90° . To provide this 90° phase shift various methods such as potential cross phasing, *RC* and *RLC* circuits, varformers, digital phase shifting in sampling meters, and electronic integrators in time-division multiplier meters are used (5). If the voltage across the single-phase circuit is given by

$$v = |V| \sin \theta_v \quad (15)$$

then the voltage in quadrature is given by

$$v = |V| \sin(\theta_v - 90 \text{ deg}) \quad (16)$$

From the point of view of measuring system design, it is irrelevant whether the phase shift is performed on the voltage or on the current signal. Only the phase displacement between the input signals is of importance. Figure 9 shows the simplified block diagram of a reactive power and energy measurement system. At first, the voltage signal is given a 90° phase shift in the phase shifter; it is then multiplied with the current signal in the analog multiplier. The output signal from the multiplier unit, that is, the reactive power signal, is introduced into an analog-to-digital converter (ADC). The integrating unit inherent in any ADC provides the power integration step to generate the reactive energy, which after some intermediate processing is shown on the display unit. It should be mentioned that, for nonsinusoidal signals, the reactive energy of the fundamental frequency is usually the dominant reactive energy component. Therefore, the measurement of that component is most important. For better accuracy, additional higher harmonic terms may be incorporated.

METER RATINGS

The rise in temperature resulting from the losses in a meter must be accounted for when rating a meter because meter life span should not be unduly shortened by overheating and consequent deterioration of insulating materials surrounding the current carrying conductors. In general, a general purpose meter should have a lower permissible temperature than a special

purpose meter to accommodate a greater factor of safety. In the past, watt-hour meters were rated only at the full load which is currently referred to as the test ampere (TA) rating or test current rating. The TA rating is indicated on the nameplate by the manufacturer and is mainly used for test constant calculation as well as for determining and adjusting the percent registration of a meter at light and heavy loads. The percent registration of a watt-hour meter is defined as the ratio of the actual registration to the value of the quantity measured during a given time interval. Currently watt-hour meters are rated into a number of classes based on their maximum capacities. A class such as 10, 20, 100, or 200 usually indicates the peak of the load range in amperes.

RECENT TRENDS

Power and energy measurement using electronic systems, microprocessors, and digital techniques is becoming increasingly popular. Such systems can provide greater versatility and accuracy, and afford the potential to transmit data without the need for a meter reader. Since electronic systems are often smaller and lighter than electromechanical systems, these systems may be more portable as well as less expensive.

Switching devices, such as silicon-controlled rectifiers (SCRs), that draw current from the line during only a portion of the cycle are becoming commonplace. This trend has added to the problems of power companies, since the switching process generates harmonics of the line frequency, which have repercussion upon both customer and the supplier of electrical power.

As mentioned earlier, electronic equipment can be used in versatile ways to aid in measuring energy. The portable tester is one of the latest electronic gadgets to help us with energy measurement. Until recently a meter could only be tested by pulling it and taking it back to the shop. However, portable testers effectively alleviated that problem. They are especially helpful when customers live in remote areas.

ACCURACY

Electric utilities all over the world use single-phase watt-hour meters to measure residential energy consumption. Since the watt-hour meter serves as the cash register for the utility, both the utility and the regulatory body have a high interest in its accuracy. In order to ensure that the customers are not charged for more energy than they consume and the utility collects the revenue to which it is entitled, watt-hour meters are sampled on an annual basis, ensuring accurate registration of energy. Typically the sampled meters are sent to a central testing facility and tests are performed to indicate how accurately the meters are performing in service. The results of these sampled tests are reported to the regulatory body to satisfy its rules and

regulations. The utility also uses statistical tools for determining the performance of the meter population.

Single-phase watt-hour meters typically used in residential service are three-wire, 240 V, and class 200 (6). These meters are generally tested at two test load points: full load (FL), which is defined as 30 A, and light load (LL), which is defined as 3 A. In order to get an indication of how a meter performs over a load range of 0 to 200 A, the test results at full load and light load are weighted to yield an overall accuracy. The overall accuracy (OA) equation is written as

$$OA = \frac{4 \times FL + LL}{5} \quad (17)$$

For example, if the full-load accuracy is 100.2% and the light-load accuracy is 95%, then the overall accuracy is 99.16%. This method is one of the two methods recognized by the utilities. The second is the true average method. In this case the overall accuracy equation is defined as

$$OA = \frac{FL + LL}{2} \quad (18)$$

The second method is not in widespread use.

Energy meters are also classified as *precision grade* or *commercial grade*, depending on the degree of accuracy with which they make measurements. Commercial grade meters are those generally met with in ordinary factory and home service installations. Precision grade meters are designed to have negligible error and are retained exclusively for test-room measurements thus making them expensive.

Because of its low cost, the inductive watt-hour meter is still the most widely used energy meter. It is practically the only meter used in the residential sector. The accuracy of most inductive watt-hour meters is typically 1% for sinusoidal operation at a fixed frequency. Recently, energy meters based either on digital sampling or on time-division multiplier became available, which are very accurate ($\sim 0.1\%$) compared to the inductive watt-hour meters.

CALIBRATION

Phantom loading of watt-hour meters is a common method for their calibration. This mainly consists in supplying sources of calibration voltage that are independent of each other at one point and are adjustable. This practice has several advantages over the use of resistive and reactive loads, such as (1) better adjustment of voltage, current, and power factor in the laboratory, (2) reduced need for large current supplies, and (3) elimination of errors that might result from failure to take into account losses or loading effects in the test and reference instruments. Figure 10 shows two circuits commonly used in the calibration of watt-hour meters. In Fig. 10(a), the current circuits are connected in series and supplied with a stepdown transformer for the test and reference instruments. This isolates the current circuit from the voltage circuit. A smaller series adjusting resistor may also be used for adjusting the load current. This method is used when it is desired to simulate a unity-power-factor load. The voltage of the current supply transformer must be high enough so that it swamps the reactance of the current circuits of the meters. The circuit in Fig. 10(b) is used when low-power-factor loads of the order of 50 are desired. By using

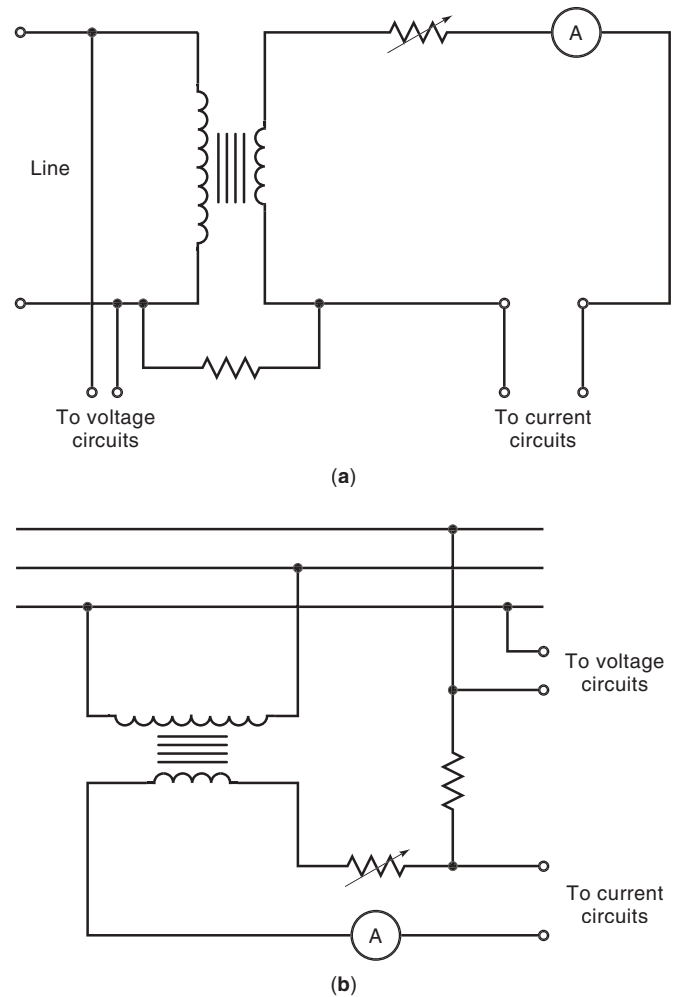


Figure 10. (a) A basic calibration circuit for a watt-hour meter, and (b) a more sophisticated version of the calibration circuit.

a three-phase supply and taking the current supply from one phase and the voltage supply from another, it is possible to simulate low-power-factor loads if the phase angle is known. The phase for the voltage supply can be selected to lead or lag the current of the secondary of the supply transformer by 60° .

PORTABLE METER READING AND STANDARDS

Presently utilities all over the world are debating whether to incorporate *automated meter reading* (AMR) in their systems. AMR offers many other advantages than just remote meter reading. An AMR system can signal outages, perform remote connect and disconnect, deter meter tampering, prevent lockouts, and facilitate time-of-use metering. In addition, utilities can offer value-added services such as forced-entry alarms and low- or high-temperature alerts. Any of these can generate revenue. Over 30 vendors are in the AMR market, offering systems of every size and kind (7). The systems vary according to the communication media used, whether they are one-way or two-way, whether they read meters at high or low speed, and whether they are fixed or mobile. Utilities have been slow to adopt automated meter reading systems because the investment is high and the payback is slow. Other reasons

include reluctance to change, discomfort with new technology, and long-term satisfaction with walkaround readers. Furthermore, AMR system requires a coordinated, horizontal effort across the utility, involving metering, billing, customer service, and information systems.

The most widely used technique for watt-hour meter testing is called portable standard or rotating standard. Portable standards are transportable watt-hour meters with multiple current and voltage range and allows both whole and fractional reading. Initially, energy is measured by using the watt-hour under test. Then the portable standard meter is used to measure the same energy. The accuracy is determined by comparing the readings of the two meters.

REFERENCE STANDARDS

For ac watt-hour meter testing, a time reference standard such as a clock or a stop watch and a wattmeter may be used. Alternatively, a watt-hour meter is used as the reference standard which is started and stopped automatically by a signal (e.g., light) generated from the meter to be tested. For dc watt-hour meter testing, the reference standard may be ammeters and voltmeters or potentiometers.

Standards and technical practices are an essential tool in any field, and energy measurement is no exception. In the USA, approximately half of all standards are generated through a voluntary consensus process (8). Most engineering standards are developed by technical societies and trade associations. The private sector rather than the government generates the standards. The American National Standards Institute (ANSI) coordinates the voluntary development of standards in the USA. ANSI is a nonprofit organization composed of corporate members, trade associations, and technical societies. It is self-designated in its role in standard coordination and does not develop any standards. ANSI reviews the documents produced by others and certifies them as American National Standards. The Institute of Electrical and Electronic Engineers (IEEE) is the primary producer of electrical documents for ANSI. Other major nongovernment standard organizations producing electrical standards are the American Petroleum Institute (API), Associated Edison Illuminating Companies (AEIC), the National Electrical Manufacturers Association (NEMA), and the National Fire Protection Association (NFPA).

CONCLUSION

In this article we started off by defining some basic terms that would allow us to understand different aspects of energy measurement. We then described different types of energy-measuring instruments in detail, and briefly discussed some of the latest trends in energy measurement. All these topics were chosen by keeping in mind the diverse readership of this encyclopedia. With our present dependence on energy it is necessary that a wide section of the population become aware of different aspects of energy measurement.

BIBLIOGRAPHY

1. E. W. Golding, *Electrical Measurements and Measuring Instruments*, 4th ed., London: Pitman, 1959.
2. D. E. Johnson, J. R. Johnson, and J. L. Hilburn, *Electric Circuit Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
3. J. K. Kolanko, Accurate measurement of power, energy and true rms voltage using synchronous counting, *IEEE Trans. Instrum. Meas.*, **42**: 752–754, 1993.
4. B. Djokic, P. Bosnjakovic, and M. Begovic, New method for reactive power and energy measurement, *IEEE Trans. Instrum. Meas.*, **41**: 280–285, 1992.
5. R. Arseneau, G. T. Heydt, and M. J. Kempker, Application of IEEE standard 519-1992 harmonic limits for revenue billing meters, *IEEE Trans. Power Deliv.*, **12**: 346–353, 1997.
6. R. P. Anderson, A method of deriving overall accuracy for single phase watthour meters, *IEEE Trans. Power. Deliv.*, **PWRD-2**: 337–341, 1987.
7. J. Marks, Much more than automated meter reading, *Electrical World*, April 1996, pp. 17–24.
8. R. L. Haynes and F. L. Messec, Codes, standards, and recommended practices for electrical systems, *IEEE Trans. Ind. Appl.*, **30**: 1506–1513, 1994.

Reading List

- A. Braun (ed.), *1996 Conference on Precision Electromagnetic Measurements Digest*, IEEE, 1996.
- E. Frank, *Electrical Measurement Analysis*, New York: McGraw-Hill, 1959.
- W. Kidwell, *Electrical Instruments and Measurements*, New York: McGraw-Hill, 1969.
- W. Alexander, *Electrical Instruments and Measurements*, London: Cleaver Hume Press, 1962.
- J. W. Snider, *Electronics for Physics Experiments*, Menlo Park, CA: Addison-Wesley, 1989.
- J. W. Snider, Metering equipment for interconnects and large loads, Research Report, Canadian Electrical Association, Toronto: SANG-AMO Canada, October 1984.
- F. Castelli, Both active and reactive power and energy transfer standard, *IEEE Trans. Instrum. Meas.*, **41**: 274–279, 1992.
- A. Domijan, Jr., A. Gilani, G. Lamer, and C. Stiles, Watthour meter accuracy under controlled unbalanced harmonic voltage and current conditions, *IEEE Trans. Power. Deliv.*, **11**: 64–70, 1996.

MOHAMMAD S. ALAM
Purdue University

WATTHOUR METERS

A watthour meter measures electric energy consumed by a utility customer in units of kilowatthours (kWh). The meter does this by measuring the integral of the active power with respect to time. This integral represents the energy delivered to the load (electric utility customer) in the interval over which the integration is done, and the units for convenience are in blocks of $1000 \times W$. The watthour is the electric energy used in 1 h when the average power during the hour is 1 W. The watt is the unit of active power and is defined as the rate of change of energy delivery to a load. The watt is the power used when a current of 1 A flows through a 1Ω resistance.

The watthour meter has been called the cash register of the utility industry and certainly from the viewpoint of the customer it has served this one purpose very well for over 100 years (1). In essence, the watthour meter transforms customer dollars to utility dollars and thus has enabled electric

utilities to bill relatively accurately for electric energy use and create (using revenue dollars earned) a multitrillion dollar transmission, distribution, and generation infrastructure worldwide.

The watthour meter measures active power, which represents the real power used by customers in end-use applications, such as heating, cooling, lighting, process control, electric machines, and running computers. Other components of energy at the power frequency level (fundamental 60 Hz to 50 Hz) include apparent and reactive powers. The measurement of these two quantities will not be addressed herein in detail, but certainly all three quantities form the basis of power network flow and define the electric product/service. For example, the reactive power, measured in kilovolt-amperes reactive (kVAR), controls primarily the voltage profile in power systems but does not contribute directly to end use. The apparent power S measured in kilovolt-amperes is all the power available to a customer, and from the ratio of active to apparent power, we may determine the power factor, which is an important factor in adjusting the reactive power. Certain types of watthour meters may measure all three quantities (active, reactive, and apparent) of power and many others such as harmonics.

Another purpose that the watthour meter is increasingly being used for is to provide both the utility and customer with information. The meter may provide not only revenue information, but also real-time demand, energy use patterns, profiling of loads and potential equipment failure information, and further control of equipment and other information technologies (Internet, cable) may be routed through the meter. The next generation of meters may well serve as the utility's information gateway to the customer.

THE DEVELOPMENT OF REVENUE METERING

With the invention of the electric light in 1879, Thomas Edison made the first electric equipment, the incandescent lamp. He knew that there was not only commercial value in lighting but also in heating homes and in transportation. Difficulties had to be overcome, especially because most people were unfamiliar with electrical technology and electricity, as well as its value compared to gas. Certainly to be of commercial value, electricity had to be measured. Edison was well aware that, in order to enjoy commercial success, he had to be able to deliver accurate measurements of electric use. The Mazda Light Company was first formed with Edison to solve the problem of accurate measurements in 1881 with the development of the chemical meter. With respect to a present-day meter, this was a primitive device. Basically, its operation hinged on the use of two electrodes. After a period of time, the electrodes in the chemical bath would interchange mass. Then the electrode would be weighed at the start and end of the billing period. This was also an attempt to make electricity more palatable to users in that the product (electricity) could be weighed much as one would weigh produce in the food market. This meter was used for a few years but became obsolete prior to the introduction of alternating current.

During and well after the time of the chemical meter, several leading people focused on satisfying customers by developing accurate metering. Among these people were Thompson, Shallenberger, Lamphier, and Duncan. The companies that these people worked for are still manufacturing watthour and

other meters. Thompson was the founder of the Fort Wayne Meter Company that eventually became part of General Electric. The work of Lamphier became the genesis of Sangamo Electric Company (today known as Schlumberger). Shallenberger was associated with Westinghouse (now ABB). Lastly, Duncan was the founder of the Duncan Meter Company (now Landis & Gyr).

The first practical wattmeter was developed by Professor Elihu Thompson in 1889. It was the first practical meter to measure true watthours and was awarded the 1889 Grand Prize at the Paris Exhibition. It could be used for alternating current (ac) or direct current (dc) electric power systems and became the industry standard. This was a significant advance over Edison's chemical meter, but it was not up to present-day standards. It is still remarkable for its innovation considering it was developed over 100 years ago. Also, in the same time period, Oliver Shallenberger developed an ampere-hour meter that was used exclusively in ac systems. This meter was adopted by many utilities because of low cost. However, with the increased use of ac, Thompson's meter was the desired choice because it measured watt-hours, which is the appropriate measure of electric energy.

The requirements of electric companies in the early days exceeded the availability of meters. Consequently, much electric energy was sold on a per-lamp cost basis. By 1890, it was recommended that this practice be halted and that in large metropolitan areas meters be used. This development in replacing the flat rate with a metered rate was made possible by the meters developed by Thompson and Shallenberger. The next several years witnessed the introduction of new meters by Duncan and Lamphier, which were primarily based on the pioneering work of Thompson and Shallenberger. In 1894, Shallenberger achieved one of the most important discoveries made in the metering field. He applied the induction meter, which would only previously measure ampere-hours, to measure watt-hours. He accomplished this via the method known as the lagging process, which continues to be at the crux of the metering field.

The first modern meter was introduced in early 1903 and is known as the Type 1 induction watthour meter. It was characterized by its compactness, high torque, ease of adjustment, large four-dial register, and long life. In 1913, the combination lag and light load plate was introduced. By 1925, F. Kinnard developed temperature-compensated magnets. Also, at this time, it was deemed necessary to respond to the rapidly increased use of electricity to form a mechanism for educating people in the utility industry on the proper maintenance and techniques associated with watthour meters. This organization was chartered with the University of Florida as the Southeastern Metermen's Association. This organization is still going strong and is now known as the Southeastern Electricity Metering Association (1). The middle 1920s saw a boom in the use of electricity and, as a result, saw the introduction of longer-range accuracy and increased current-carrying capacity in meters. Standardization of mounting dimensions was formed in 1934. In particular, the standardization of the "S" type meter and an agreement that manufacturers would produce two standard types of single-phase meters. One was the type "A," and the other was known as the type "S" meter. By 1938, the change to installing meters outdoors was the result of better materials being used in construction resulted in new problems. The main problem was in decalibration of the retarding magnets caused by current surges (due to switching and lightning) causing the meter

to speed up. This problem was rectified by the introduction of harder magnets, which were able to better resist knockdown under surge conditions, such as Alnico II. By 1948, the General Electric Company introduced the I-50 meter, which had magnetic suspension. This eliminated the need for replacement of the expensive jewel bearings in meters and increased calibration stability. The continued growth of residential loads resulted in two further improvements, one in 1955 extended the meters range from 60 A to 100 A, and the other improvement in 1957 increased the range to 200 A. Today we see up to class 320 meters used for 400 A service and class 600 meters (bolt-in type) are available.

Up to the 1970s, the many improvements in metering were primarily the result of the dominant manufacturers. After this time, innovative new firms, such as Scientific Columbus, Trans-Data, and Process Systems, among others, have produced novel breakthroughs in metering. The use of electronic and solid-state technologies has been on a steady increase. 1975 witnessed a major event in the introduction of the JEM-1 and solid-state watthour standard (SC-10) from Scientific Columbus. By 1982, the first commercially available solid-state register was produced. Today we see a revolution in the way meters may be applied to not only provide billing data but also to serve customers as an information gateway in energy use and communications.

MATHEMATICAL DEFINITIONS USED FOR THE MEASUREMENT OF POWER

The basic mathematical expressions recommended to be used in metering under single-phase sinusoidal conditions are given herein and are based on Ref. 2. In particular, it should be noted that Budeanu's theory is the approach most familiar to those in industry and the one taught in universities, and it is certainly the method most commonly used in instrumentation. Unfortunately, Budeanu's theory is not adequate for nonsinusoidal situations (3,4). Instead, the recommendations of IEEE Working Group on Nonsinusoidal Situations, chaired by A. E. Emanuel, should be used (2). Power theory under nonsinusoidal situations will not be developed herein; rather the sinusoidally valid approach used historically will be developed. Let us consider the single-phase case assuming a sinusoidal voltage and current.

For voltage, we have

$$v = \sqrt{2} V \sin(\omega t)$$

which, if applied to a linear load, will result in a current of the form

$$i = \sqrt{2} I \sin(\omega t - \theta)$$

where

I = root-mean-square (rms) value of current

V = rms value of voltage

$\omega = 2\pi f$

f = frequency in hertz

θ = phase angle

The instantaneous power in watts is

$$p = vi$$

p is composed of two components, an active and reactive component.

$$p = p_a + p_q$$

where

$$p_a = VI \cos \theta [1 - \cos(2\omega t)] = P[1 - \cos(2\omega t)]$$

and

$$p_q = VI \sin \theta \sin(2\omega t) = Q \sin(2\omega t)$$

The power p_a , which is an instantaneous quantity, is a result of the active current component, and represents the rate of change of energy

$$w_a = \int p_a dt = Pt - (p/2\omega) \sin(2\omega t)$$

w_a flows in one direction from the source to the load, and its rate of change is positive. On the other hand, the power p_q is a result of the reactive part of the current, and is the rate of change of energy

$$w_q = \int p_q dt = (Q/2\omega)[1 - \cos(2\omega t)]$$

w_q is an oscillatory quantity between the source and load and contributes no net energy to the load.

The active power in units of watts is

$$P = (1/kT) \int_{\tau} p dt = VI \cos \theta$$

where

$T = 1/f$, the cycle in seconds

k = an integer number

τ = the time when the measurement begins

The reactive power in units of vars is

$$Q = (\omega/kT) \int_{\tau} i \left[\int v dt \right] dt = VI \sin \theta$$

The quantity Q is the magnitude of the oscillating power p_q . With an inductive load, Q is positive, and with a capacitive load, Q is negative.

The apparent power in units of volt-amperes is

$$S = VI = \sqrt{P^2 + Q^2}$$

In addition, the power factor, which is a unitless quantity, is

$$\text{pf} = P/S$$

COMPONENT FEATURES AND OPERATING PRINCIPLES

Types of Meters

Watthour meters record energy and measure it in kilowatt-hours. The meters come in many types.

1. *Demand meters* record the demand and/or maximum demand and are used to ensure equitable charges to customers. Demand is the rate at which energy is consumed and is measured in kilowatts. The maximum demand is the greatest demand seen during a period of monthly billing. The demand interval is the period of time upon which the maximum demand is determined in the register and is usually taken over 1 min to 60 min intervals. Demand meters usually are three-phase and transformer rated for commercial loads, but they may also be self-contained. Two types of demand meters are (a) integrating demand meters and (b) thermal demand or lagged meters. The integrating demand meters record an average power during consecutive demand intervals. The watt-hour meter is the driver behind the register, which in turn rotates proportionally to the amount of watt-hours in the interval. This meter would show the maximum demand during a certain time period. On the other hand, the thermal demand meter has a pointer that changes in accordance with the temperature variation in its elements due to the current flow. Because the meter responds to temperature changes caused by current flow, it responds more slowly than a normal meter. The main difference between these two meters is with respect to the demand interval. One follows a heating response curve, and the other follows regular intervals. Both types of meters give comparable results with the exception of certain loads, which have large peaks during a short time period (in this case, the thermal meter would be slow to respond). On the other hand, thermal meters operate on a heating curve similar to equipment on the line and thus are preferred by some utilities. Many utilities also use fully solid-state meters with digital displays to measure demand.
2. *Electronic meters* are solid-state devices and have no moving parts. In these meters, the voltage and current waveforms are sampled at specified intervals and digitized. The signals are then processed using appropriate power definitions.
3. *Hybrid mechanical or hybrid electronic meters* may have a memory function and use electronic displays and have a mechanical rotor.
4. *Load profile meters* are an electronic or hybrid meter with an on-board memory so that the energy used for a certain time interval may be calculated and then downloaded to a computer for further analysis.
5. *Mechanical meters* are in large part used in residential applications and are primarily single-phase and self-contained.
6. *Primary meters* are put in service on the primary side of distribution transformers, usually 12 kV. In this case, the customer would own the transformer and its associated losses. Also, primary meters are used when a customer

has multiple points of service and it may be best to use a single meter rather than several.

7. *Pulse meters* are either mechanical or electronic and are used mostly by customers with energy management systems. In this type of meter, a pulse is set to be equivalent to a certain number of watt-hours, and the number of pulses in an interval will be the same as the average demand in the interval. A solid-state data recorder may be used to store and collect the pulses and has the ability of storing data to be read by computer via phone lines or handheld devices.
8. *Secondary meters* are located on the customers or secondary side of distribution transformers. In this case, the utility owns the transformer.
9. *Self-contained meters* are designed so that all the loads current flows through the meter. Depending on the size of the load, a self-contained meter may either be (a) plain meter or (b) demand meter.
10. *Time-of-use meters* record and store energy and demands during on-peak, off-peak, and other times as needed.
11. *Transformer-rated meters* record a fraction of the energy going to the load. Meters are not capable of directly being used on voltages above 480 V and currents greater than 400 A. Potential transformers (PT) and Current transformers (CT) are used in conjunction with the meter so that the meter is capable of measuring large loads (5).

Meter Form Types

The classification of meters into basic form types is necessary for meter selection and to use it in the correct manner with respect of current circuits and external wires. Table 1 (see also ANSI C12.10) illustrates many of the form types.

Andre E. Blondel's theorem says that "If energy is supplied to any system of conductors through N wires, the total power (energy) in the system is given by the algebraic sum of readings of N wattmeters (watthour meters), so arranged that each of the N wires contains one current coil, the corresponding potential coil being connected between that wire and some common point. If the common point is on one of the N wires (ground may also be included in N), the measurement may be made using $N - 1$ wattmeters (watthour meters)" (1). What this means is that energy may be metered using one less current and potential element than the wires in the system. Hence, we would use three stator meters for a three-phase four-wire system. Four kinds of watt-hour meters used in the United States typically apply to Blondel's theorem; these are forms 5, 9, 12, and 16. However, design modifications resulting from economics are also typically used by utilities. These include, for example, the $2\frac{1}{2}$ stator forms 6, 8, 14, and 15. The accuracy of these compromised design forms is contingent upon keeping the phase displacement between line voltages fixed. Figure 1 shows a simplified example of using a form 16s meter in a three-phase four-wire application.

Meter Components and Operation

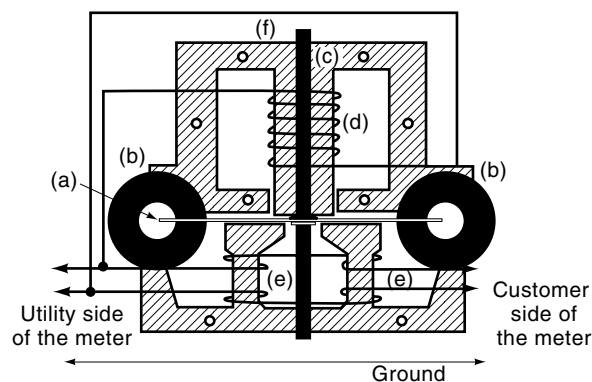
A single-phase mechanical (induction) watt-hour meter is in essence a small induction motor that is used to measure kilowatt-hours. Figure 2 illustrates a three-wire single-phase induction watthour meter. The induction disk rotates because

Table 1. American National Standard for Watthour Meters: Basic Meter Form Types

Form	SC/TR	No. Stators	No. Current Circuits	No. External Wires
1s	sc	1	1	2
2s	sc	1	2	3
3s	tr	1	1	2
4s	tr	1	2	3
5s	tr	2	2	3 (or 4)
6s	tr	2	3	4 wye
7s	tr	2	3	4 wye
8s	tr	2	3	4 delta
9s	tr	3	3	4 wye
10s	tr	3	3	4 wye (alt)
11s	tr	3	3	4 delta
12s	sc	2	2	3
13s	sc	2	2	3 (or 4)
14s	sc	2	3	4 wye
15s	sc	2	3	4 delta
16s	sc	3	3	4 wye
17s	sc	3	3	4 delta
35s	tr	2	2	3
45s	tr	2	2	3 (or 4)
1a	sc	1	1	2
2a	sc	1	2	3
3a	tr	1	1	2
4a	tr	1	2	3
5a	tr	2	2	3 (or 4)
6a	tr	2	3	4 wye
8a	tr	2	3	4 delta
9a	tr	3	3	4 wye

Note: tr = transformer rated, sc = self-contained, s = socket base, a = bottom connected.

of the torque developed by the magnetic field generated by currents flowing through the meter elements. The stator (stationary frame), which becomes an electromagnet when energized, has two coils. The coil in series with the line produces a magnetic field in proportion to the line current. The other coil, which is connected across the line, produces a magnetic field proportional to the voltage. Because the two coils are on the same stator frame, they develop a magnetic field that produces a torque on the disk which is proportional to the load power. Further, there exists a permanent magnetic field developed by the braking magnets; it creates a damping torque in proportion to the disk speed. The rate the disk turns now becomes proportional

**Figure 2.** Main Components of a single-phase, three-wire, mechanical induction watt-hour meter. (a) Rotating disk; (b) braking magnets; (c) stator frame; (d) voltage coil; (e) current coils; (f) shaft.

to the measured power. The number of revolutions during a specified time period is indicative of the power, and the power may now be displayed on the register through appropriate gearing between the register and the disk. A revolution made by the disk represents the watthours and is specified by the meters Kh factor, where $Kh = (\text{rated current} \times \text{rated voltage}) / (\text{rotations per minute} / 60)$.

The main components of the induction watt-hour meter consist of (a) base/cover, (b) disk and spindle, (c) damping magnet, (d) electromagnet or stator, and (e) register. The cover is usually constructed of glass or plastic and is sealed to protect the meter from the elements and tampering. The disk is constructed of aluminum, fitting between the magnetic poles of the current and potential coils, and is mounted on a shaft. Eddy currents on the disk, induced by the potential and current coils magnetic fields, in turn produce a flux that interacts with the two coils that then cause the disk to rotate. The disk drives two elements: (a) the gears and dial pointers by which the revolutions are added to kilowatts of energy used and (b) the shafting and gears in combination with a timing motor that add the revolutions in each demand interval to determine the kilowatt demand. The damping magnet produces a braking force on the disk, which is placed between the poles of the magnet, so that the speed of the disk becomes proportional to the power used by the load. The stator serves to direct the magnetic fields developed by the current and potential coils. The register is constructed of dials

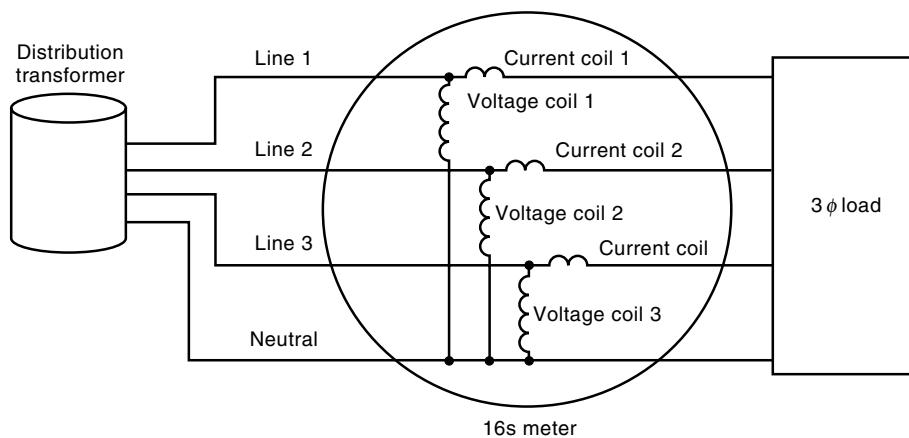
**Figure 1.** Connection of a form 16s polyphase watthour meter in a three-phase four-wire power distribution system. (Courtesy of RJC Associates/CTGI, Southeastern Electricity Metering Association Course 140 Class Notes, 1996)



Figure 3. State-of-the-art solid-state polyphase revenue meter—the ABB Alpha meter. (ABB Power T&D Company, Electric Metering Systems, Raleigh, NC)

that record the amount of energy. The dials are driven by the rotation of the disk through a set of gears.

Metering is made more attractive if the register is electronic. Electronic demand registers were made possible with the introduction of nonvolatile memory. Mechanical registers use an externally operated recorder and translator to do rolling demand calculations. Electronic registers can determine rolling demand internally. The main part of an electronic register is a microprocessor. If an electronic register is used with a mechanical meter, the disk rotation is typically sensed by optical devices. These optical sensors create pulses for each rotation of the disk or by detection of disk shaft motion. Hence, a pulse represents some energy used. The electronic register may perform tasks to establish time intervals, count and sum pulses for display information, and count and sum pulses for demand intervals.

A state-of-the-art solid-state meter, typical of those produced by leading manufacturers, is shown in Fig. 3. In general, solid-state meters give tremendous flexibility by providing sound economic benefits and useful diagnostic information. The ABB Alpha meter is an integral meter and register. With this meter, it is possible to gather, process, and store energy use and demand information on a four rate time-of-use (TOU) or demand basis. The TOU and demand data can be shown in watthours and either apparent energy (VAh) or reactive energy (VARh).

The basic design, as shown in Fig. 4 with the components indicated, is compact. The chassis assembly houses the base, current and voltage blades, connecting cables to the circuit board, and lightning arrester. The electronic housing contains the meter and register electronics on a single circuit board. The circuit board includes the voltage range power supply and dividing resistors. The housing also includes the LCD, nameplate, and battery. The circuit board can also be provided with (optional) relaying outputs. The cover provides for visual metering and an optical port and demand reset. Further, with a wide voltage range capability, it becomes possible to reduce the need (and reduce inventory requirements) for the many form types shown in Table 1. The Alpha meter can meter both form 8s and 9s transformer-rated applications, and the self-contained version can meter form 14s, 15s, and 16s applications. Of the 18 typical types of mechanical meters used today, the solid-state meter shown in Fig. 3 consolidates the form number down to five—reducing inventories up to 50%.

THE FUTURE OF METERING

The future of revenue metering may be considered to be on our doorstep in that many of the components are here today but remain to be integrated into an acceptable system and developed with additional capabilities. For example, the ABB PowerPlus Alpha meter is capable of multiple functions, in addition to serving as a watt-hour meter, such as: voltmeter, ammeter, distortion indicator, power quality monitor, VAR meter, watt meter, phase rotation indicator, VA meter, phase angle meter, and circuit wiring checker. This increase in functionality opens many new applications branches up for the utility. The Alpha meter may be used for diagnostic testing, which involves determining the validity of the electrical service. The meter may be used as a power quality monitor. Such capabilities allow monitoring of voltage and current total harmonic distortion, time stamping of power quality disturbances, and recognition of momentary voltage sags. With regard to revenue metering, these meters record energy and demand values for both real and reactive, and real and apparent quantities. Further, load profile and event logs are available. The communications features included on the main circuit board allow for automatic meter reading and control. Depending on the pass code authority, the users (utility and customer) may access billing records, power quality events, and energy use and load

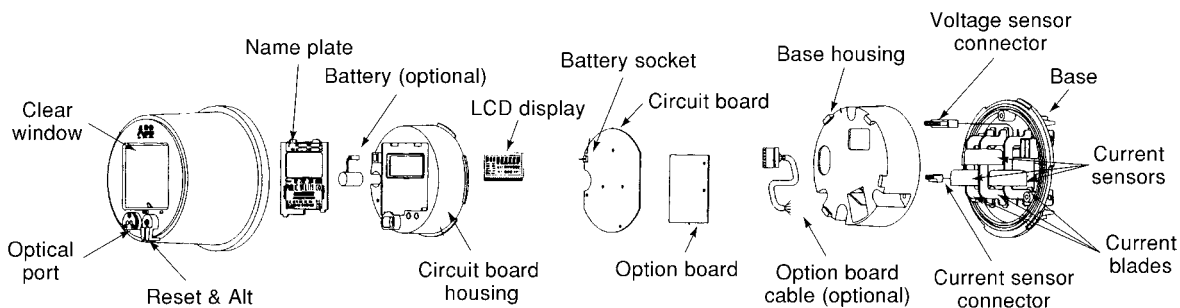


Figure 4. Main components of a solid-state polyphase revenue meter. (ABB Power T&D Company, Electric Metering Systems, Raleigh, NC)

profile information remotely via a telecommunications system to a personal computer. The telecommunications system may be telephone modem, radio, power line carrier, cable, or even satellite.

What the development of new telecommunications and metering infrastructure means is that it will become possible to market new services. This is of vital concern to utilities as competition and deregulation unfold. The telephone modem interfaces, with each meter being able to hold multiple encrypted passwords, will permit direct customer use metered data and support credit card dialing so that billing can be done to appropriate accounts. Further, the meters can also be programmed to follow a certain call in schedule, or call automatically if there are errors occurring. Several novel areas of future application are discussed in the following material. All these areas of meter applications are actively being studied by researchers and manufacturers.

Automation of the Home or the Smart House Concept

Revenue meters, such as the Alpha meter, which are designed for compatibility with protocols to access wide and local area networks like CEBus, permit great flexibility in load control. This then permits information to be transferred to home heating, ventilating, and air conditioning (HVAC) controllers and receivers. These elements then function as relays to the customers appliances. An entire new opportunity for sharing information with many home electronics-based devices ranging from pool pumps to televisions becomes possible."

Real Time Pricing

The utilities via the interactive network can communicate directly with home (and industry) electric systems. The customer would determine criteria for energy usage, and the utility would then via the telecommunications network download real-time pricing for the customer-specified energy usage. With a system such as this in place, it becomes possible to price-control appliances automatically to a customer's desired comfort and economic specifications.

Automatic Meter Reading

Automatic meter reading (AMR) technologies have been in existence for several years. The AMR techniques have typically been done by means of radio or power line carrier schemes. Novel methods under study now include the use of low earth-orbiting satellites. Two-way communication to meters in remote locations worldwide becomes feasible with the addition of communication elements installed in the meter. In general, the major manufacturers are intending to support several communications systems to enhance the application use of AMR. The main driving force behind AMR is the cost (when compared to standard ways of reading meters) and the need for having several items metered, monitored, and controlled. With the advent of retail and wholesale wheeling in a competitive electric utility environment, AMR becomes more of a necessity than an option.

Challenges in Metering

We have come full circle from the early days of metering, described at the start of this article, in that a new product is being offered. From the past to the present, the product has been electricity. We must look beyond electricity metering to combined service offerings of such items as television, energy management or load control of appliances, security, and computing. Figure 5 illustrates this concept. The communications gateway using AMR may be telephone, radio, power line carrier, cable (as illustrated), or satellite. The challenges in making this complex system work are great not only because there are hundreds of millions of customers potentially involved but also because of the limitations of communications bandwidth, measurement technology, and conflicting standards. For example, it may be implied from Fig. 5 that an explosive growth in the use of electrotechnologies is expected to occur. In large part, these electrotechnologies are power electronics based. Power electronics (PE) are incorporated in such elements as variable speed motor drives (used in HVAC and many other applications), PCs, high-efficiency lighting ballasts, process controls, and a full range of appliances. The proliferation of PE-based devices means a consequent increase of power system harmonic

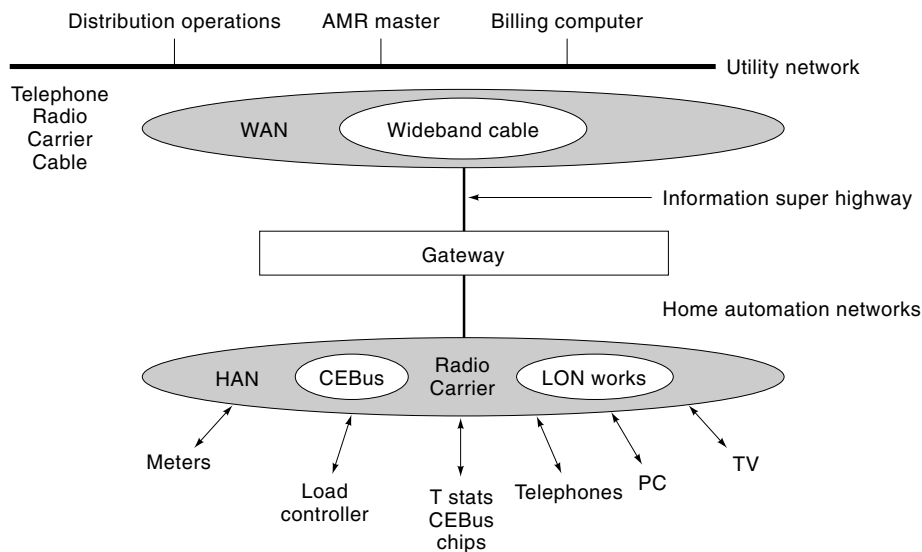


Figure 5. Automatic meter reading, billing, and control/service provision to most electrical devices and communications via cable and the information super highway. Other communication gateways possible are telephone modems, radio, power line carrier, and low earth-orbiting satellites. (ABB Power T&D Company, Electric Metering Systems, Raleigh, NC)

and transient current and voltage waveform distortions and unbalance (6). Metering designed for 60/50 Hz operation may have errors under these conditions and this topic is addressed in the next section herein (7). There is a need to determine correct power definitions for nonsinusoidal and unbalanced conditions (3,4,8). Also there is a need to quantify properly and allocate equitably such costs as may be caused by these distortions so as to maintain a good level of power quality. In fact, it is noted in IEEE P1459 that “there is not yet available a generalized power theory that can provide a simultaneous common base for: energy billing, evaluation of electric power quality, and help in detecting the major sources of waveform distortion.” (2).

Calibration and Testing of Watt-Hour Meters. The term *calibration* is understood in the metering industry to define a process whereby the accuracy of the meter is adjusted. The term *testing* is used to define a process by which the accuracy of the meter is determined. Calibration is conducted carefully by each utility company, and results are traceable to the National Institute of Science and Technology. The traceability is accom-

plished by having the utilities transfer standard meters annually compared to known standards at NIST. Then the meters may be adjusted after a correction factor is determined using a watt-hour transfer standard in the utility meter departments primary measuring instrument, such as a Knopp model ST-31 automatic watthour meter testing system. Historically there have been two methods of calibrating polyphase meters: (a) series-parallel (single-phase) and (b) true polyphase. It has been considered to be acceptable for polyphase watt-hour meters to be adjusted using series-parallel calibration. However, in evaluating electromechanical watthour meters in Refs. 9 and 10 it has been found that under full load, three-element meters tend to run more quickly if used in polyphase loads than in series-parallel condition. On the other hand, two-element meters tend to run slow in polyphase load conditions when compared to series-parallel conditions. It should be noted that NIST standards and calibration are done for sinusoidal situations and that this does not necessarily apply very well to nonsinusoidal and unbalanced voltage and current conditions (11-13). These nonsinusoidal situations are becoming more prevalent as was previously discussed.

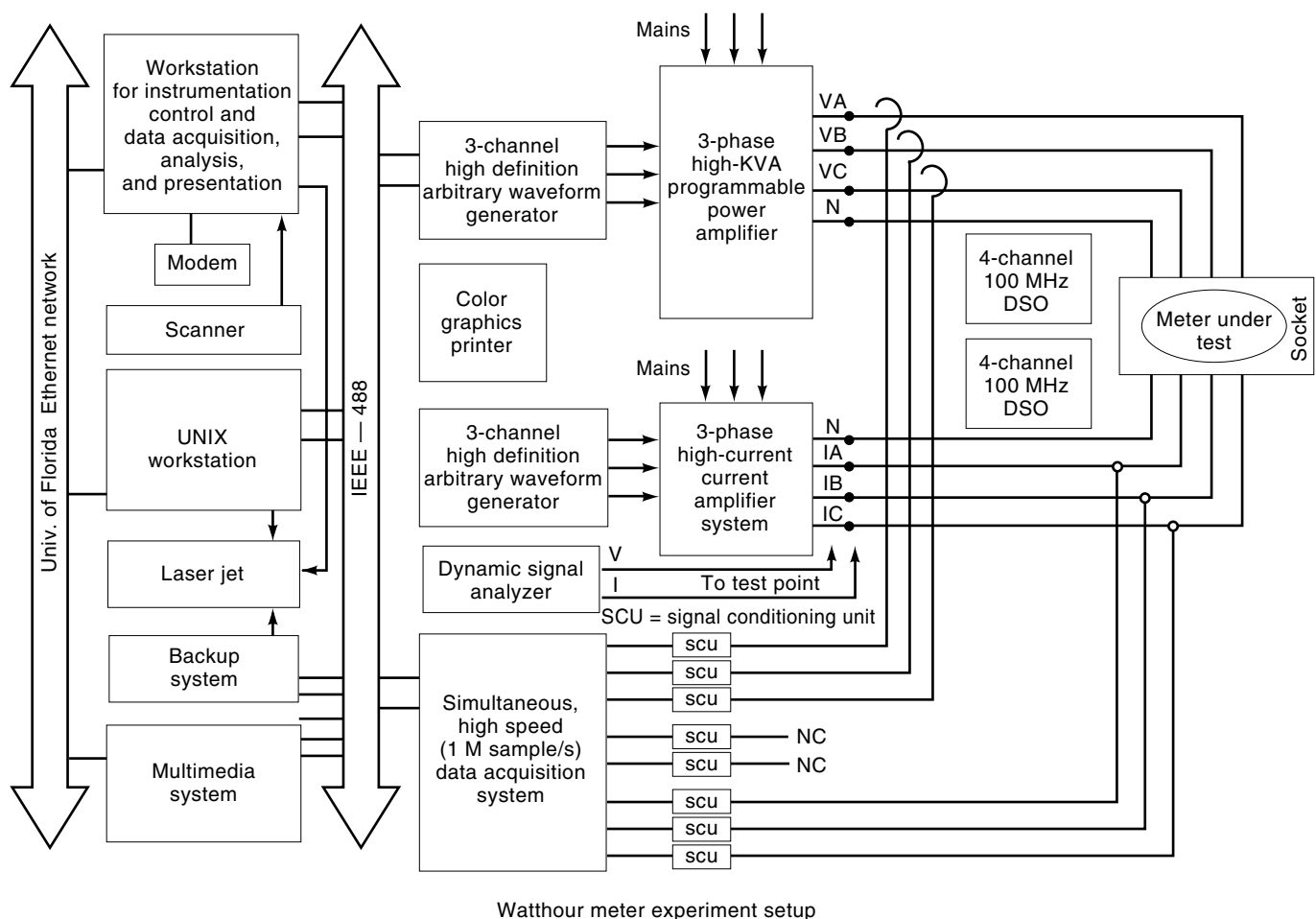


Figure 6. Watthour meter experimental setup for testing the meter’s performance under non-sinusoidal and unbalanced voltage and current waveform conditions. The unique test system at the Power Quality Lab at the University of Florida is capable of producing real-world arbitrary polyphase and synchronized voltage and current waveforms to apply to meters (and other power equipment such as relays, transformers, lighting, motors, and drives) under test to determine, for example, percent registration errors.

Table 2. Three-Phase Watthour Meter Percent Registration Errors^a

	Office Building		Pump Station	
	Magnitude (rms)	THD (%)	Magnitude (rms)	THD (%)
Volts-phase A	122.52	2.96	125.40	3.78
Volts-phase B	123.44	2.68	115.83	2.66
Volts-phase C	121.41	2.09	115.20	2.12
% Unbalance	0.85	18.8	5.55	32.5
Current-A	4.75	4.66	3.39	96.90
Current-B	4.72	4.51	5.03	104.2
Current-C	4.58	2.94	5.71	83.88
% Unbalance	2.21	27.0	28.0	11.7

Meters Tested	Office Building Registration Error (%)	Pump Station Registration Error (%)
Hybrid meter 1	-5.57	-8.75
Digital meter 2	-4.76	-10.09
Solid-state meter 3	-4.37	-9.54
Solid-state meter 4	-3.04	-7.27
Digital meter 5	-0.94	+0.52
TDM meter 6	-0.19	-2.26
Electronic meter 7	-0.18	-3.24
TDM meter 8	-0.13	-2.56
TDM meter 9	-0.04	-0.70

^a The meters were tested when subjected to unbalanced and nonsinusoidal voltage and current waveforms from a pump station (2 up to 150 hp adjustable speed driven pumps) and a typical office building (168,000 ft² with a 980 kW demand with a usual mix of lighting, personal computers, copy machines, and other office equipment loads).

The testing of watt-hour meters, as well as many other power elements, under nonsinusoidal situations may be accomplished with a test system as shown in Fig. 6. Voltage and current waveforms generated mathematically or through field measurements may be downloaded to the computers on the left-hand side of the figure. These computers then digitize the waveforms and send them through the IEEE-488 (GPIB) network to arbitrary waveform generators (one for current and one for voltage). Power amplifiers then boost the analog signals to high power levels so that they may be applied to the meter under test. The readings of the meter under test are compared with a high-speed data acquisition system that is calibrated for each harmonic measurement. The data acquisition system then feeds the signals to computers for analysis and modeling purposes. Results for testing several typical watt-hour meters discussed herein are shown in Table 2. These results are for various voltage and current unbalances and total harmonic distortions. Further results may be found in Ref. 7. Directions for research indicated by these results include development of meter technologies appropriate to nonsinusoidal situations, development and use of appropriate power definitions to process nonsinusoidal, and unbalanced waveforms when computing watt-hours.

Unbundled Power Quality Services: Revenue Metering Implications. Some utilities are expecting to be able to provide various levels of power quality services to customers. One type of service may provide "premium quality" or "UPS" grade power to a customer, and another may provide low grade power at a

lesser cost to the customer. In this situation, it becomes important for the utility to be able to have the metering technology available to meter appropriately and bill the load under nonsinusoidal situations (12). The pressing business and research challenge is to design meters and develop definitions of power and billing schemes that are correct for various grades of electric services (13–16).

ACKNOWLEDGMENTS

The author is grateful for serving as conference chairman for the Southeastern Electricity Metering Association and its annual conference, now in its 73rd year, for the information learned on metering practice and technology, which formed the inspiration to create this article. In particular, the author appreciates the fine talks given at a number of SEMA conferences by D. R. Knight, Metering Department, Jacksonville Electric Authority, on the subject of metering history, which formed the basis of the development of revenue metering material herein.

BIBLIOGRAPHY

1. A. Domijan, Electricity metering, power solutions, customer relations, management and technology school, *72nd Annu. Conf.*, Southeastern Electricity Metering Assoc., Gainesville, FL: Univ. of Florida, 1997.
2. A. E. Emanuel, Standard definitions for the measurement of electrical power quantities under sinusoidal, nonsinusoidal, balanced or unbalanced conditions, IEEE P1459/D1/B, IEEE Working Group on Nonsinusoidal Situations, New York, 1998, pp. 1–31.
3. P. S. Filipsky, Y. Baghzouz, and M. D. Cox, Discussion of power definitions contained in the IEEE dictionary, *IEEE Trans. Power Delivery*, **9**: 1237–1244, 1994.
4. L. S. Czarnecki, What is wrong with Budeanu's concept of reactive and distortion power and why it should be abandoned, *IEEE Trans. Instrum. Meas.*, **IM-36**: 480–487, 1987.
5. A. Domijan, Voltage and current transducer harmonic sensitivity analysis, *ASHRAE Trans.*, **104** (1a): 550–558, 1998.
6. A. Emanuel, On the assessment of harmonic pollution, *IEEE Trans. Power Deliv.*, **10**: 1693–1698, 1995.
7. A. Domijan et al., Watt-hour meter accuracy under controlled unbalanced harmonic voltage and current conditions, *IEEE Trans. Power Deliv.*, **11**: 64–78, 1996.
8. A. E. Emanuel (ed.), *NonSinusoidal Situations: Effects on the Performance of Meters and Definitions of Power*, IEEE Tutorial Course, IEEE Catalog No. 90EH0327-7-PWR, 1990.
9. A. Domijan, J. Miller, and R. D. Knight, A study of polyphase and single phase series calibration testing techniques applied to polyphase watt-hour meters, *IEEE Trans. Power Deliv.*, **8**: 1712–1720, 1993.
10. R. D. Williams et al., Series-parallel versus true polyphase calibration of electricity meters, *IEEE Trans. Power Deliv.*, **13**: 351–355, 1998.
11. D. Czarkowski and A. Domijan, Performance of electric power meters and analyzers in adjustable speed drive applications, *ASHRAE Trans.*, **103** (1): 235–240, 1997.
12. A. Domijan, The implications of measurement errors in high end instruments to unbundled power quality services, *Proc. Nat. Sci.*

- Foundation Conf. Unbundled Power Quality Services Power Ind.*, Key West, FL: 1996, pp. 77–81.
13. A. Domijan and D. Czarkowski, Challenges in electrical power measurements of adjustable speed drives and motors used in HVAC equipment, *ASHRAE J.*, **39**: 32–41, 1997.
 14. A. Domijan, D. Czankowski, and J. H. Johnson, Power measurements of adjustable speed drives, *Proc. of Ind. Applic. Conf.*, New Orleans, LA, No. IAS97314, 1997, pp. 1–7.
 15. A. Domijan et al., Directions of research on electric power quality, *IEEE Trans. Power Delivery*, **8** (1): 429–436, 1999.
 16. A. E. Emanuel, Nonsinusoidal situations: Effects on the performance of meters and definitions of power, *IEEE*, Piscataway, NJ, IEEE Catalog No. 90EH0327-7-PWR, pp. 1–210, 1990.

ALEXANDER DOMIJAN, JR.
University of Florida

INSTRUMENTATION FOR CHEMISTRY AND PHYSICS

PHOTOMETERS

A *photometer* is an instrument for measuring light. The science of light measurement is called *photometry*. Light in the present context is electromagnetic radiation that can be detected by the human eye, extending from a wavelength of about 360 nm at the blue end of the spectrum to about 830 nm at the red end; the actual limits depend to some extent on the individual. The terminology of photometry is covered in Refs. 1 and 2.

THREE BASIC CLASSES OF PHOTOMETERS

The earliest form of photometers used the human eye as a detector. The eye, because of its accommodation to light and other physiological factors, cannot be used to quantify light with accuracy, but it does rather precisely discern differences in brightness. This is the basis of the *visual (subjective) photometer*. This type of photometer, first developed in the nineteenth century, allows an observer to visually compare light from two (test and reference) sources. The sources illuminate adjacent screens as they are viewed simultaneously. The distances between the sources and screens are adjusted until a match in brightness of the screens is observed. The light output of the test source can then be calculated from the known intensity of the reference (known) source and the measured distances between the sources and the screens using the inverse square law (explained below). One embodiment of visual photometer is called a *bench photometer*; shown schematically in Fig. 1.

Beginning in about 1940, another type of photometer using an electronic detector instead of the eye came into use, the *photoelectric (objective or physical) photometer*. In this type of photometer, a photoelectric detector converts light into an electrical current. In its most basic form, this current, which is proportional to the amount of incident light, is measured by observing the deflection of the needle of a galvanometer, as shown schematically in Fig. 2. Photoelectric photometers range from small, battery-powered portable instruments to

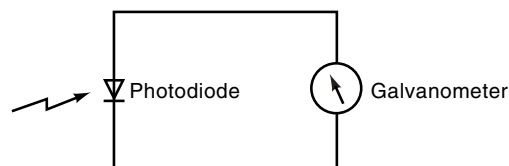


Figure 2. Schematic of the basic photoelectric photometer. An electrical current generated within the photodiode is proportional to the intensity of the light being measured. The deflection of the galvanometer is proportional to the light intensity.

large, precise laboratory instruments. Such physical photometers are made to mimic the human eye in terms of spectral response. The relative spectral responsivity of the average human eye was defined by the International Commission on Illumination (CIE) in 1924, and is generally called the $V(\lambda)$ function, see Ref. 3.

In addition to the two basic types of photometers, a number of specialized photometers have been developed. Rather than measure the quantity of light directly, these instruments either measure parameters that have some relationship to the amount of incident light or have some relationship to the amount of light absorbed or scattered from a light beam. Unlike the former two types of photometers, these specialized photometers may not necessarily measure light according to the spectral response of the human eye. All three types (visual, photoelectric, and specialized) are described in detail in this article.

UNITS OF PHOTOMETRY

The photometric quantities used in this article are *luminous flux* [light output of a source, in lumens (lm)], *luminous intensity* [luminous flux emitted per unit solid angle, in candelas (cd)], *luminance or brightness* [luminous flux emitted per unit area per unit solid angle, in candelas per square meter (cd/m^2)], and *illuminance* [incident luminous flux per unit area in lux (lx)]. The units are Système International (SI) (see Radiometry).

THE DISTINCTION BETWEEN PHOTOMETERS AND RADIOMETERS

The distinction needs to be made between a photometer and a radiometer. While the photometer is designed to have a spectral response range equivalent to that of the human eye, a radiometer generally has a spectrally uniform response and is used to make measurements at wavelengths outside and, in some cases, inside of the visible range as well. The equivalent to photometry in this sense is radiometry. In some countries, the translated word for photometer has the same meaning as radiometer. A few exceptional cases where a specialized photometer senses radiation outside of the visible range will be noted in this article.

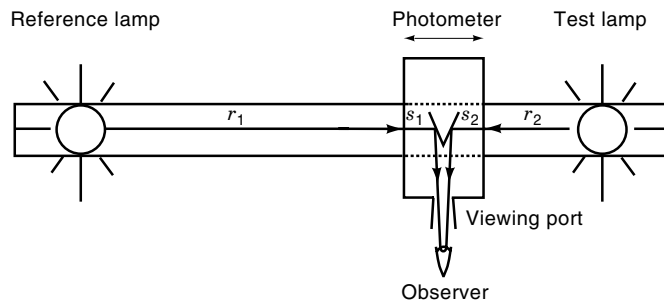


Figure 1. Schematic of a bench photometer. The distances between the photometer head and the lamps, r_1 and r_2 , are adjusted until screens s_1 and s_2 appear to the observer to be of equal brightness.

VISUAL PHOTOMETERS

As described above, visual photometers depend on the ability of the human eye to match the brightness of two adjacent screens, one illuminated by a reference source, the other by a test source. We will describe the bench photometer mentioned above in more detail. The luminous intensity of the test lamp, assumed to be a point source of light, is calculated from the positions of the test and reference lamps and the known intensity of the reference lamp using the inverse square law,

$$E = I/r^2$$

This law states that the illuminance (E) at the given distance (r) from a light source is directly proportional to the luminous intensity of the source (I) and inversely proportional to r^2 .

When the observed illuminance levels on the two screens are the same, the unknown intensity can then be calculated using

$$I = I_0(r/r_0)^2$$

where I and I_0 are the luminous intensities of the test and reference lamps, respectively, and r and r_0 are the corresponding distances from the screen.

In addition to its use to measure unknown sources, the bench photometer can be used together with calibrated light attenuating filters to demonstrate the inverse square law. A commercially available bench photometer is shown in Fig. 3. This photometer is normally attached to an optical bench, which allows it to be moved between sources located near each end of the bench. Fluorescent disks are used as screens in



Figure 3. A commercially available bench photometer. The photometer is moved in relation to the two light sources located near the ends of an optical bench until equal brightness on two fluorescent disks is observed through the eyepiece. (Courtesy of PASCO scientific.)

the instrument shown in the figure; the conical tube is the eyepiece.

Another form of visual photometer, which is similar to the bench photometer, is called the *wax or grease-spot photometer*. A flat surface is coated with either wax or grease to make it translucent. Light coming from both sources can be observed from one side of the translucent material until both the greased (or waxed) areas have the same observed brightness.

The *Lummer–Brodhun photometer*, or contrast light box, is a more sophisticated and compact form of visual photometer. This instrument consists of a several prisms (as well as mirrors in some designs) arranged to allow a viewer to see adjacent screens through an eyepiece while moving the sight box between the two sources. This is a more compact, flexible configuration than the bench style of instrument, but works on the same principle. The basics of visual photometry, including the *Lummer–Brodhun photometer*, are discussed in Hausmann and Slack (4).

If the colors of the two lamps are not matched, it is difficult for an observer to compare to the brightness of the two screens. To solve this problem, the *flicker photometer* was developed. In this instrument, light from the two sources is alternately imposed on a screen at such a rate that the observer can no longer discern that there are two colors. Measurement is made when the sources are adjusted so that the observer is no longer aware of flicker.

Visual photometers are rather cumbersome to use and depend on the observer's subjective ability to make contrast or flicker distinctions. Photometry became a much more exacting and easily implemented science with the advent of the photoelectric detector. Photoelectric instruments have largely supplanted visual photometers; visual instruments are rarely used today except for educational purposes.

PHOTOELECTRIC PHOTOMETERS

Photoelectric photometers depend on the ability of certain non-metallic materials or combination of materials to generate an electric current in the presence of light. The first photoelectric detector used to any great extent was the selenium cell. Selenium is a polycrystalline, nonmetallic element and is used in photovoltaic (i.e. generating an electric current without an external electrical power source in the presence of light) detection devices. Selenium cells were widely used for illuminance meters, photographic light meters, and many other applications until the early 1980s. These early day detectors could drive a sensitive galvanometer movement without use of an external power source, making them ideal for portable light meters. The selenium detector has a spectral response (i.e., response as a function of wavelength) that overlaps the visible region and the spectral response of most photographic films.

Today the silicon detector has replaced selenium cells. Silicon is an element used as a detector in single-crystalline form. The material is doped with small amounts of impurities to form a barrier junction where current is generated by entering photons. Silicon has a significant response that overlaps the visible region. This detector can be used in either photovoltaic mode or photoconductive mode, the latter requiring an external current source. When used in either mode, the electric current is proportional to the incident light over several decades. A special optical filter is used to match the spectral

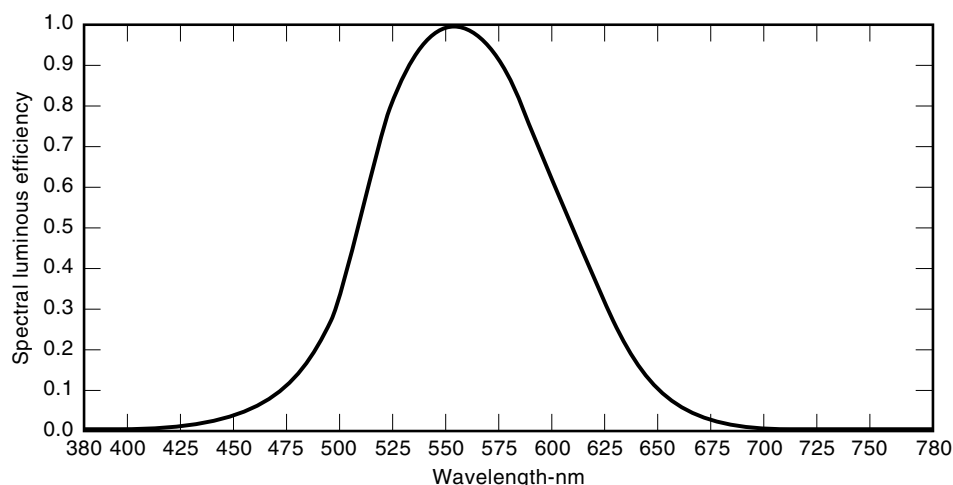


Figure 4. Photopic spectral luminous efficiency function, $V(\lambda)$, for CIE 1924 standard photometric observer.

response to that of the average human eye, the V - λ or $V(\lambda)$ function (see Fig. 4). An operational amplifier (also called a current-to-voltage converter or transimpedance amplifier) is generally used to convert the detector's output current to a voltage proportional to the light signal. This output voltage is detected using a calibrated readout device (analog or numerical display) (see Fig. 5).

Photomultiplier tubes may be used as detectors in place of silicon detectors for increased sensitivity, especially toward the short-wavelength end of the visible spectrum.

A representative high-accuracy laboratory photometer is shown in Fig. 6. This instrument may be configured to measure luminous flux, luminous intensity, luminance, or illuminance. The unit on the right is used to measure luminance or luminous intensity; the operator aims the detector head at the area to be measured by looking through a viewfinder. To measure illuminance, a cosine receptor

consisting of a transmitting diffuser or an integrating sphere is used as an alternative detector head. Some laboratory photometers can be computer-controlled and can be configured with a number of program-selected detector heads connected to one controller. An instrument designed to measure only illuminance or luminance is called an illuminance meter (or illuminometer) or luminance meter respectively.

An instrument for measuring the total luminous flux is called an *integrating sphere photometer*. It consists of two joined hemispheres coated on the inside with a light-scattering, diffuse white material, typically barium sulfate. The spheres can be opened so that a light source can be placed inside. After the hemispheres are closed to form a sphere, all of the light emitted by the source is captured inside of the sphere and scattered uniformly in all directions by a number of interreflections between the sphere wall surfaces. A detector is attached to the side of the sphere and produces a signal proportional to the total flux emitted by the source. One or more diffuse white baffles are placed inside the sphere, so the detector does not receive light directly from the source; only light that has been reflected from the sphere wall (i.e., the light scattered a number of times

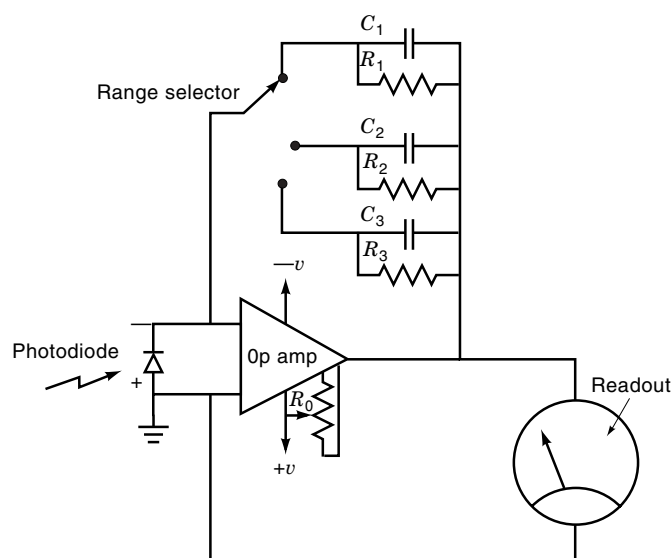


Figure 5. Schematic of silicon detector operated in photovoltaic mode with operational amplifier and readout device. Resistors R_1 to R_3 are used to set the sensitivity range, and resistor R_0 is used to set the zero. Bypass capacitors, C_1 to C_3 , are used to suppress alternating-current noise.

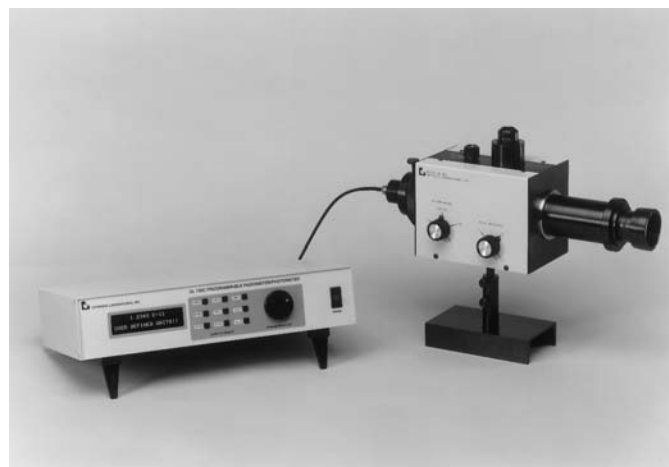


Figure 6. A high-accuracy laboratory photometer. The programmable controller is on the left, and a detector head for measuring luminance or luminous intensity is on the right. (Courtesy of Optronic Laboratories.)



Figure 7. Two-meter-diameter integrating sphere photometer. (Courtesy of the UK's National Physical Laboratory. Copyright © Crown 1998. Reproduced by permission of the Controller of HMSO.)

off of the sphere's surface and baffles) will reach the detector. An example of an integrating sphere photometer is seen in Fig. 7.

The fundamentals of photoelectric detectors and transimpedance amplifiers are covered in Deboo and Burrous (5).

SPECIALIZED PHOTOMETERS

The word *photometer* has its roots in the Latin word *photometrum*, from *phot-* (light) + *metrum* (meter), and is used in the broadest sense to describe a number of instruments used to make light measurements. The visual or photoelectric photometers described previously are instruments for directly measuring one or more of the photometric quantities. The present section will describe examples of different types of photometers that measure quantities directly related to the amount of light or the amount absorbed or scattered by various substances.

The *photographic light meter* or *exposure meter* is an example of a photometer calibrated in other than photometric units. In one form of light meter, the deflection of a needle is proportional to light level. The instrument is constructed so that when a manually aligned pointer coincides with the needle, the photographic exposure (*f*-stop and exposure time) corresponding to the preset film speed can be read directly from a dial. One variant of this concept includes light meters integral to the camera itself, so that the aperture and speed can be adjusted to match needles while looking through the camera's viewfinder. Modern exposure meters, such as the one shown in Fig. 8, use digital displays to indicate exposure. Some light meters are manufactured to read either incident or reflected light; some are calibrated in photometric units as well as units of exposure.



Figure 8. A modern photographic exposure meter with digital readout. The translucent white diffuser is used for incident light measurements. (Courtesy of the Minolta Corporation.)

A number of photometers are designed for measuring the optical properties of substances. The word *colorimeter* is used to describe photometers for chemical analysis based on light absorption by either the test substance or a chemical product of the test substance obtained by use of a reagent. These instruments consist of one or more light sources, narrow band-pass interference filters for selecting wavelengths specific to the substance being measured, and a detection system. Some of them employ a transparent sample holder (cell, ampoule, or cuvette). A measurement is taken with an empty sample holder in place, and a second measurement is taken with a sample in the holder. The ratio of the two measurements is a measure of the opacity of the sample at the selected wavelength and can be converted to the amount of a specific substance in the sample.

An example of this type of photometer is the *water analysis photometer*. A typical instrument of this type has five filters in the visible region, uses a blue-enhanced silicon photodiode detector, and can measure 40 water parameters. An example is shown in Fig. 9.

Another example of this type of photometer is the *flame photometer*, also called an *atomic-absorption photometer*. Rather than using liquids, the flame photometer burns the test substance in propane or natural gas. Here the strength of absorption lines from the burned substance at filtered wavelengths are a measure of the amount of unknown in a given amount of substance. An *optical-emission spectrometer* is similar to a

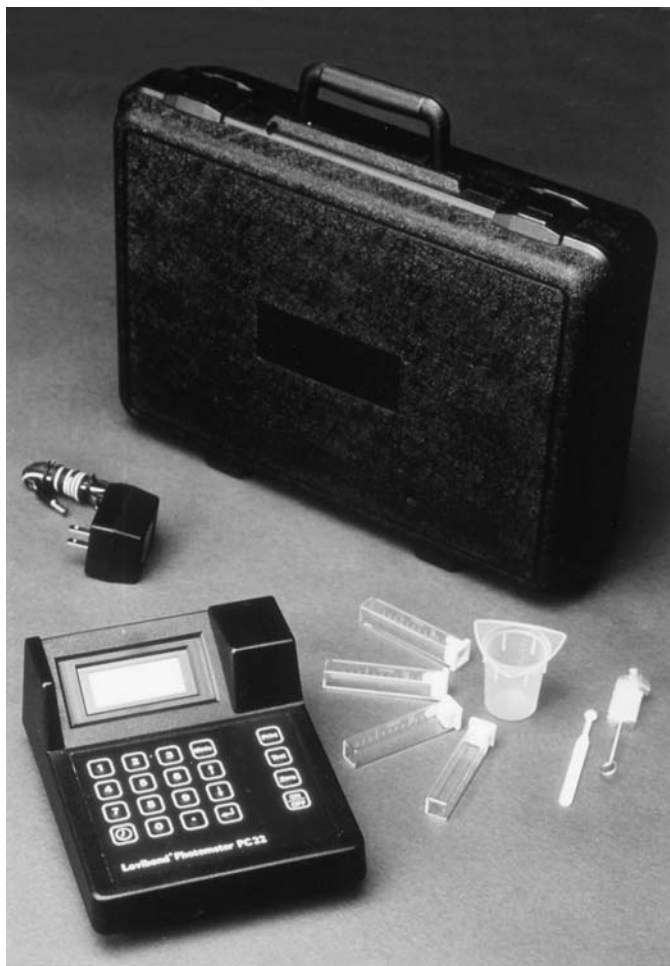


Figure 9. A single-channel, microprocessor-controlled photometer used to measure impurities in water. The transparent cells shown on the right each hold 10 ml of water and reagent. (Courtesy of VWR Scientific Products.)

flame photometer except that an electric spark rather than a flame is used to vaporize the sample.

A *light-scattering photometer* makes use of the light-scattering properties of particles suspended in a jet of gas. Light from particles is primarily scattered in the forward direction, so the detector is aimed at angles looking toward the forward direction of the incident light. This technique can be used to obtain particle size distributions in clean-room environments, clouds, smoke stacks, and so on. Lasers are generally used in this application, because they are monochromatic (single wavelength) and highly collimated so that the scattering volume can be accurately defined. An instrument for measuring airborne particles is called a *nephelometer*.

Some photometers are very specific to a given profession or industry. For example, a specialized photometer for measuring blood hemoglobin is called a *hemoglobinometer*. A specialized photometer for evaluating the brightness of paper and similar products is called a *brightness meter*; it measures the diffuse reflectance of the paper in the blue range of the spectrum. A similar instrument, the *glossmeter*, is used to measure specularly reflected light from a flat sample, with the incident and reflected light making the same angle with the test surface. A *saccharimeter* is a type of photometer for analyzing the

concentration of sugar solutions based on the angle of rotation of plane-polarized light passing through the sample. Photometers utilizing polarization are known as *polarimeters*.

A *spectrophotometer* consists of a light source, a monochromator, a sample space, and a photodetector. In one type, wavelengths are selected by using a scanning monochromator, where a diffraction grating is rotated to select the appropriate wavelengths. Another version uses a grating and diode array detector, which simultaneously detects multiple wavelengths at discrete intervals. These instruments are made to operate in the infrared (*infrared spectrophotometer*), ultraviolet (*ultraviolet spectrophotometer*), and visible regions of the spectrum. Sometimes a spectrophotometer will have two independent channels (reference and test), referred to as a *dual-beam spectrophotometer*. A portable, abridged dual-beam, array detector spectrophotometer is shown in Fig. 10.

Additional information on specialized photometers can be found in Horton (6).

SUN PHOTOMETERS

The *sun photometer* is one of the more interesting and complex types of photometers and deserves special mention. This instrument is used to measure certain optical properties of the earth's atmosphere, and information about atmospheric constituents can be derived from these measurements. Since there is increasing interest in the effects of ozone depletion, atmospheric pollution, and water vapor on visibility, health effects, air quality, and global warming, there is increasing interest in this type of photometer. A sun photometer consists of a basic filtered photometer in which the viewing field-of-view has been limited (typically to about 2°) by use of a tube containing one or more apertures. Sun photometers can be used either on the ground (terrestrial) or from an airborne platform. The optical path between the instrument and the sun must be cloud-free.

The instrument is aimed at the sun either manually or with an automatic solar tracker, so that it is measuring



Figure 10. Portable, dual-beam, array detector spectrophotometer used for making precision color measurements. The sample material is illuminated by a pulsed-Xe lamp built into the instrument. (Courtesy of GretagMacbeth, © 1998. GretagMacbeth™ is a trademark of GretagMacbeth LLC. All rights reserved. Color-Eye is a registered trademark of GretagMacbeth LLC. All rights reserved.)

incoming solar energy over a narrow wavelength band (typically 10 nm wide) of radiation coming directly from the sun (plus a small scattered component). Since certain atmospheric constituents attenuate and/or scatter solar radiation in a wavelength-dependent way, information about the atmospheric constituents can be measured at the appropriate critical wavelengths. Sun photometers are used to measure atmospheric aerosol optical thickness (turbidity), ozone, and water vapor. Measurements taken at discrete wavelengths can be entered into atmospheric transmission models that calculate an approximation to the entire solar spectrum.

Operation of the sun photometer is based on Beer's law, namely,

$$S(\lambda) = S_0(\lambda)e^{-\delta(\lambda)M}$$

where

$S(\lambda)$ = the terrestrial signal at wavelength λ

$S_0(\lambda)$ = the extraterrestrial (i.e., outside the earth's atmosphere) signal at λ

$\delta(\lambda)$ = the total optical depth at λ

M = the relative air mass

Taking the log of both sides of this equation gives,

$$\ln S(\lambda) = \ln S_0(\lambda) - \delta(\lambda)M \quad (1)$$

which in turn yields a linear equation with dependent variable $S(\lambda)$, independent variable M , slope $\delta(\lambda)$, and y -axis intercept $S_0(\lambda)$. A plot of this equation (Fig. 11) is called a Langley plot, in honor of its inventor.

The relative air mass M is defined as unity at the zenith; neglecting atmospheric refraction, it is the relative atmospheric path length through the atmosphere, namely,

$$M = 1/\cos(\theta)$$

where θ is the angle between the zenith and the solar position.

The total optical path, $\delta(\lambda)$, is composed of three components:

$$\delta(\lambda) = (p/p_0)\delta_R(\lambda) + \delta_0(\lambda) + \delta_a(\lambda) \quad (2)$$

where

p/p_0 = ratio of surface atmospheric pressure to reference atmospheric pressure, $p_0 = 101.3$ kPa

$\delta_R(\lambda)$ = Rayleigh (molecular) scattering optical depth at λ

$\delta_0(\lambda)$ = ozone optical depth at λ

$\delta_a(\lambda)$ = aerosol optical depth at λ , also called turbidity

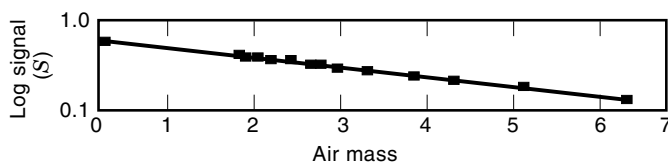


Figure 11. Langley plot of log signal versus air mass. The y -axis interception of the best linear fit to the data gives the calibration constant V_0 for the instrument.

$\delta_a(\lambda)$ can be determined from any reading $V(\lambda)$ under cloud-free conditions, solving Eq. (1) for $\delta(\lambda)$ and calculating $\delta_a(\lambda)$ from Eq. (2) since $\delta_R(\lambda)$ values are available as a function λ and $\delta_0(\lambda)$ can be estimated based on location on the earth's surface and date.

A sun photometer is calibrated by taking successive readings throughout a clear, stable (pristine) day, determining a best-line fit to the Eq. (1), and then calculating $V_0(\lambda)$, the y -axis intercept. The World Meteorological Organization (WMO) recommended wavelengths are 368, 500, 778, 675, and 862 nm, with either 5 nm or 10 nm band pass (see Ref. 7). Typical field of view for these instruments is 2.3° . A four-channel sun photometer with solar tracker is shown in Fig. 12.

In addition to measuring aerosol optical depth, sun photometers can be used to measure the amount of water vapor in the atmosphere using channels centered in the near infrared at 862 nm and either 942 nm or 1020 nm. The 862 nm channel is an atmospheric transmission window; that is, there is no significant absorption of solar radiation by water vapor or aerosol, while there is a significant water vapor absorption band centered at 942 nm and 1024 nm. The amount of precipitable water vapor as a function of optical transmission can be derived from fundamental equations of optical transmission. A good reference on sun photometry is found in Ref. 8. Note that in this case the wavelengths used for water vapor determination are outside of the visible region.

A simple hand-held sun photometer can be constructed for around \$25. Rather than use an expensive filter and detector combination, a red light-emitting diode (LED) is used as



Figure 12. Automatic four-channel sun photometer. (Courtesy Eko Trading Co., Ltd.)

detector. In this reverse role, the LED generates a voltage proportional to the incident optical signal at its normal radiating wavelength. This instrument can be used to monitor atmospheric pollution and haze [see Carlson (9) and visit the Concord web site (10)].

CALIBRATION OF PHOTOELECTRIC PHOTOMETERS

Calibration of photoelectric photometers is based on a unit of luminous intensity, the candela. The candela was originally defined in 1948 by the Conférence Générale des Poids et Mesures (CGPM) as 1/60 of the luminous intensity per square centimeter of a blackbody at the solidification temperature of platinum (2042 K). This calibration was based on basic physical principles—that is, the Stefan–Boltzmann radiation law for blackbody radiators. A point source of 1 cd luminous intensity radiates one lumen of luminous flux into a solid angle of one steradian. However, this old definition using the platinum blackbody was difficult to achieve. The advent of the highly accurate electrical substitution radiometer made it possible in 1979 to redefine the candela as “the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and that has a radiance intensity in that direction of (1/683) watt per steradian.”

In the United States, the candela is maintained on a set of eight *standard photometers* maintained at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD. Each standard photometer consists of an Si detector, a $V(\lambda)$ filter, and an electronic amplifier. The $V(\lambda)$ filter is matched to each detector to give a response representative of the average human eye. Each detector is calibrated for its radiometric response by interpolation of its spectral response (A/W) measured using a high-accuracy cryogenic radiometer (HACR) at selected stabilized laser beam wavelengths over the visible portion of the spectrum. The illuminance responsivity scale (A/lx) is realized by adding a precision aperture (to define the detector's sensitive area) and an accurately characterized $V(\lambda)$ filter. Since the illuminance can now be accurately measured and the geometry can be well established, luminous intensity (cd) of transfer lamps can be accurately measured using the standard photometers.

The NIST luminous flux unit (lm) is defined from the detector-based candela using a calibrated integrating sphere photometer. Calibration is implemented by introducing a precisely measured amount of flux into the sphere from external standard lamps. Lamp standards are then calibrated by placing them inside of the calibrated integrating sphere.

An NIST standard of luminance is realized by use of a calibrated integrating sphere source. Using the standard photometers, precision apertures at the sphere's exit port, and precise aperture-to-photometer distance, the luminance at the aperture plane is established. A reference luminance meter is used as a working standard for routine calibrations.

The NIST provides photometric calibration services for photometer heads, luminance meters, and illuminance meters. Artifacts used to calibrate photometers include lamps calibrated in units of luminous intensity, luminous flux, and illuminance at specified distances. Integrating sphere calibration sources are calibrated by NIST in units of luminance (see Fig. 13). Intercomparisons of photometers and artifacts between national standardizing laboratories are used to ensure the integrity

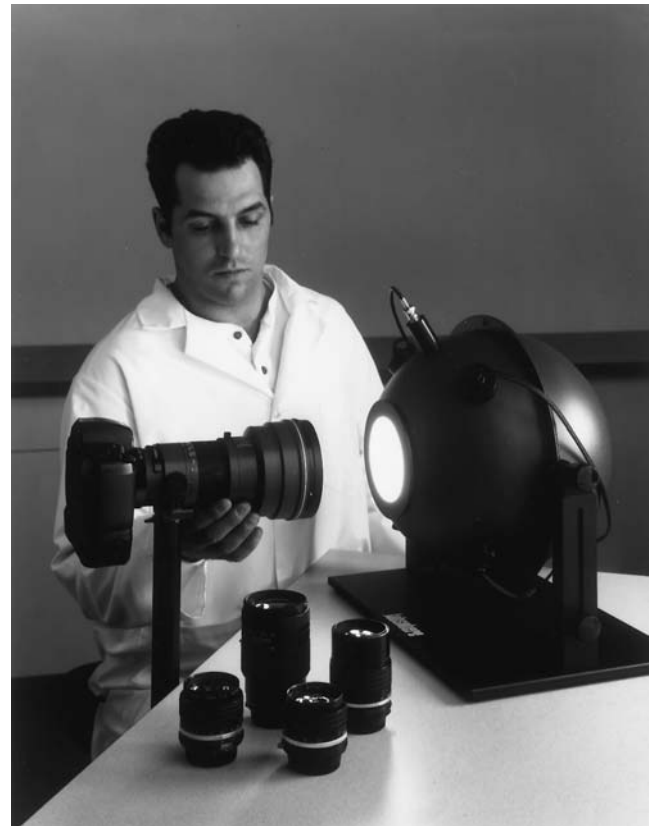


Figure 13. Integrating sphere uniform source for luminance calibration of photometers and imaging systems. (Courtesy of Labsphere Inc., North Sutton, NH)

of standards throughout the world. A detailed description of photometric units and photometric calibration techniques used at NIST can be found in Ohno (11).

BIBLIOGRAPHY

1. *American National Standard Nomenclature and Definitions for Illuminating Engineering*, ANSI/IES Rep. RP-16-1986, New York, N.Y.: American National Standards Inst./Illuminating Eng., 1986.
2. CIE, *International Lighting Vocabulary*, CIE Pub. No. 17.4, IEC Pub. 50(845), 1987.
3. CIE, *The Basis of Physical Photometry*, CIE Pub. No. 18.2, 1983.
4. E. Hausmann and E. P. Slack, *Physics*, 3rd ed., New York: Van Nostrand, 1948, pp. 624–626.
5. G. J. Deboo and C. N. Burrous, *Integrated Circuits and Semiconductor Devices: Theory and Application*, 2nd ed., New York: McGraw-Hill, 1977, chap. 7.
6. G. A. Horton, Photometer, in S. P. Parker (ed.), *Optics Source Book*, New York: McGraw-Hill, 1988.
7. World Meteorological Society, *Guide to Meteorological Instruments and Methods of Observation*, 5th ed., WMO Pub. 8, 1983, sect. 9.3.22.
8. G. E. Shaw et al., Investigations of atmospheric extinction using solar radiation measurements made with a multiple wavelength radiometer, *J. Appl. Meteorol.*, **12**: 374–380, 1973.
9. S. Carlson, The amateur scientist: When hazy skies are rising, *Sci. Amer.*, **276**: 106–107, 1997.
10. [Online]. Available www: <http://www.concord.org/haze>.

11. Y. Ohno, *Photometric Calibrations*, NIS Special Publication **250-37**, Washington, DC: US Government Printing Office, 1997.

Reading List

The following books cover many topics about photometers in detail:

- C. DeCusatis (ed.), *Handbook of Applied Photometry*, New York: Springer-Verlag, 1998.
 W. R. McCluney, *Introduction to Radiometry and Photometry*, Norwood, MA: Artech House, 1994.

THEODORE W. CANNON
 National Renewable Energy
 Laboratory

PHOTOMETRIC LIGHT SOURCES

Any emitter of electromagnetic radiation is a photometric light source. Electromagnetic radiation is produced whenever a charged particle, such as an electron, is accelerated (synchrotron radiation) or decelerated (bremsstrahlung radiation). This occurs in a number of ways. For instance, heat is the irregular motion of electrons, atoms, and molecules. The higher the temperature, the more rapid the motion. Such thermal motion produces irregular oscillatory charge motion, and each oscillation at a particular frequency acts as a tiny “antenna” that emits and receives electromagnetic radiation. Other prevalent physical processes that produce electromagnetic radiation are the quantum energy transitions within materials by electron orbital energy transitions within atoms, recombination of free electrons with atoms, or transitions from excited bonding states of atoms within molecules. Absorbed energy raises a system to an excited state, and electromagnetic radiation is released upon transition to a lower energy state. The frequency of the radiation is related directly to the energy transition. As energy transitions are quantized characteristics of the material makeup, so is the frequency of the electromagnetic radiation produced.

Almost all photometric light sources used in practice result from one of two categories of physical phenomena: (1) incandescence and (2) luminescence. Light from incandescent bodies, such as burning wood or coal, molten iron, and filament wire heated by an electric current, results from thermal motion. Luminescence observed in fluorescent lamps, X-ray fluoroscope screens, organic substances found in fireflies and glowworms, lightning, high intensity discharge lamps, light-emitting diodes, lasers, and, electroluminescent phosphors results from light emission after energy has been absorbed by a material raising its energy state and then subsequently transitioning to a lower energy state. Luminescence is sometimes called “cold light” because a number of photometric light sources based on this phenomenon operate in the vicinity of room temperature. In some cases the distinction between incandescent and luminescent phenomena becomes blurred, such as in shock tubes at very high temperatures, where the collisions of atoms are so violent that electrons dissociate from the atoms and then recombine.

The functionality of a light source is largely determined from the wavelength (frequency) distribution of electromagnetic radiation that it produces. The top diagram of Fig. 1 shows the wavelength range (log scale) and standard nomenclature used

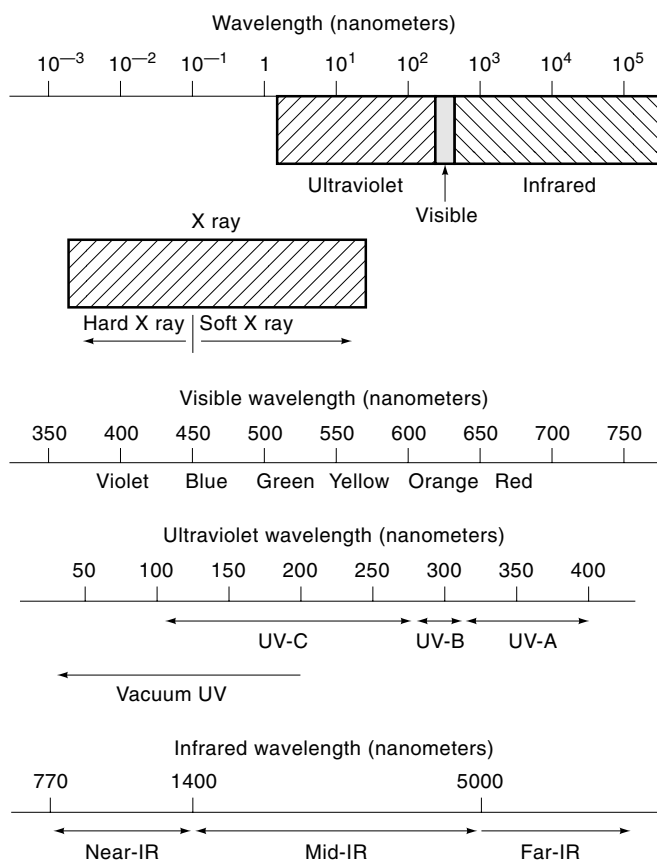


Figure 1. Nomenclature for important divisions and subdivisions of part of the electromagnetic spectrum.

for various spectra of electromagnetic radiation. The three diagrams below in Fig. 1 show further subdivisions of specific wavelength spectra. The term *photometric light sources* is applied to sources of electromagnetic radiation emitting at any of the wavelengths in Fig. 1, although the term “light” is usually only applied to electromagnetic radiation visible to the human eye.

The visible spectrum with a wavelength from roughly 400 to 700 nm produces different color perceptions which are chromatically completely saturated for single wavelength distributions. Multiple wavelength distributions in this range produce less saturated colors, white light, and other colors, such as earth tones. Longer wavelength infrared electromagnetic radiation comprises much of the thermal emission from materials at room temperature to well above 5000 K. The subdivision of infrared radiation into near, intermediate and far is standard terminology that has grown out of various measurement applications, although the divisions are arbitrary. Ultraviolet electromagnetic radiation has a wavelength shorter than visible light and the main difference of this type of radiation from that at longer wavelengths is its ability to ionize atoms by removing electrons. This gives ultraviolet radiation the ability to influence chemical reactions, such as in plant life, and it also makes this type of radiation potentially harmful to human and animal life. Photobiologists have assigned significance to the UV-A, UV-B, and UV-C subregions of the ultraviolet spectrum according to their interaction with biochemical phenomena. Of great importance to ultraviolet spectroscopy is the spectral region below 200 nm

in which ultraviolet radiation is absorbed by air. Hence the term *vacuum* ultraviolet requiring the almost complete evacuation of air from instrumentation measuring or producing such radiation. At still shorter wavelengths (higher frequency) is X-ray electromagnetic radiation. Infrared, visible, and ultraviolet radiation are produced by ionization/recombination in the outer valence electrons of atoms. The X-ray region occurs at the higher ionization/recombination energies involving the inner electron orbitals of atoms. X rays are energetic enough to ionize and also to penetrate many materials. "Hard" and "soft" X rays are differentiated by their penetrating power. The top diagram of Fig. 1 shows other important forms of electromagnetic radiation, longer wavelength millimeter and still longer radio waves to the right and high-frequency energetic gamma rays from cosmic sources to the left.

The generation of electromagnetic radiation by photometric light sources is classified into two categories: (1) systems and processes that produce electromagnetic radiation covering a broad continuous spectrum of frequencies and (2) those that emit electromagnetic radiation of discrete frequencies characteristic of the materials making up the light source. For most photometric light sources, the production of continuous and discrete spectra are concurrent. A classic example of a continuous spectrum-emitting source is an ideal blackbody radiator, which is used as a fundamental reference standard. The term blackbody refers to an object that absorbs all frequencies of electromagnetic radiation, and hence appears black when emitting little radiation at cooler temperatures. The rule of reciprocity states that a body radiates strongly at those frequencies it absorbs. Analogous to the oscillation of charges from thermal motion as "tiny antennas," an antenna is part of an electric resonance circuit that transmits and receives at the same frequency. A piece of black coal is a good example of a blackbody. When heated to increasingly high temperatures, it first glows red, then yellow, and finally white. The continuous distribution of frequencies depends solely on its absolute temperature and the frequency at which the maximum radiative energy increases in proportion to absolute temperature. When glowing red, a blackbody emits a large portion of infrared radiation. When glowing white, a significant amount of ultraviolet radiation is emitted.

Because of the inhibition of certain oscillatory modes of electric charges, in practice all materials are *selective* radiators with respect to emission at a particular wavelength compared with the emission from an ideal blackbody at exactly the same temperature. The ratio of the emission energy of a material to the emission energy of a blackbody at the same temperature and at a particular wavelength is called the wavelength *emissivity* of the material. Emissivity also depends on the direction in which radiation is emitted relative to the body. The closer the emissivity of a material to 1.0, the better the approximation to a blackbody at that wavelength. For instance, tungsten in incandescent filaments has high emissivity in the visible spectrum, but lower emissivity in the infrared. As a visible photometric light source, heated tungsten filaments closely approximate the output of a blackbody. Glass is transparent because it transmits (i.e., lacks absorption) visible light radiation and as a result has low emissivity in the visible spectrum. Heated glass looks relatively colorless compared with heated opaque materials. However glass has even higher emissivity than tungsten in the mid-infrared. References 1–3 provide thorough background on the topics discussed thus far.

Other sources of continuous frequency spectra of electromagnetic radiation are high-energy electron accelerators, appropriately called synchrotrons because they use magnets to accelerate free electrons in circular orbits. The first experimental investigation was conducted in 1947 on the 70 MeV General Electric synchrotron to produce visible light radiation. Subsequently higher energy synchrotrons were developed which produce broad continuous spectrums at higher frequencies into the ultraviolet and X-ray spectral regions.

Sources of discrete frequencies or line spectra are produced from energy transitions characteristic of the atomic and/or molecular structure of the material comprising the light source. Some important examples include the five principal visible spectral lines of mercury gas from a high-intensity mercury discharge lamp and the 589nm strong spectral line from a high-pressure sodium lamp. The intense line spectra for these lamps occur against a lower grade background of a radiative frequency continuum. Laser light is produced by light amplification by stimulated emission of radiation and is coherent (i.e., all emitted light is in the exact same phase), monochromatic or containing only a few monochromatic lines, and is intensely focused within a narrow aperture beam. Atoms of a material are stimulated to an excited state by using a noncoherent light source or with electricity, and synchronous radiative emission occurs from the corresponding synchronously generated energy transitions. An example is the red HeNe laser with its characteristic 632nm radiation.

BLACKBODY RADIATION

In the 1890s the German physicist Wilhelm Wien came up with the idea of constructing a good approximation to an ideal blackbody photometric light source by using an enclosed oven to create a cavity with a small hole through which the cavity can be viewed. The radiation coming out of this hole is very close to the equilibrium blackbody electromagnetic radiation corresponding to the oven temperature. Any external radiation that enters the hole is internally scattered and reflected within the cavity so that almost all of it is ultimately absorbed, and extremely little finds its way back outside the hole. Wien carefully measured the radiative energy dW per wavelength interval $d\lambda$ for the blackbody radiative distribution.

The theory to accurately explain the shape of this empirically observed blackbody radiative distribution came shortly later in 1900 when Planck used the radically new idea that internal electric charge oscillators within the blackbody emit only electromagnetic radiation in finite quantized amounts of energy of size $E = h\nu$, where ν is the frequency of the radiation. With $c = \lambda\nu$, the Planck radiation law accurately predicts that the radiative energy per unit wavelength for a blackbody is given by

$$\frac{dW}{d\lambda} = \frac{8\pi ch\lambda^{-5}}{e^{\frac{hc}{\lambda kT}} - 1}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant from thermodynamics. The value of Planck's constant $h = 6.626 \times 10^{-34}$ J·s is determined from the best fit with empirical data. The properties of this distribution are consistent with Wien's law which states that the product of the wavelength with maximum radiative energy and the absolute temperature is an

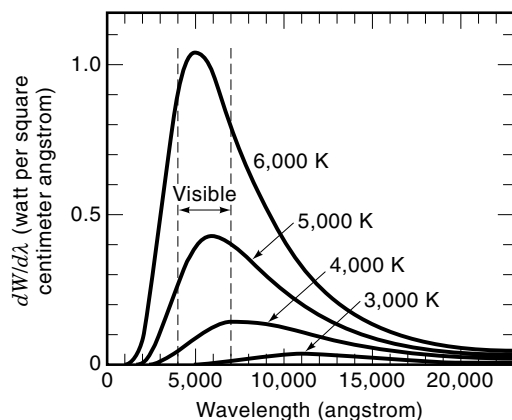


Figure 2. Spectral power distribution for a blackbody at various absolute temperatures, predicted by Planck's radiation law.

absolute constant $\lambda_m T = 0.2898 \text{ cm} \cdot \text{K}$ and are consistent with the Stefan-Boltzmann law that the total radiant energy emitted per second and per unit area W of a blackbody is proportional to the fourth power of the absolute temperature, $W = \sigma T^4$, where $\sigma = 5.67 \times 10^{-12} \text{ W}/(\text{cm}^2 \cdot \text{K}^4)$. Figure 2 shows wavelength distribution plots for a blackbody at various absolute temperatures.

LUMINESCENCE

In contrast to incandescence, radiative emission from luminescent phenomena occurs at relatively cool temperatures. The process by which an appropriate material absorbs energy and then subsequently releases electromagnetic radiation upon an electronic transition to a lower energy state or its original ground state occurs in a variety of ways depending on (1) how the electronic transitions occur and (2) in what form energy is initially absorbed by the material. In all cases the energy transitions involve the outer valence electrons of atoms comprising the material. The field of luminescence is very broad, so discussion here is limited to that most practically relevant to photometric light sources.

For luminescence there is an important distinction between two different types of electronic transitions upon absorbing energy, *fluorescence* and *phosphorescence*. The process of radiative emission from fluorescence occurs very quickly, on the order of 10 ns after initial excitation. The high-energy excited state upon absorption is unstable and therefore rapidly transits back to a lower state. Sometimes there are metastable excited states in fluorescent materials that are much longer-lived before transiting to a lower state. Radiative emission from such a phenomenon is called phosphorescence and, from the time of initial excitation, transition back to a lower energy state takes anywhere from milliseconds to several days.

The predominant sources of energy initiating luminescence in photometric light sources are ultraviolet or visible electromagnetic radiation and electricity. The former is called *photoluminescence* and the latter is called *electroluminescence*. In photoluminescence the wavelength of emitted light typically has a wavelength longer (i.e., lower frequency and therefore lower energy) than the incident radiation. Other sources of energy for luminescence are electron guns, such as in television

tubes, and particles emitted by radioactive materials which is called *radioluminescence*.

MEASUREMENT OF LIGHT SOURCES

The terms "intensity" and "color" of a light source are often used loosely, but each of these has a formal International Standard (SI) definition that is universal in practical application. Units of radiosity apply to energy of light radiation whereas units of luminosity apply to the perception of the energy of light radiation by the human eye. For instance radiant intensity from a light source refers to the radiant flux per unit solid angle, usually expressed in watts per steradian. The SI unit for luminous intensity is the candela. One candela is the perceived luminous intensity of monochromatic radiation at $540 \times 10^{12} \text{ Hz}$ (approximately 555 nm) at the radiant intensity of 1/683 watt per steradian. The *lumen* is the SI unit of luminous flux. One candela is equal to one lumen per steradian. The human spectral photopic response curve is approximately a bell shape which is maximal at 555 nm (yellow-green light) and falls to zero near 400 nm on one side and near 700 nm on the other side. Because the spectral response of the human eye is variable from 400 to 700 nm, the relationship between watts and lumens, therefore, is wavelength-dependent. The ratio of lumens to watts is termed *luminous efficacy*. Maximum efficacy of 683 lm/W is achieved at a wavelength of 555 nm and falls off in proportion to the spectral response to blue and to red. Many visible light sources are rated in terms of the luminous efficacy of their spectral outputs. See reference (7) for more details.

Instruments called *radiometers* are used to measure radiant power (watts) of light sources that is cumulative over wide ranges of wavelengths including the ultraviolet, visible, or infrared spectral regions. Radiometers employ a variety of detectors, such as thermocouples, thermopiles, and pyroelectric detectors. Visible radiometers use a filter transmitting light according to the human spectral photopic response for measuring luminous power (lumens). By incorporating a device called a *monochromator* which separates or disperses the various wavelengths of the spectrum through prisms or gratings, one can determine the radiant power in a very small range of wavelengths. Such an instrument called a *spectroradiometer* measures the spectral power distribution for a light source, which is the radiant power per unit wavelength, as a function of wavelength. From the spectral power distribution, one determines all radiometric, photometric, and colorimetric properties of a light source in the absolute sense.

Radiometers are limited to measuring light power within a small aperture. An instrument called an *integrating sphere photometer* measures the total radiant power emitted by a light source over the entire 4π steradian spherical solid angle. The most common is called the Ulbricht sphere which is about 3 m in diameter and whose inner coated surface, ideally, is coated by a perfectly diffusing substance with uniformly nonselective reflectance. White magnesium oxide and barium sulphate are coating materials that come close to these properties. The light source is placed at the center of the sphere, and every point on the inner surface then reflects to every other point so that the flux incident on each unit area along the inner wall is uniform regardless of the angular distribution of the emitted power from the source. The total flux E received per unit area of the sphere from reflection and multiple interreflections is related to the

total flux F emitted by the source inside the sphere, according to the following expression:

$$E = \frac{\rho F}{4\pi r^2(1 - \rho)}$$

where r is the radius of the sphere and ρ is the total energy reflectance coefficient of the material (4). In practice this formula requires a number of photometric corrections due to nonuniform reflectance in wavelength and nonuniformity across the interior surface of the sphere. Another way of computing total luminous flux is by comparing the measurement of flux incident on a unit area of the inner surface of the sphere for two light sources. If the total emitted power of one of the light sources is known, then the total power of the other light source can be determined.

COLORIMETRY

Perception of the color of light radiation by the human eye was first internationally standardized by the Commission Internationale de l'Eclairage (CIE) in 1931. A system for the CIE standard observer was defined in terms of color matching functions used to convert a spectral power distribution into tristimulus color values. A set of three color matching functions $\bar{C}\bar{I}X(\lambda)$, $\bar{C}\bar{I}Y(\lambda)$, and $\bar{C}\bar{I}Z(\lambda)$ convert a spectral power distribution $P(\lambda)$ into tristimulus values X , Y , and Z according to the following equations

$$X = \int_{\lambda} P(\lambda) \bar{C}\bar{I}X(\lambda) d\lambda$$

$$Y = \int_{\lambda} P(\lambda) \bar{C}\bar{I}Y(\lambda) d\lambda$$

and

$$Z = \int_{\lambda} P(\lambda) \bar{C}\bar{I}Z(\lambda) d\lambda$$

where integration is from 400 to 700 nm. For a given fixed set of color matching functions $\bar{C}\bar{I}X(\lambda)$, $\bar{C}\bar{I}Y(\lambda)$, and $\bar{C}\bar{I}Z(\lambda)$ there is an infinity of spectral distributions $P(\lambda)$ that produce exactly the same tristimulus values X , Y , and Z . Therefore each tristimulus set of values corresponds to an equivalence class of spectral distribution functions called *metamers* with respect to the color matching functions. The chromatic coordinates x , y , z for a color are defined as the ratio of each tristimulus value of the color to their sum:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

and

$$z = \frac{Z}{X + Y + Z}$$

Chromatic coordinates have two degrees of freedom because specifying x and y automatically determines z on the plane $x + y + z = 1$.

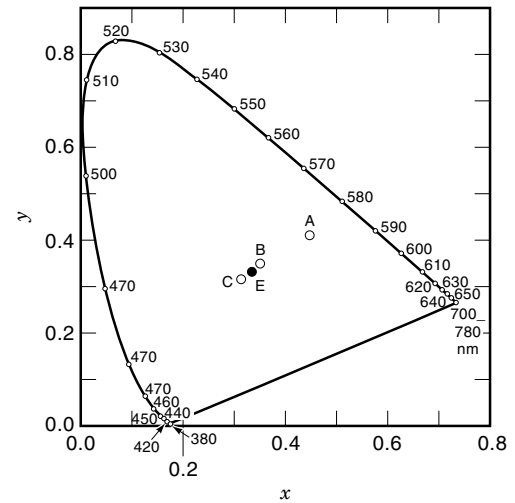


Figure 3. 1931 CIE chromaticity diagram showing chromatic coordinates for the pure spectral colors, CIE standard light sources A, B, and C, and the equal-energy source E (4).

The color matching functions developed for the 1931 CIE standard observer are based on earlier work on color matching for chromatic coordinates of the pure spectral colors. Details of the derivation for these color matching functions along with tabulated and graphed values are in (4). Two important criteria are (1) that the middle color matching function $\bar{C}\bar{I}Y(\lambda)$ is the photopic spectral response function for the human eye and (2) all three color matching functions are nonnegative for all wavelengths in the visible spectrum. The first criterion implies that the tristimulus value Y represents the photopic intensity. Figure 3 shows a plot of the pure monochromatic spectral colors in the x - y CIE chromatic space together with the plot of the CIE standard light source A at $x = 0.448$, $y = 0.407$, the CIE standard light source B at $x = 0.348$, $y = 0.352$, and the CIE standard light source C at $x = 0.310$, $y = 0.316$. The definitions and constructions of these CIE standard light sources are discussed in the subsection "Incandescent Sources." The "equal-energy" source at point E where $x = y = 1/3$ appears as an achromatic gray tone. In 1964 color matching functions were developed to improve performance for a CIE standard observer viewing colors over a wider field of view of 10° , compared with the 1931 CIE standard observer which performs well over a narrow field of view of 2° .

The trace of points in CIE chromatic space that represent x - y values for spectra produced by blackbodies is known as the *Planckian locus* depicted as a solid curve in Fig. 4 for absolute temperatures ranging from 1515 K up to infinity. An important parameter for a visible photometric light source, particularly an incandescent filament source, or a high-intensity discharge lamp is its correlated color temperature or simply its color temperature. The idea is that, given a spectral output for a light source, what is the absolute temperature of a blackbody producing the most similar spectral output sensed by a standard observer. For lamps emitting a spectrum with CIE x - y coordinates in the vicinity of the Planckian locus, the color temperature is the "nearest" point on the Planckian locus to this CIE point. This was formalized by Judd (6) using isothermal lines which were recomputed by Kelly (5) in the 1931 CIE chromatic space shown in Fig. 4. Most often the color temperature of a visible

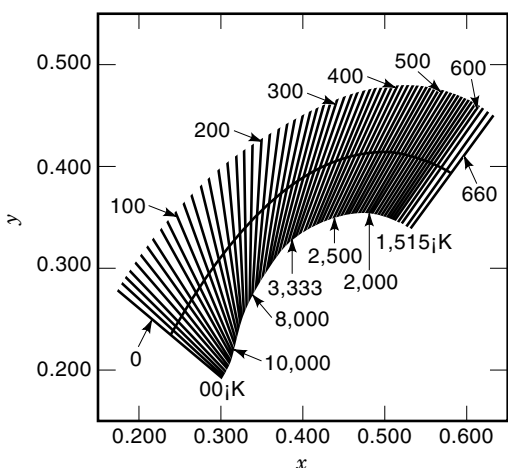


Figure 4. 1931 CIE chromaticity diagram showing the Planckian locus and isotherm lines (4).

light source is not the same as the actual operating temperature of the source. Even for incandescent tungsten filament lamps heated to different temperatures which lie quite close to the Planckian locus, the filament temperature is always lower than the color temperature (e.g., by about 40 K at color temperatures around 2850 K) because of the wavelength-dependent emissivity of tungsten. Visible light sources producing CIE x - y coordinates far from the Planckian locus, such as some fluorescent lamps, are still assigned a color temperature although this is more for convenience in interior lighting design than for physical meaning.

VISIBLE LIGHT SOURCES

A diverse variety of designs are available today as sources of visible light radiation. These include incandescent filament lamps, high-intensity discharge lamps, short-arc lamps, fluorescent lamps, light-emitting diodes, electroluminescent phosphor lamps and even nuclear light sources. A number of these designs simultaneously produce a significant amount of electromagnetic radiation at wavelengths longer and shorter than visible light also making some of them good infrared and ultraviolet sources. The choice of design of a visible light source for a particular application depends on a number of factors including spectral output, radiant intensity requirements, power consumption, luminous efficacy, temperature conditions, and durability.

Incandescent Sources

Incandescent filaments are the most commonly used source of visible light radiation. An electric current is passed through a filament which glows because of thermal motion. The filament is encased in a glass envelope evacuated to prevent oxidation of the filament. The criteria for a filament material are high melting point, low vapor pressure, high strength, high ductility, and suitable radiative and electrical resistance characteristics. Earlier designs for incandescent filaments used carbon, osmium, and tantalum, but tungsten with its high melting point of 3655 K and good adherence to the other desirable properties for a filament is now the most widely used material. Subsequent designs

have alloyed tungsten with metals, such as rhenium, for desired spectral characteristics, and thorium is used in tungsten filaments for rough service applications.

More recently the tungsten-halogen lamp prolongs lifetime and operates at a higher filament temperature up to 3500 K. Metal halides are added inside the glass envelope at low pressures creating a cycle of dissociation and recombination that reduces gradual blackening of the envelope by tungsten evaporated from the filament. The filament heats the metal halide gas disassociating it into metal and halogen. Tungsten evaporated from the filament combines with the halogen. When tungsten halide contacts the hot filament, the tungsten recombines with the filament, disassociating it from the halogen and completing the cycle. Tungsten-halogen bulbs are available with power up to 1000 W using a power supply between 8 and 13 V at high current. Because the output varies roughly as the eighth power of the current, the stability of the power supply is quite stringent.

Standard sources for colorimetry use incandescence from tungsten filaments with added features. Standard source A is a tungsten filament operated so that the correlated color temperature is at 2856 K. Standard source B approximates noon sunlight with a correlated color temperature of approximately 4874 K. Standard source B uses the standard source A in conjunction with a special filter one centimeter thick consisting of layers which include compounds of copper and cobalt. Standard source C approximates daylight provided by a combination of direct sunlight and clear blue sky with a correlated color temperature of approximately 6774 K. Standard source C is of construction similar to standard source B except that the special filter has different concentrations of copper and cobalt compounds. In addition there are D-type standard illuminants D_{55} , D_{65} , and D_{75} which approximate different phases of natural daylight with respective correlated color temperatures of 5500 K, 6500 K, and 7500 K.

High-Intensity-Discharge Sources

When an electric arc is struck within a gas, some of the atoms and molecules of the gas ionize producing free electrons whereas others are excited to high energy states. Subsequent quantum electronic energy transitions within excited atoms and molecules to lower energy states causes the emission of various line spectra characteristic of the gas. A broad continuous spectrum of light radiation is also produced by a number of phenomena including (1) energy transitions from stable excited electronic states of a molecule to lower repulsive states, (2) the recombination of free electrons with ions, and, (3) the acceleration and deceleration of free electrons within the electric arc. Except for (3), all of these phenomena that produce light radiation are examples of electroluminescence. For light production in the visible spectrum, the most common high-intensity-discharge lamps are mercury, metal halide, and high-pressure sodium.

The basic construction of a high-intensity-discharge lamp consists of a starting electrode and a pair of operating electrodes contained in an evacuated glass tube. The tube contains the principal gas or gases responsible for light emission and a rare gas, such as argon or xenon, which is easily ionized, to help initiate the electric arc. The principal gas or gases have very low vapor pressure at room temperature and require very high voltage to create an electric arc between the operating

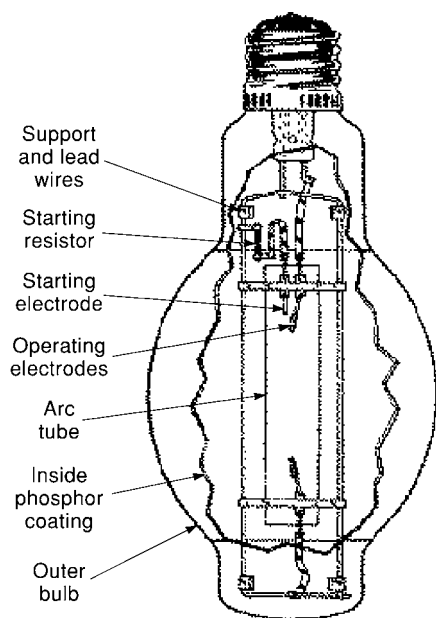


Figure 5. A 400 W phosphor-coated mercury lamp (7).

electrodes. Instead, an arc is initially struck by ionizing the rare gas with the starting electrode heated up. This produces heat which in turn raises the vapor pressure of the principal gas or gases facilitating the creation of an electric arc between the operating electrodes at lower voltages. Adding rare gases also reduces the warm-up time to full operational output.

Figure 5 shows the standard construction of a mercury lamp. The mercury lamp uses mercury gas and approximately 10 to 30 torr pressure of argon gas to initiate the electric arc. When in full operation, the vapor pressure of the mercury gas is usually between two and four atmospheres. The characteristic mercury spectrum consists of five visible principal lines at 404.7, 435.8, 546.1, 577, and 579 nm. This occurs against a continuous broad spectrum which extends well into the ultraviolet. Altogether a mercury lamp appears bluish-white. Increasing the operating vapor pressure of the mercury pushes the continuous spectrum to longer wavelengths (i.e., more red is added). Figure 5 shows the mercury lamp constructed with two glass envelopes. The inner envelope (arc tube) contains the electric arc and the outer envelope shields the arc tube from convection drafts and is usually filled with an inert gas, such as nitrogen, to prevent oxidation of internal parts. The outer envelope also provides an inner surface for a phosphor coating which converts the high amount of emitted UV radiation to visible light by photoluminescence (fluorescence), thereby increasing the efficiency of light power output.

Metal halide lamps are constructed similarly to mercury lamps. Combinations of metal halide gases are added with mercury and argon to produce the spectra of metallic elements. This improves color balance and increases luminous efficacy. Three typical combinations of halides used in metal halide lamps are (1) sodium, thallium, and indium iodides; (2) sodium and scandium iodides; and (3) dysprosium and thallium iodides. Because strong characteristic lines of sodium (589 nm) and thallium (535 nm) are so close to the maximum efficacy wavelength of 555 nm, these particular halides are added to increase efficacy. The use of metal halides in discharge lamps is to combine spectral lines of metals providing two desirable advantages. First, al-

though fused silica comprising the inner arc tube chemically reacts with some metals by themselves, it does not react with the metal halides when they are at the cooler wall temperature. The second advantage is that whereas some metals cannot be vaporized at temperatures which fused silica withstands, they are vaporized when they approach the high temperature central core of the arc tube.

High-pressure sodium lamps are produced by passing an electric current through sodium vapor above 200 torr pressure. Xenon is used as a starting gas. Because of the small diameter of the tube used for sodium lamps, no starting electrode is present, and an arc is initialized by a high-voltage, high-frequency pulse. A high-pressure sodium lamp is golden-white with all frequencies present and the strong 589 nm line characteristic of sodium. Low-pressure sodium lamps principally radiate the 589 nm spectral line and appear as an almost monochromatic yellow. For best efficacy, the vapor pressure for a low-pressure sodium lamp is about 0.005 torr. Although the low-pressure sodium lamp has very high efficacy, it requires much more power than the high-pressure sodium lamp per watt of output.

Short-Arc Sources

Short-arc lamps are basically high-intensity-discharge lamps with very high gas vapor pressure and a small interelectrode distance compared with the diameter of the enveloping tube. Depending upon rated wattage, the length of the electrode-stabilized arc varies from about 0.3 up to 12 mm. The output of these arcs has the highest luminance and radiance of any continuously operating visible light source (i.e., up to well above 30,000 W) and are the closest to a true "point" source. These lamps are used primarily in searchlights, projectors, display systems, and optical instrumentation, such as spectrophotometers.

Mercury and mercury-xenon short-arc lamps are available from 30 to 5000 W operating under the same principle as the mercury high-intensity-discharge lamp. With the same pressure of argon as a starting gas as the standard mercury lamp, the mercury short-arc lamp requires several minutes to achieve full operation. Adding over one atmosphere of xenon gas, the resulting mercury-xenon lamp reduces warm-up time to full operation by about half. The spectral power distribution for both the mercury and the mercury-xenon lamps are essentially the same in the visible spectrum.

Xenon short-arc lamps are filled with approximately five atmospheres of xenon gas and are available from 5 W up to 32,000 W output. Their appearance closely resembles a phase of daylight similar to the visible range of the blackbody spectral distribution at 6000 K. The time to full operational output is quite short, reaching 80% immediately after the initial start.

Fluorescent Sources

The most commonly used fluorescent sources combine light production from electroluminescence and photoluminescence to create a low-temperature discharge lamp. As with the mercury lamp, both mercury and argon gas fills a tube except at much lower pressure. At the operational temperature of 40°C, the vapor pressure of the mercury gas is maintained at about 0.008 torr, and argon gas used to initiate the electric arc is maintained at from one to three torr depending on energy-saving measures. At this low operating pressure, the mercury gas pri-

marily emits a strong characteristic line in the UV at 253.7 nm which is invisible to the human eye. The inside of the tube for the fluorescent source is coated with phosphors designed particularly to fluoresce when stimulated at this wavelength in the UV and reradiate visible light with desirable color properties. Two common phosphors are zinc silicate and magnesium tungstate. In addition to fluorescence, these phosphors exhibit the longer-lived phenomenon of phosphorescence which helps to reduce the stroboscopic effect of ac current operation.

Like most gas-discharge lamps, the electric arc of fluorescent sources must be carefully regulated with a current limiting device called a ballast. The two electrodes hermetically sealed at opposite ends of the tube are designed for either "cold" or "hot" cathode operation, more correctly, respectively, called *glow* and *arc modes* of discharge operation. The tradeoff for these modes of operation is voltage versus current. Electrodes for glow (cold cathode) operation are coated with an electron emissive material, and the standard operating current is on the order of a few hundred milliamperes at 50 V. Electrodes for arc mode (hot cathode) operation are constructed from tungsten wire and, in operation, the current is on the order of 1.5 A at about 10 to 12 V. The lower voltage for the arc mode makes lamp operation more efficient and this is used most frequently.

Light Emitting Diodes

The light emitting diode (LED) is a p - n junction semiconductor device which emits radiation when a forward applied voltage yields a flow of current (see Figs. 6(a) and 6(b)). The material comprising an LED is a specially prepared semiconductor of high purity to which small amounts of other elements are added as controlled "impurities." One type of impurity creates an excess of electrons to produce n -type material, and another impurity has a shortage of electrons (i.e., "holes") which act

as positive charges to produce a p -type material. When a dc voltage is applied to a p - n junction with polarity such that the n -type is negative and the p -type is positive, electrons are forced to meet at the junction and recombine with holes. Light radiation is produced by electroluminescence caused by the recombination of electrons in the conduction band with holes in the valence band. The energy gap crossed during this recombination determines the wavelength of the emitted radiation. Changing the energy gap and, therefore, the wavelength of emitted radiation is achieved by changing the composition of the added impurities. The ratio of the number of emitted photons to the number of electrons crossing the p - n junction is called the quantum efficiency. LEDs are made to produce electromagnetic radiation in various parts of the spectral range from 400 up to 2000 nm. Visible LEDs are used for numeric displays or for indicator lamps. Near-IR LEDs are used as opto-isolators or as sources in optical communication systems. LEDs generally operate in the range of 1 to 3 V at currents in the range of 10 to 100 mA.

The first visible LED with extensive applications and still in wide use today is based on gallium arsenide phosphide ($\text{GaAs}_{1-x}\text{P}_x$). The variable x controls the respective fractions as the elements used. One value used in practice is $x = 0.4$. For smaller values of x the spectral output is largely red light. As x is increased, the quantum efficiency is decreased but the wavelength distribution shifts lower toward 555 nm increasing efficacy. This LED can be made to emit longer wavelength radiation in the dark red and near-infrared by raising x above 0.45 and adding nitrogen. Other LED materials in use include aluminum gallium arsenide ($\text{Al}_x\text{Ga}_{1-x}\text{As}$), and, gallium indium arsenide ($\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$). Spectral output distributions are shown in Fig. 6(c) for some LEDs used in practice.

Semiconductor lasers are based on a variation of the light-emitting diode by carefully constructing structures at

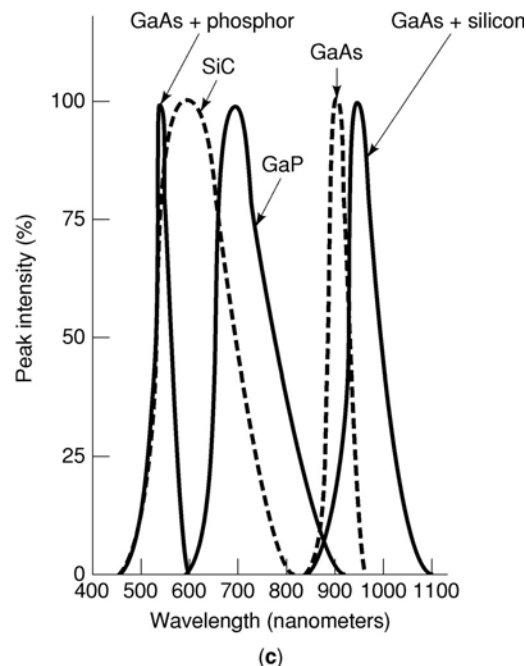
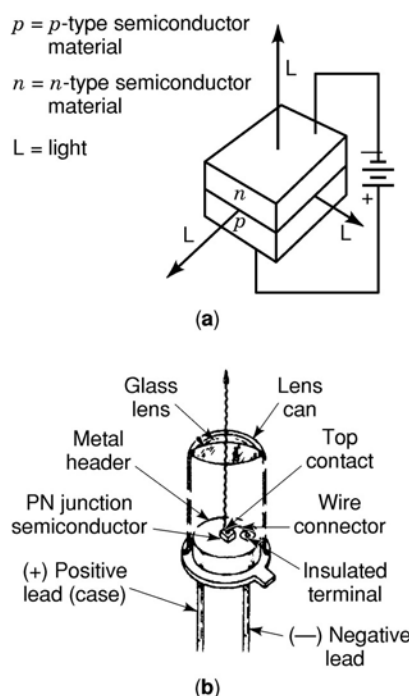


Figure 6. (a) A semiconductor junction. (b) A cross-sectional view of a typical light-emitting diode. (c) Spectral output of several light-emitting diode semiconductor materials (7).

the p - n junction that promote stimulated emission and photomultiplication of photons which are then emitted in a single focused direction. One such laser uses ($\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$) where parallel mirrors perpendicular to the plane of the p - n junction are readily obtained by cleaving along this material's natural cleavage planes. Photons bounce back and forth between these parallel mirrors at opposite ends of the p - n junction causing electrons to recombine with holes and producing more and more photons which are partially transmitted as laser light by one of the mirrors.

Electroluminescent Phosphor Sources

Special phosphors exist that convert alternating current energy directly into visible light radiation purely by electroluminescence. This skips the intermediate step in discharge tubes of first having to create light indirectly to activate the phosphor. An electroluminescent phosphor, such as zinc sulfide or zinc sulfoselenide is placed between the two conducting layers of a plate capacitor. Standard 60 Hz alternating current at 120 V excites electrons within the phosphor that emit visible light radiation upon transition back to their ground state. The color of the visible light produced is controlled by adding different relative concentrations of activators, such as copper, lead, and, manganese. This produces low-luminance lamps for decorative lighting, night lights, switchplates, instrument panels, clock faces, telephone dials, thermometers, and signs. Colors of these lamps are blue, green, yellow, or pink. Rated wattage varies with applied voltage, frequency, and temperature.

Nuclear Sources

A nuclear light source is a self-powered device that uses a radioactive material to create light radiation. One design for such a source uses tritium gas, an isotope of hydrogen, to fill a sealed glass tube whose inner wall is coated with a phosphor. Tritium emits low-energy beta particles (i.e., electrons) which strike the phosphor causing it to fluoresce. This type of fluorescence is the same mechanism by which an electron gun creates a visible image on a television screen. Glass is impervious to the beta radiation so that it does not present a radiation hazard. The half-life of tritium is 12.3 years, although in practice half-intensity of these light sources is reached in about six to seven years, and they have a useful life of about 15 years (7).

INFRARED AND ULTRAVIOLET LIGHT SOURCES

Many of the same design concepts described for visible light sources are also employed as man-made sources of infrared and ultraviolet radiation. Incandescent tungsten filaments, in fact, radiate up to 70% of their power in the infrared region. Most of this is in the near-infrared between 800 and 1100 nm despite the lower emissivity of tungsten at these wavelengths. Tungsten-halogen lamps also have significant output in the 350 nm to 400 nm range. Figure 6(c) shows the spectral output for two types of GaAs light emitting diodes in the near-infrared. Synchrotrons are ideal for producing a broad continuum of radiation throughout the entire ultraviolet region, but the engineering of high-energy electron accelerators makes this practical primarily for research in ultraviolet spectroscopy. Only some commercial applications are becoming feasible.

Discharge lamps are tuned to emit different amounts of electromagnetic radiation in the ultraviolet and in the infrared depending on the type and vapor pressure of gas(es) used. Xenon has intense spectral lines in the range from 780 to 1000 nm. A xenon-discharge lamp is a good near-infrared source together with a continuous infrared emission tapering off at about 2600 nm. A xenon-discharge lamp also has good continuous emission in the 200 nm to 400 nm range. A mercury-discharge lamp is used as a good near-infrared and mid-infrared source with continuous emission up to 2600 nm. As exploited for photoluminescent stimulation in visible light fluorescent sources, mercury discharge lamps are good ultraviolet light sources. In addition to the intense 253.7 nm spectral line, mercury lamps have good continuous emission from 240 to 400 nm. The most commonly used source in absorption spectrometers is the deuterium discharge lamp which has three to five times more power output than the previously used hydrogen-discharge lamp. The deuterium-discharge lamp has good continuous emission from 180 to 400 nm in the ultraviolet. Most photometric light sources for the vacuum ultraviolet region below 200 nm are high-intensity-discharge tubes filled with rare gases including helium, neon, argon, krypton, and xenon. For more detailed information on ultraviolet light sources consult (8,9).

SOURCES OF X RAYS

X rays were first discovered by Wilhelm Conrad Roentgen in 1895 while observing the effect radiation emanating from a high-voltage vacuum tube had on a piece of barium platino-cyanide. Roentgen deduced that invisible radiation was emitted from the vacuum tube causing the barium platinocyanide to fluoresce. Even placing various materials between the vacuum tube and the barium platino-cyanide, such as black paper, wood, and cardboard, did not prevent the fluorescence. Roentgen then observed the bones of his own hand when placing it between the tube and a sheet of cardboard coated with barium platino-cyanide producing the world's first fluoroscopic screen. He called the invisible emissions X radiation before it was soon discovered that this was yet another form of electromagnetic waves only at higher energies.

The basic designs of X-ray sources widely used today for medical imaging are variations of the original *hot cathode tube* invented by W. D. Coolidge in 1913 at the General Electric Company Laboratories. A low-voltage heating circuit operates a hot filament cathode, and the intense heat produces a source of free electrons by thermionic emission. A second circuit creates a large kilovoltage potential difference between the negatively charged cathode and a positively charged anode target. The free electrons are accelerated toward the anode target striking the anode at high velocity, typically exceeding half the speed of light. The cathode, anode and the space between them are contained in a high vacuum to prevent collisions with gas atoms and therefore prevent slowing of electrons. Two physical mechanisms produce X-ray emission, bremsstrahlung (braking) radiation from the sudden deceleration of electrons as they strike the anode target and characteristic radiation resulting from high-speed electrons which displace inner orbital electrons of the atoms composing the anode. The bremsstrahlung radiation is emitted over a continuous and broad wavelength range whereas the characteristic radiation consists of limited discrete spectra.

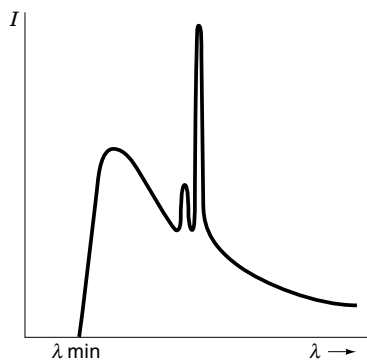


Figure 7. Typical spectrum emitted by an X-ray tube showing the continuous bremsstrahlung radiation spectrum with some characteristic radiation lines superimposed. $\lambda_{\min}$ indicates the short wavelength limit of the spectrum (8).

The broad continuous range of wavelengths for bremsstrahlung radiation is caused by the nonuniform rate at which kinetic energy is lost when each free electron strikes the anode target. The minimum wavelength emitted (i.e., maximum frequency) is given by equating the maximum energy of an electron with charge e accelerating in a potential voltage V to the energy of light at frequency ν , namely,

$$eV = h\nu_{\max}$$

With $\nu\lambda=c$,

$$\lambda_{\min}(\text{angstroms}) = \frac{12400}{V(\text{volts})}$$

A typical spectral distribution emitted from an X-ray tube is shown in Fig. 7 showing strong discrete lines from characteristic radiation against the broad spectrum of bremsstrahlung radiation. For potentials between the cathode and anode exceeding 2000 V, the maximum intensity over the continuous spectrum occurs at $1.5\lambda_{\min}$ and the falloff at longer wavelengths is as $1/\lambda^2$. The most useful range in many radiological applications is 0.1 to 0.5 angstroms. For characteristic radiation with a short wavelength the material composing the target anode must have a high atomic number. Tungsten with an atomic number of 74 and a high melting point of 3655 K is generally used for the target anode. If the incident-free electrons were decelerated along their direction of incidence, classical electromagnetic theory predicts that the intensity maximum of bremsstrahlung radiation occurs at 90° . However, intensity maxima have been observed for different electron energies in the range from 50° to 65° , more consistent with that predicted by quantum mechanics. Good background reading is (10).

BIBLIOGRAPHY

1. M. Born and E. Wolf, *Principles of Optics*, Oxford, UK: Pergamon, 1981.
2. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, New York: McGraw-Hill, 1976.
3. R. Siegel and J. R. Howell, *Thermal Radiation Heat Transfer*, 2nd ed. New York: McGraw-Hill, 1981.
4. G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. New York: Wiley, 1981.

5. K. L. Kelley, Lines of constant correlated color temperature based on MacAdam's (u, v) uniform chromaticity transformation of the CIE diagram, *J. Opt. Soc. Am.*, **53**: 999, 1963.
6. D. B. Judd, Estimation of chromaticity differences and nearest color temperature on the standard 1931 ICI colorimetric coordinate system, *J. Opt. Soc. Am.*, **26**: 421, 1936.
7. *IES Lighting Handbook*, New York: Illuminating Eng. Soc. North Amer., 1993.
8. J. A. R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy*, New York: Wiley, 1967.
9. A. Knowles and C. Burgess, *Practical Absorption Spectrometry*, London: Chapman and Hall, 1984.
10. J. Selman, *The Fundamentals of X-Ray and Radium Physics*, Springfield, IL: Charles C. Thomas, 1971.

LAWRENCE B. WOLFF
The Johns Hopkins University

pH MEASUREMENT AND CONTROL

The acidity or alkalinity of solutions is indicated by a pH value which it is important to control in many chemical and biotechnical processes, especially in water and wastewater treatment. The significance of pH in industrial processes lies in the fact that many phenomena, such as flotation, precipitation, and coagulation, occur in a certain pH range. The inhibition of corrosion is another application of pH control. In industrial applications of biotechnology, such as fermentation, the pH is an essential factor in the growth process.

The neutrality of pure water is an important factor in human life. Municipal water is kept slightly alkaline (about pH 8) to avoid corrosion. The quality of wastewater has a significant influence on environmental water ecology. Changes in the acidity of water influences, for instance, the solubility of compounds and the growth of flora and fauna. The acceptable pH range of wastewater is usually pH 6 to pH 9. An important pH control application is the preparation of wastewater flowing through a sewer system to chemical or biological purification plants. Chemical purification occurs best at a pH of 5.5 in aluminum sulphate precipitation and at a value of 9 in ferrosulfate precipitation. Fast pH changes are unacceptable in wastes flowing in an active sludge purification plant because of the bacteria death risk. For these reasons reliable measurement and good control of pH are very important aspects of a wide range of systems.

GENERAL PROPERTIES OF LIQUIDS

Liquid phase systems consist of a solvent and dissolved components. The most common solvent is water, the properties of which are well known. Most chemical constants related to pH are connected with systems where water is the solvent.

When solid compounds are introduced into water, there is an upper limit of how much substance can be dissolved, called the solubility. When the solvent contains the maximum amount of dissolved component or more (the solid phase contains the excess component), the solution is saturated.

The concentration [.] of a solution is defined as the amount of dissolved substance per the amount of solvent. Concentration

is usually given in moles per liter (sometimes moles per cubic foot, grams per liter, mass or volumetric percentages, etc.).

Compounds that form ionic bonds dissolve into ions in water. A typical example is the dissolution of table salt, sodium chloride:



Many physical and chemical phenomena are related to the activity $\{i\}$ of a given ion. In dilute solutions activity coincides with concentration, but in more concentrated solutions there can be significant differences. In the literature the activity is sometimes called the efficient concentration. The relationship between concentration and activity of a substance i is often assumed to be the following:

$$\{i\} = f_i \cdot [i] \quad (2)$$

The activity coefficient f is assumed to be 1.0 in very dilute solutions. For somewhat more concentrated solutions the activity coefficient can be estimated with the help of ionic strength. For very concentrated solutions calculating f becomes rather complicated.

There are many excellent books in the field of ionic equilibrium and pH systems. The notations used in this article are close to those of Butler (1) and Bates (2). These books are also recommended for a more detailed treatment of the subject.

Chemical Reactions

Most chemical reactions are reversible. There are always two reactions taking place simultaneously, an advancing reaction and a reverse reaction. Consider the general reaction



A reversible reaction consists of an advancing reaction $aA + bB \rightarrow cC + dD$ and a reverse reaction $cC + dD \rightarrow aA + bB$.

A reaction that takes place in a single step is called a unit reaction. For unit reactions the rate of reaction depends on the activities of the participating components. For an advancing reaction the rate of reaction is given by the following:

$$r_1 = k_1 \cdot \{A\}^a \cdot \{B\}^b \quad (4)$$

Correspondingly for the reverse reaction

$$r_{-1} = k_{-1} \cdot \{C\}^c \cdot \{D\}^d \quad (5)$$

The reaction rate constants k_1 and k_{-1} are strongly dependent on temperature. In some cases the relationship can be described by the well-known Arrhenius equation.

At equilibrium the reaction rates of the advancing and the reverse reactions are the same ($r_1 = r_{-1}$), and the overall reaction does not go in either direction:

$$k_1 \cdot \{A\}^a \cdot \{B\}^b = k_{-1} \cdot \{C\}^c \cdot \{D\}^d \quad (6)$$

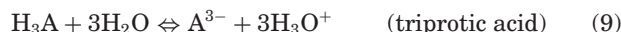
The ratio of the reaction rate constants is called the equilibrium constant K .

$$\frac{k_1}{k_{-1}} = \frac{\{C\}^c \cdot \{D\}^d}{\{A\}^a \cdot \{B\}^b} = K \quad (7)$$

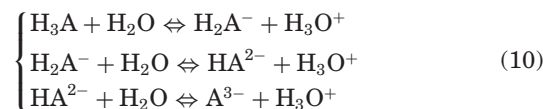
A large value of the equilibrium constant indicates that the overall reaction (3) tends to go to the right toward the reaction products, and a small value that the overall reaction tends to go to the left toward the source materials. Similarly to reaction rate constants, the equilibrium constant also depends on temperature.

Acids and Bases

Acids contain hydrogen that is released under certain conditions. Acids are monoprotic or polyprotic (diprotic, triprotic, etc.) depending on the number of hydrogen ions H^+ that they can donate. The hydrogen ion H^+ forms different compounds with the water molecule, but usually in acid-base systems all of the hydrogen ion combinations (e.g., H_3O^+ , H_9O_4^+) are symbolized by a simple hydrogen ion H^+ or an oxonium ion H_3O^+ :



Acids are generally symbolized with letter A and bases with letter B. Polyprotic acids donate hydrogen ions in several stages, and therefore the latter reaction (9) can be separated into three unit reactions:



Bases accept hydrogen ions and similarly to acids they are also called monoprotic or polyprotic:



The hydrogen ion is actually a proton and therefore in the literature acids are defined as proton donors and bases as proton acceptors. Every acid HA has a corresponding base A^- , and together they form an acid-base pair. A substance that either accepts or donates a proton is called an ampholyte. The most important ampholyte is water (as can be seen from all of the previous examples).

Acid-Base Constants. Equilibrium constants can also be written for acids and bases. For monoprotic acids that react as in Eq. (8), the equilibrium constant is defined as follows:

$$K = \frac{\{\text{A}^-\} \cdot \{\text{H}_3\text{O}^+\}}{\{\text{HA}\} \cdot \{\text{H}_2\text{O}\}} \quad (12)$$

The reaction takes place in an aqueous solution and the activity (and the concentration) of water can be considered a constant

and is combined with the equilibrium constant:

$$K_a = K \cdot \{H_2O\} = \frac{\{A^-\} \cdot \{H_3O^+\}}{\{HA\}} \quad (13)$$

The modified equilibrium constant is called the acid constant. For bases a corresponding base constant is developed. For reaction in Eq. (11) the base constant is given by the relationship

$$K_b = \frac{\{HB^+\} \cdot \{OH^-\}}{\{B\}} \quad (14)$$

Polyprotic acids and bases have equilibrium constants for each dissociation step. The acid constants of a triprotic acid that reacts by the reaction mechanism Eq. (10) are as follows:

$$K_{a1} = \frac{\{H_2A^-\} \cdot \{H_3O^+\}}{\{H_3A\}}, K_{a2} = \frac{\{HA^{2-}\} \cdot \{H_3O^+\}}{\{H_2A^-\}}, \quad (15)$$

$$K_{a3} = \frac{\{A^{3-}\} \cdot \{H_3O^+\}}{\{HA^{2-}\}}$$

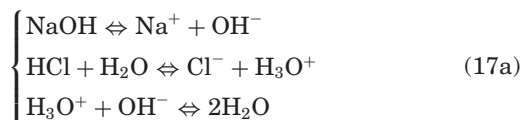
Acid and base constants describe the strengths of acids and bases. A large acid constant indicates a strong acid that donates practically all of its protons. Weak acids have a small acid constant, and only some of the protons that could be released are actually donated.

The equilibria of weak acids and bases can be moved to either side of the reaction by adding or removing reaction products or source materials. A typical method of removing reaction products is neutralizing a solution by adding bases to acids and vice versa. Neutralization is treated in more detail in the next section.

Neutralization. The oxonium ion H_3O^+ (hydrogen ion H^+) and the hydroxide ion OH^- neutralize each other and produce water:



Consider an example where sodium hydroxide and hydrochloric acid are mixed together:

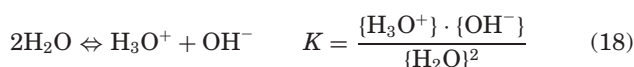


The overall reaction is as follows:



Mixing equal amounts of sodium hydroxide and hydrochloric acid yields table salt and water.

Ionic Product of Water. The neutralization reaction in Eq. (16) is also reversible. The reverse reaction is called autoprotolysis of water:



The equilibrium constant is modified similarly to the acid and base constants:

$$K_w = K \cdot \{H_2O\}^2 = \{H_3O^+\} \cdot \{OH^-\} \quad (19)$$

The modified equilibrium constant is called the ionic product of water [at 25°C $K_w = 1.0 \cdot 10^{-14}$ (mol/L)²].

The Definition of pH. In acidic solutions there are more oxonium ions than hydroxide ions, and vice versa in alkaline solutions. If the system is neutral, there are equal amounts of both ions. The previous section stated that the product of oxonium and hydroxide ion activities (efficient concentrations) is constant.

$$\begin{array}{ll} \{H_3O^+\} > \{OH^-\} & \text{acidic} \\ \{H_3O^+\} = \{OH^-\} = \sqrt{K_w} = 1.0 \cdot 10^{-7}, \text{ (at 25 deg C)} & \text{neutral} \\ \{H_3O^+\} < \{OH^-\} & \text{alkaline} \end{array}$$

For practical reasons the logarithmic scale is more appealing (ten-base logarithm $\log_{10}$ is marked with lg). pH is defined as

$$pH = -\log_{10}\{H_3O^+\} = -\lg\{H_3O^+\} \quad (20)$$

At 25°C the neutrality of a solution is given as follows:

$$\begin{array}{ll} pH < 7 & \text{acidic} \\ pH = -\lg 10^{-7} = 7 & \text{neutral} \\ pH > 7 & \text{alkaline} \end{array}$$

The pH usually varies between 1 and 14, but higher or lower values are also possible.

Weak acids and bases have equal amounts of dissociated and undissociated forms when the oxonium ion activity is equal to the acid or base constant, as can be seen from Eqs. (13)–(15). There are important practical phenomena when this happens (such as maximum buffering, explained in the next section) and therefore the pH which enables that is significant. Often the equilibrium constants K_a and K_b are given in pK_a and pK_b values whose definitions are identical to Eq. (20):

$$pK_a = -\lg K_a, pK_b = -\lg K_b \quad (21)$$

pK_a and pK_b values indicate directly important pH ranges that have practical significance.

Buffer Solutions. In many practical situations the pH should be close to an optimal value. This is typically the case for biochemical processes. The solution should be insensitive to small additions of acids or bases and, if this applies, the solution is called a buffered solution. Many natural systems contain buffers (e.g., blood) but buffers can also be produced artificially from weak acids and bases and their salts.

A typical buffer solution is produced from acetic acid (HAc) and sodium acetate (NaAc). The resulting solution contains both acetic acid and acetic ion (Ac^-). The reaction is as follows:



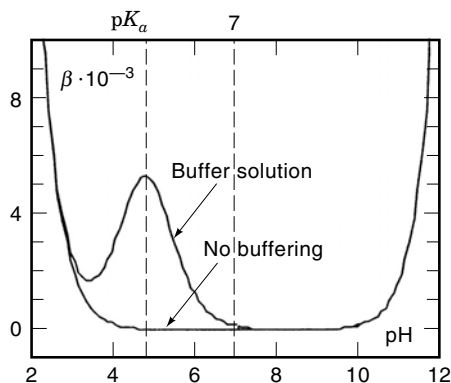


Figure 1. The buffer index of acetic acid solution.

If acid is added, the free hydrogen ion (oxonium ion) reacts with excess acetic ion to form undissociated acetic acid (reverse reaction). The pH is not significantly affected until the buffer solution runs out of acetic ions. Correspondingly, if base is added, the free hydroxide ions react with hydrogen ions to form water. The acetic acid ionizes more (advancing reaction) to satisfy the equilibrium condition. The pH changes only a little until the buffer solution runs out of undissociated acetic acid. The buffer solution can be considered a reservoir of ions that are released when needed in the reaction.

The efficiency of a buffer solution is described by the buffer index β defined as follows:

$$\beta = \frac{d[B]}{dpH} = -\frac{d[A]}{dpH} \quad (23)$$

The differential change in the concentration of a strong base B versus the differential change in the pH (or negative differential change in the concentration of a strong acid A vs. the differential change in the pH) corresponds to the reverse of the titration curve's slope. In Fig. 1, the buffering index of the acetic buffer solution described earlier is presented as a function of pH. The maximum buffering is reached at $pH = pK_a = 4.75$. The curve (buffer solution) is compared to the corresponding curve of a solution (no buffering) that contains only water and strong acids and bases.

Gibbs Free Energy

Deep understanding of the chemical equilibrium can be achieved only by thermodynamics, but that is beyond the scope of this article. The elementary principles rely on the Gibbs free energy defined by the relationship

$$G = H - TS \quad (24)$$

where H is the enthalpy, T is the absolute temperature, and S is the entropy. Any free system is autonomously moving toward small values of G , and it does not stop until it has reached the minimum Gibbs free energy, that is equilibrium.

The practical significance of this fundamental truth is the fact that many parameters and constants related to Gibbs free energy are tabulated or easily calculated enabling the calculation of equilibrium constants in changing situations, such as varying temperatures.

The free energy of any substance X is calculated as follows:

$$G_x = G_x^0 + RT \ln\{X\} \quad (25)$$

where G_x^0 is the standard free energy and R is the gas constant. For Eq. (3) $aA + bB \rightleftharpoons cC + dD$ the change in Gibbs free energy is given by

$$\Delta G = cG_C + dG_D - aG_A - bG_B = G^0 + RT \ln \frac{\{C\}^c \{D\}^d}{\{A\}^a \{B\}^b} \quad (26)$$

ΔG is called the reaction work (work done by the chemical reaction).

Chemical Energy and Electrical Energy

Chemical energy is transformed into electrical energy in a galvanic element consisting of two electrodes in contact with liquid. If a metallic element is introduced into an ionized salt solution containing the same metal ion, there is an equilibrium. The electrons and metal ions remain in different phases close to each other creating an electrical potential.

The potential of the galvanic element is a direct measure of the cell work and correspondingly also of the change in the free energy. The free energy change is calculated as follows:

$$\Delta G = -nFE \quad (27)$$

where n is the number of electrons transferred in the cell reaction, F is the Faraday constant, and E is the cell potential. By combining Eqs. (26) and (27), Nernst's equation results:

$$E = E^0 - \frac{RT}{nF} \ln \frac{\{C\}^c \{D\}^d}{\{A\}^a \{B\}^b} \quad (28)$$

where the standard potential E^0 of the cell is defined as follows:

$$E^0 = -\frac{G^0}{nF} \quad (29)$$

The standard potential of any electrode is determined by combining it with a reference electrode in a galvanic cell. The standard hydrogen electrode (shown in Fig. 2) is arbitrarily chosen as a general reference electrode with a zero potential. The standard hydrogen electrode consists of a platinum plate that is constantly saturated with hydrogen by a continuous hydrogen gas stream.

The standard potentials of electrodes measured against a standard hydrogen electrode are found in numerous tables.

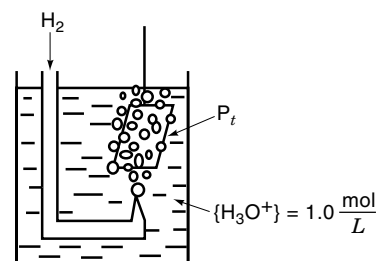


Figure 2. Standard hydrogen electrode.

PERMEABLE GLASS ELECTRODE

Permeable Glass Membrane

The most common pH measuring electrode is the glass electrode. Around 1906 Cremer found that some types of glass gave a potential difference whose magnitude depends on the acidity of the liquid in which the glass is immersed (3). Later Haber and Klemensiewicz proved that this potential difference, within a fixed pH range, follows Nernst's potential in the same way as the hydrogen electrode (4).

Corning Glass Works Inc. developed a well known glass material called Corning 015 consisting of 72.2 mol% SiO₂, 6.4 mol% CaO and 21.4 mol% Na₂O. This glass can be used for pH measurement in the range from 0 to 9. After that glasses containing Li₂O, Cs₂O, BaO, and La₂O₃ were developed. With these glasses more alkaline pH values could be measured.

Glass can be considered an "undercooled" electrolyte consisting of an irregular SiO₂ structure containing other components which move in the interspaces. These components give the membrane an electron balance; see Fig. 3. When immersed in aqueous solution, glasses exchange the metal ions in the glass with the H⁺ ions in the solution. As a result of this reaction a "gel layer" is developed on the surface of the glass membrane. This gel layer is the equivalent of the metal ion layer in Nernst's theory and therefore is essential for the operation of the glass electrode. After one or two days the electrode reaches an equilibrium, and the resulting gel layer has a thickness between 10 nm and 100 nm. This mechanism depends on several factors, such as the composition of the glass and the temperature at which the glass is immersed. Many theories have been developed on the subject of gel layer formation (2).

The electrical potential with the glass electrode is generated by a method somewhat different from that with metal electrodes. In glass electrodes only the ions change location not the electrons.

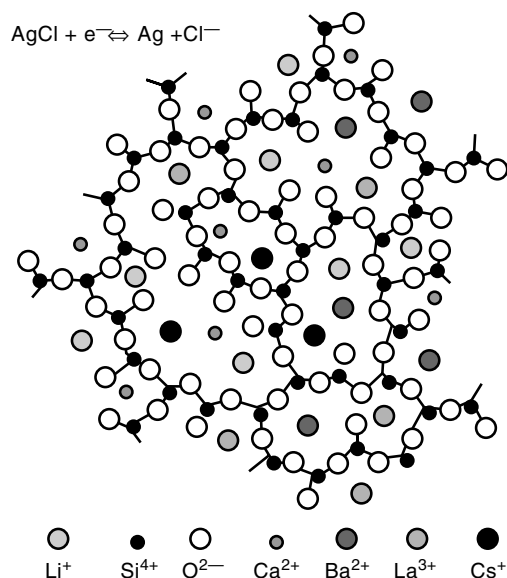


Figure 3. The texture of pH glass.

Construction of the Glass Electrode

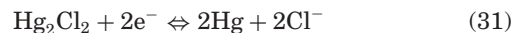
The construction scheme of the glass electrode is shown in Fig. 4. The electrode consists of a glass tube, into which a bulb-shaped membrane of a specific pH glass is welded. The electrode is filled with a buffer solution containing Cl⁻ ions. Usually the pH of this electrolyte is 7. An inner electrode, constructed of a metal pin covered by its sparingly soluble salt, is connected to the output cable of the electrode. The inner electrode is immersed in the electrolyte containing anions of the salt.

A typical inner electrode is the silver-silver chloride electrode, in which AgCl salt covers the Ag pin. When the electrolyte contains Cl⁻ ions, the following electrode reaction occurs:



A potential difference between the inner electrode pin and the electrolyte is formed as a redox reaction Eq. (30); see Fig. 4. The potential difference is calculated from Nernst's equation.

Another common inner electrode is the calomel electrode in which the output cable has a galvanic connection with mercury, which is in contact with mercury chloride (Hg₂Cl₂). This compound is known as calomel. The electrode reaction is as follows:



The potential difference between the inner electrode and the electrolyte is calculated from Nernst's equation. A thallium amalgam-thallium inner electrode is used especially when the temperature of the process liquid to be measured is high. In the electrode thallium amalgam is in contact with thallium chloride.

As mentioned earlier the pH-sensitive membrane is produced from special glass. Its thickness is usually 50 μm to 200 μm, but in measuring very aggressive solutions it can even be 1 mm. After immersion in water, the glass electrode can measure the process solution. A potential difference between the process liquid and the glass surface is created, and this difference is a function of the activity of H₃O⁺ ions and thus also a function of the pH. Figure 5 shows the structure of a typical glass electrode with a calomel inner electrode.

As shown in Fig. 4, the following three potential differences (electromotive forces) are created in the glass electrode:

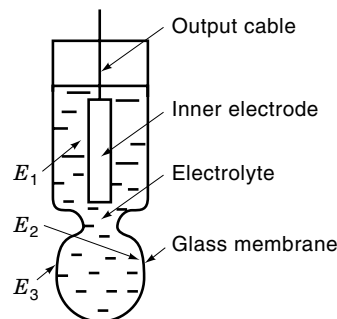


Figure 4. The construction scheme of the glass electrode.

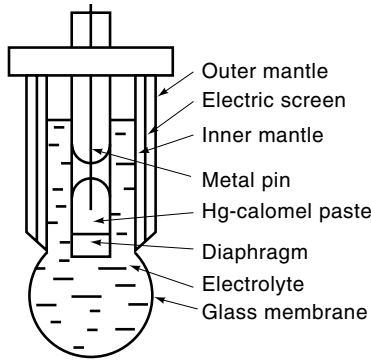


Figure 5. The construction of the glass electrode.

E_1 between the inner electrode and the electrolyte

E_2 between the inner surface of the glass electrode and the electrolyte

E_3 between the outer surface of the glass electrode and the liquid to be measured

As mentioned earlier E_1 is a constant potential difference. The potential difference E_3 is the significant indicator of the pH. Nernst's equation holds quite well:

$$E_3 = E_0 + (RT/F) \ln\{H_3O^+\}_{OUTER} \quad (32)$$

where E_0 is constant at a specific temperature and $\{H_3O^+\}_{OUTER}$ is the activity of the H_3O^+ ions in the process liquid. A pH-sensitive gel layer is also developed on the inner surface of the glass membrane. The potential difference E_2 also obeys Nernst's equation:

$$E_2 = E_0 + (RT/F) \ln\{H_3O^+\}_{INNER} \quad (33)$$

where $\{H_3O^+\}_{INNER}$ is the activity of the H_3O^+ ions in the electrolyte inside the electrode. Because the electrolyte is a buffer solution, the potential difference E_2 is constant. At constant temperature only E_3 is a variable, and it changes when the H_3O^+ content changes in the measured process liquid. By noting that $\ln\{H_3O^+\} = \ln 10 \lg\{H_3O^+\}$ and $pH = -\lg\{H_3O^+\}$, Eq. (32) can be rewritten as

$$E_3 = E_0 - (RT \ln 10 / F) pH_{OUTER} \quad (34)$$

and Eq. (33) as

$$E_2 = E_0 - (RT \ln 10 / F) pH_{INNER} \quad (35)$$

The total potential in the output cable of the electrode is given by

$$E_M = E_1 - E_2 + E_3 \quad (36)$$

Now Eq. (36) can be rewritten as

$$E_M = E_{M0} - (RT \ln 10 / F) pH_{OUTER} \quad (37)$$

where

$$E_{M0} = E_1 + (RT \ln 10 / F) pH_{INNER} \quad (38)$$

Equation (37) shows that the potential of the glass electrode is a function of the pH and the temperature. If the glass electrode potential obeys Eq. (37), it is said to be ideal.

Resistance of the Glass Membrane

The definition of pH requires the potential difference of the measuring and the reference electrodes. If it is measured with a semiconductor converter, a small electric current is created through the electrodes. The current penetrates the pH-sensitive membrane whose resistance is quite high (about 10 MΩ to 500 MΩ at room temperature) and varies as a function of temperature. For instance, the resistance of a membrane produced from Corning 015 glass is about 200 MΩ at room temperature and 1000 MΩ at 10°C.

The high resistance of the glass electrode causes practical problems in constructing measuring circuits. Different kinds of glasses are developed for different measurement conditions. The important considerations are the acidity and temperature ranges of the liquids to be measured. The resistance of the glass membrane, however, should be as small as possible in all cases.

The Asymmetry Potential

The reason for the asymmetry potential is the difference in the behavior of the outer and inner surfaces of the glass membrane. If the inside and outside liquids are identical and there are identical electrodes on both sides, a potential difference still exists between the electrodes. This is called the asymmetry potential. It is only a few millivolts, but it changes slowly during aging of the electrode. Following are the reasons for the asymmetry potential:

Different strains in the inner and outer surfaces of the electrode

During production thermal effects decrease the cation content in the outer surface of the membrane and that is why the penetration of water differs on different sides of the membrane

The outer surface is vulnerable to mechanical and chemical effects

Fouling of the membrane

Alkaline and Acidic Errors

The pH measurement error in strong alkaline liquids is called the alkaline error or sometimes the sodium error. In alkaline solutions above pH 9, other monovalent cations, such as Na^+ and Li^+ influence the electrode potential causing measured values too low. Alkaline error increases with high temperatures. Nowadays special electrodes with a small alkaline error are available. They are well suited for measurements in strong alkaline conditions.

In very concentrated acid solutions the output voltage does not completely obey Nernst's law. Tests show that the gel layer absorbs acid molecules and consequently that the hydrogen ion concentration (and activity) increases, resulting an increase in

pH. Acid error is significant only at very low pH values, and it is relatively small compared to the alkaline error. Acid error does not depend very much on temperature.

REFERENCE ELECTRODE

Purpose

pH measurement requires both a measuring electrode and a reference electrode, which gives the reference potential against which the potential of the measuring electrode is compared. The reference potential has to remain constant when the properties of the process liquid change. On the other hand, the reference electrode builds a galvanic contact from the process liquid to the pH converter.

Silver–Silver Chloride (Ag–AgCl) Electrode

This electrode is named after its inner electrode. The inner electrode has the same structure as the corresponding measuring electrode, that is silver coated with silver chloride, and it is immersed in a KCl solution. The structure of the electrode is shown in Fig. 6. Galvanic contact is created through a diaphragm in the glass mantle of the reference electrode. The diaphragm is produced from ceramics, Teflon, or different kinds of fibers. The KCl solution flows very slowly through the diaphragm, and that is why care must be taken to ensure that there is always enough KCl solution. The pressure inside the electrode has to overcome the pressure of the process liquid at all times. Otherwise the flow through the diaphragm changes direction, and the electrode becomes poisoned.

The concentration of the KCl solution decreases in the diaphragm, and a diffusion potential exists. This potential disturbs the measurement, and its value must be minimized. KCl solution has been chosen as the electrolyte for the particular reason that the sizes and thus the mobilities of K^+ and Cl^- ions are almost equal, and thus the diffusion potential is very small. The usual electrolytes are 1 molal, 3 to 3.5 molal or saturated KCl solutions.

Calomel Electrode

Another commonly used reference electrode is the calomel electrode, also named after its inner electrode. Figure 7 shows two different calomel electrode constructions. Again the inner

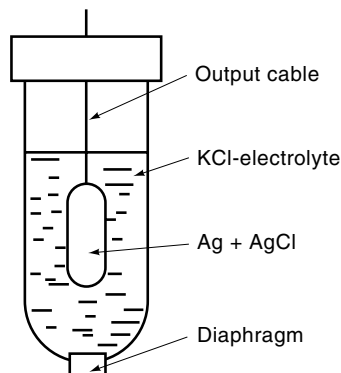


Figure 6. The structure of the Ag–AgCl reference electrode.

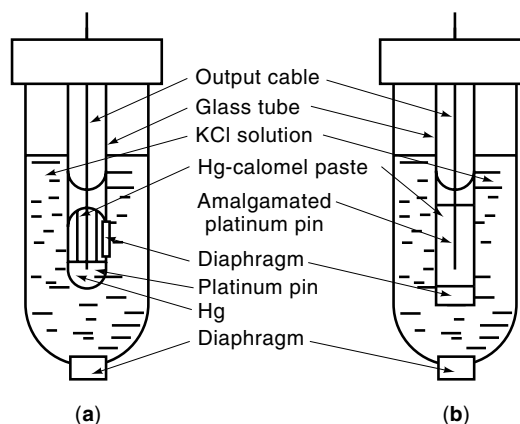


Figure 7. Constructions of the calomel reference electrode.

electrode is immersed in KCl solution. In the construction of Fig. 7(a) the body of the inner electrode consists of a closed glass tube with mercury at the bottom. A platinum pin connected to the output cable is immersed in mercury, which is further connected to a paste prepared from mercury, calomel, and KCl. The electric contact with the KCl solution is realized with a diaphragm on the side of the inner glass tube. Figure 7(b) shows another construction. An amalgamated platinum pin is in contact with Hg-calomel paste, and the electric connection is realized with a diaphragm. In both cases calomel and mercury make up the electrode. The potential difference is the same as in the calomel measuring electrode.

Thallium–Thallium Chloride (Tl–TlCl) Electrode

The Ag–AgCl electrode operates up to temperatures of 100 °C and the calomel electrode up to 80 °C. The Tl–TlCl-electrode can be used for higher temperatures. It measures pH up to 135 °C or even 150 °C. The construction of the electrode is analogous to the other reference electrodes. The potential difference is created according to the following reaction:



Special Constructions

A steady flow of KCl solution out of the electrode is essential. The amount of the flow can be decreased by adding a gel to the KCl solution (5–7). Another technique is to impregnate the KCl solution into a piece of wood (5). Both techniques decrease the possibility of electrode poisoning.

If the process liquid includes compounds which react with the KCl solution, an additional chamber can be constructed between the KCl chamber and the process liquid. It is connected by a diaphragm in both directions and filled with a special electrolyte (6). There is also another approach to avoid KCl-reactions, a nonpermeable reference electrode [SOLPHET™ (26)]. In drossy circumstances the diaphragm must be constructed large enough to assure good functioning. In very drossy circumstances the diaphragm can be replaced by a capillary flow tube (6,8) or a cotton wool layer, (6). Then the consumption of KCl solution is increased, but the fouling effect is avoided. One manufacturer produces a special electrode which controls

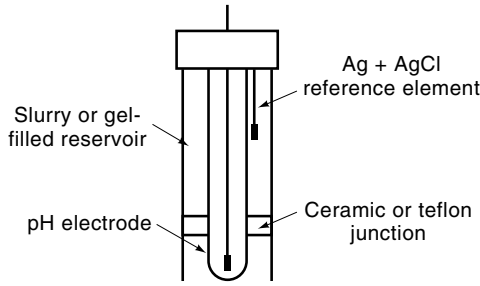


Figure 8. A combined electrode structure.

the pressure of the KCl solution chamber with a bellows connected to the pressure of the process liquid. The elasticity of the bellows maintains the KCl solution chamber at a slightly higher pressure than the process pressure thus avoiding fouling of the diaphragm. Pressures below atmosphere are also possible (6).

COMBINED ELECTRODE CONSTRUCTION

As an alternative to the normal configuration with separated glass and reference electrodes, a system may be used that combines both electrodes into one unit. With this construction the high-ohmic glass electrode is screened by the electrolyte of the reference electrode, and the metal screen can be omitted. The advantage of the combined electrode is the ease of maintenance and the ability to measure pH in very small quantities of liquid. The construction of the joints into the pressurized flow tubes is also simpler. A major disadvantage is that because of the construction of the electrode it is not possible to provide the selection of glass types for many applications. Figure 8 shows the construction of a combined electrode.

MEASUREMENT OF THE pH

The Measurement Circuit

The measurement circuit includes the measuring electrode, the reference electrode, and the signal converter (Fig. 9). The electrodes constitute the voltage source, and if the measurement signal is a standard signal, say 4 mA to 20 mA, the converter is a pH transmitter. The following potential differences exist:

E_3 between the outer surface of the glass electrode and the process liquid

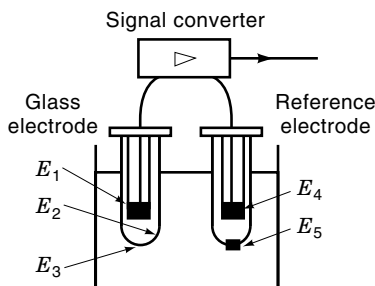


Figure 9. The pH measurement circuit.

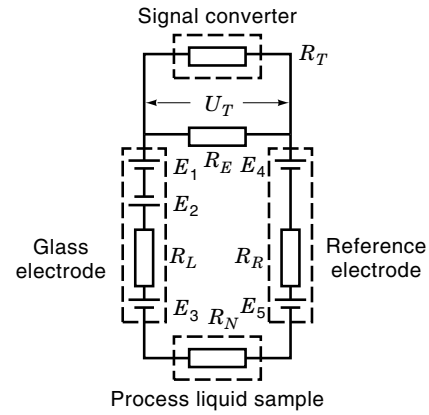


Figure 10. The electric circuit for pH measurement: U_T is the voltage at the pole of the signal converter; R_T is the input resistance of the signal converter; R_E is the isolation resistance between the connecting cables; R_R is the resistance of the reference electrode; R_N is the resistance of the process liquid; and R_L is the resistance of the membrane of the glass electrode.

E_2 between the inner surface of the glass electrode and the electrolyte

E_1 between the inner electrode of the glass electrode and the electrolyte

E_4 between the inner electrode of the reference electrode and the electrolyte

E_5 the diffusion potential of the reference electrode

An electric circuit shown in Fig. 10 is created in the measurement system. The source voltage of the system is the sum of all of the potential differences:

$$E_S = E_1 - E_2 + E_3 - E_4 - E_5 \quad (40)$$

Usually the inner electrodes and the Cl^- -ion activities in the electrolytes of both electrodes are identical, and thus $E_1 = E_4$. In usual constructions E_5 is so small that it can be neglected and thus

$$E_S = E_3 - E_2 \quad (41)$$

The membrane resistance R_L is 10 M Ω to 1000 M Ω , and the resistances R_R and R_N are much smaller and therefore are neglected. The isolation resistance R_E between the connecting cables must be very large. If the decrease of the isolation resistance caused by moisture can be avoided, its value can be assumed insignificant in the circuit. With these assumptions the voltage in the input cables obeys the following equation:

$$U_T = [R_T / (R_L + R_T)] E_S \quad (42)$$

Because the value of the voltage U_T should be as equal as possible to E_S , the input resistance R_T must be very large, about 1 T Ω to 10 T Ω .

Table 1. Values of the Coefficient k

t (°C)	k (mV/pH-unit)
0	54.20
25	59.16
50	64.12
75	69.08
100	74.04

The Effect of Temperature

By substituting Eqs. (34) and (35) in Eq. (41), the following form is obtained:

$$E_S = [(RT \ln 10)/F](\text{pH}_{\text{INNER}} - \text{pH}_{\text{OUTER}}) = k(\text{pH}_{\text{INNER}} - \text{pH}_{\text{OUTER}}) \quad (43)$$

The electrolyte of the measuring electrode is a buffer solution, and thus pH_{INNER} is a constant with a usual value of 7. The coefficient k in Eq. (43) is a function of the absolute temperature T . When the temperature is constant, Eq. (43) defines a straight line with a negative slope. The values of the coefficient k as a function of temperature are listed in Table 1. Figure 11 shows the source voltage as a function of temperature for $\text{pH}_{\text{INNER}} = 7$. The lines intersect at the so-called isopotential point. Ideally this should be $E_S = 0$, $\text{pH}_{\text{INNER}} = 7$ per Eq. (43). In a real situation with existing electrodes, the isopotential voltage differs from zero as shown in Fig. 11. The pH of the intersection point can also differ from 7. The following conditions affect the isopotential voltage:

- Small differences between the inner electrodes
- The asymmetry potential of the glass electrode
- The diffusion potential of the reference electrode
- Fouling of the glass membrane

The temperature dependence of the electrodes can be compensated for by using a Pt 100 or a thermistor immersed in the process liquid. The sensor reads the temperature, and based on this information the signal converter compensates. The temperature sensor can be a third electrode in the conventional assembly, or it can be integrated into the same body as the two other electrodes in a combined electrode.

A more serious problem is caused by the temperature dependence of the pH of the process liquid to be measured. That cannot be compensated for easily because the effect is a function of the liquid composition which is why it is usually omitted.

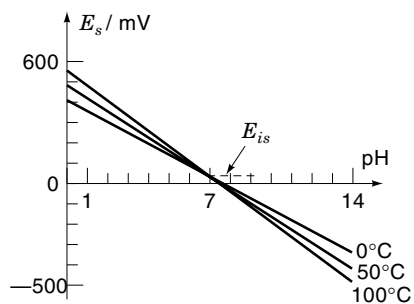


Figure 11. $E_S = f(\text{pH}_{\text{OUTER}})$ of the real electrodes at different temperatures.

Signal Converter

The special properties of the pH measurement circuit establish some requirements for the signal converter or the transmitter, for example high input impedance, temperature compensation, and elimination of electrode ageing. Cheaper transmitters are realized by the analogue technique and the more sophisticated by the microprocessor technique. In both cases high input impedance is easy to realize by semiconductor techniques. Besides the temperature dependence the isopotential voltage must be compensated for. Automatic temperature compensation is included in almost all applications, and only in laboratory devices is compensation realized with a potentiometer. The output signal of the transmitter is usually 0.4 mA to 20 mA in industrial applications. If personal computers are used in on-line control applications, a voltage output signal can also be used, for example 0.1 V to 5 V or 0.2 V to 10 V.

The microprocessor technique improves functioning of the transmitters in many ways:

- The measurement scale can be changed during continuous use
- Advanced diagnostics are available
- The hold function of the signal works during maintenance
- Better possibilities for temperature compensation
- Improved measurement accuracy
- Digital signal transfer is possible

Diagnostics can include the functions of both the transmitters and the electrodes. By measuring the resistance of the glass membrane, the resistance of the reference electrode, the change of the isopotential voltage, and some other properties, conclusions can be drawn as to whether the electrodes are broken, fouled, aged, or not immersed in the process liquid (6,8–10). When the conventional analog transmitting technique is used, the diagnostic measurements necessitate additional wiring. The new fieldbus technique using digital data transmission makes the transmission of the diagnostic data possible without additional wiring, and thus it will be of great advantage in the maintenance of measurements in the future. In special cases when the composition of the process liquid is accurately known, the change of the pH of the liquid as a function of the temperature can also be compensated for. This is the case, for example, in alkaline turbine water including one base and demineralized water (6). When the conductivity of the liquid is very low, about 1–15 $\mu\text{S/cm}$ (in 25 °C), the addition of neutral salt improves the possibility of accurate pH-measurement. If the conductivity is yet lower, less than 1 $\mu\text{S/cm}$ (in 25 °C), the conductivity measurement is a more reliable choice to measure the concentration of the protolyte present in the liquid than pH.

Special Developments

If the process liquid is very drossy or contains glass etching substances (e.g., hydrofluoric acid, HF), an antimony measuring electrode can be used. When the antimony electrode is immersed in the process liquid, a potential is created which obeys Nernst's law. The applicable pH range is only from 2 to 9, but antimony electrode measurement is much faster than that of a glass electrode. A typical time constant (response time from 0 to 63% of the change) in glass electrode measurement is about 5 s. The antimony electrode is easy to clean because of its metallic structure.

To avoid the nonideal features of the conventional measurement-reference electrode pair, a differential pH measurement method was developed (11). This method uses two identical glass electrodes. One of these electrodes measures the process liquid, and the other is immersed in a buffer solution connected to the process liquid via a salt bridge. A metallic reference electrode is also immersed in the process liquid. The resulting source voltage is the difference between these two measuring electrode source voltages and the same reference potential, and thus the actual reference potential is suppressed. Thus fouling of the reference electrode and the small leakage currents in the liquid caused by the inaccuracy of isolation between the reference electrode and the liquid ground are avoided.

Some manufacturers have developed sensors using the semiconductor technology since the year 1970. pH can be measured with an ion-selective field effect transistor, ISFET, which is developed following the construction principle of the conventional MOSFET-transistor. The metal gate electrode has in this construction been replaced by a special oxide layer, which is sensitive to the concentration of ions (in the case of pH-electrodes to hydrogen ions) at its surface. There are many different oxide layer materials for hydrogen ion sensing, for example, silicon dioxide, silicon nitride, aluminium oxide, and tantalum oxide. A conventional reference electrode can be used. This kind of combined pH-electrode can be stored dry for long periods. Because it does not include glass material, it is suited to measurements of food and drugs. The electrode dynamic is faster than that of the glass electrode, and the time constant is less than 3 seconds. The measurement can be done successfully from very small samples and from very dry materials, for example, meat and jelly. Other benefits are the high input impedance making the conventional operational amplifier capable to function as the transmitter and the possibility of cleaning the electrode with a toothbrush. The main disadvantage is the smaller selectivity to hydrogen ions (7,27,28).

Another manufacturer has developed a combined enamel electrode consisting of an enamel measuring electrode and an enamel Na-selective electrode, used as the reference electrode. With this electrode the pH of a dough or paper mass can be measured up to 14% consistency (12). There is also a nonpermeable combined electrode for masses with very low water content [SOLPHET™ (26)]. The combined electrode can be filled with gel in order to avoid fouling of the measuring sensor part in very drossy circumstances. On the other hand, this type of construction, which does not include an air bubble, makes it possible to install the electrode in an upside down position, if needed, (29).

A technique which may be important in the future is optical glass fiber measurement. The color of light shows the pH. The fiber optic sensor can measure the resulting mean value from a large region not only pointwise. This technique is not yet ready for reliable pH measurement, but it is used in a moisture alarm for concrete structures.

PRACTICAL pH MEASUREMENT

Calibration of pH Measurement

Calibration is done with standard buffer solutions. Two or three different buffer solutions are needed. A buffer having a pH near 7 is always used because at that value the theoretical

voltage is about zero. When the measurement range is acidic, a buffer solution with a pH near 4 is used. On the alkaline side, a buffer with a pH near 10 is needed. Alternatively all three buffers are used. The buffer and the process liquid sample must be at the same temperature. The sensors must be rinsed with ion-exchanged water between immersions in the different buffers. The zero potential point is set with buffer 7, and the slope of the line is tuned with the other buffers. Specific electrical calibrators are sometimes used, but they do not take into account the individual differences of the electrodes.

The ionic strength affects the movement of ions in the diaphragm and, thus, the reference potential and pH-value to be measured. For this reason the ionic strength of the buffer solutions should be on the same level as in the sample to be measured. The buffer solutions are usually commercially available in two ionic strengths. The effect of the ionic strength can also appear in the growth of the activity coefficients. Thus liquids having the same hydrogen ion concentration can have different pH-measurement results. The most common buffer solutions with different ionic strengths are presented in Ref. 2.

Practical Definition of pH

Practical definition of pH compares the measured value of the process sample with that of the buffer solution. When these voltages are denoted as E_S and E_{SB} , the definition is constructed by using Eq. (43):

$$pH_{\text{OUTER}} = pH_{\text{BUFFEROUTER}} - F(E_S - E_{SB})/(RT \ln 10) \quad (44)$$

Installation, Use, and Maintenance of Electrodes

In an industrial environment electrodes must be protected against breakage with robust mounting chambers. The most common types, the immersing chamber and the bypass flow chamber, are shown in Fig. 12. The combined electrode assembly shown in Fig. 13 is often used in the pulp and paper industry. In this case the pulp flowing in the pipes can have even 15% consistency. In this case measurement is possible in the thin liquid film formed on the tube wall. In abrasive circumstances an automated sample system must be used. The system automatically takes a sample, measures the pH, ejects the sample, washes the electrodes, and takes a new sample. A cheaper possibility for these kinds of measurement cases is a measurement

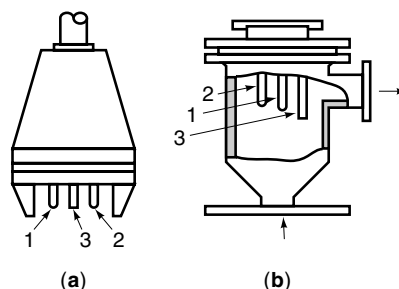


Figure 12. Immersion chamber (a) and by-pass flow chamber (b). 1 = Measuring electrode. 2 = Reference electrode. 3 = Temperature sensor.

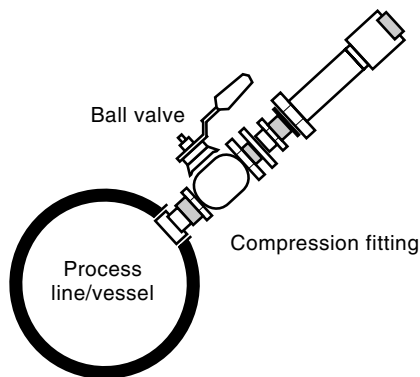


Figure 13. A combined electrode assembly in a tube line.

and maintenance assembly, which is installed on the process vessel wall by welding. The combined glass electrode is in the connection with the process liquid only during the measurement period and the connection is automatically closed when the electrode is moved to the rinsing position by a pneumatic actuator, (30). In the mounting grounding the process liquid is important. In mountings made of steel no problems exist. Otherwise a special ground electrode is often used. The glass electrode is very sensitive to pressure and flow changes in the system and therefore the installation has to be considered with great care. An ideal position is far away from any disturbing equipment, for example, pumps, and there should be straight pipe line before and after the measuring point. If there are severe pressure shocks in the system they should be filtered by using, for example, levelling equipment. A good example of advanced sampling, added pipe delay, and pressure filtering is given in Ref. 31.

Cleaning is important in industrial measurement. It can be automated to function periodically but in more demanding situations manual maintenance is also needed. The following automated cleaning methods are used:

- Mechanical spherical brushing of the sensor head
- Water spraying periodically
- Chemical spraying periodically
- Ultrasonic washing
- Small ceramic particles in a cage around the sensor head

The electrodes can be washed manually with different chemicals depending on the cause of fouling. Sterilization can be done chemically or physically with hot steam if needed. The form of the sensor head is important in fouling. A flat sensor head is suitable for measuring the pH of thick samples in pipes. Other sensor head profiles are useful for other reasons. For example, it is possible to measure very small liquid samples with a spherical sensor head.

The reference electrode consumes KCl solution, and additional solution must be added periodically. Microprocessor-based transmitters with diagnostics help in planning sensor maintenance. When the glass electrode is not used in measurement, it is usually kept in a buffer with pH 4. If it is stored dry, it must be soaked in distilled or ion-exchanged water for at least 24 hours before use.

pH MODELING

pH value is a significant factor in many chemical and biological phenomena. At the same time it is considered a basic measurement in industry and it can be measured on-line whereas many other important measures are still performed off-line. There are many different approaches to pH modeling for better understanding of chemical systems. pH models are divided into static and dynamic models. Modeling methods can be theoretical or experimental.

Static Modeling

A static model is obtained when a system has reached equilibrium and does not change further with time. Acid-base unit reactions can be considered instantaneous, and therefore static modeling is a very natural approach. Static models include titration curves and distribution diagrams. Experimental static models are often basic tools for product quality control and process state indicators. Experimental methods have been used for both qualitative and quantitative analysis.

Titration Curves. Titration curves are traditional pH models. They describe the pH as a function of the difference in the acid-base content. Experimental titration curves usually present the pH as a function of added acid or base volume. Titration curve pH vs. added reagent volume V_A depend on the sample volume, whereas pH vs. added reagent concentration $[HA]-[B]$ does not. The only difference is in scaling. The shape of the titration curve is determined by the participating chemical components. A group of typical titration curves is shown in Fig. 14 where mixtures of acetic acid (HAc), ammonia (NH_3), and water are treated with strong acids and bases. The solutions containing HAc have a titration bump at the pK_a of acetic acid, the solutions containing NH_3 have a titration bump at the pK_b of ammonia, and pure water shows a typical strong acid, strong base titration curve. Polyprotic acids donate more than one hydrogen ion, and the titration curves have correspondingly more than one step as in Fig. 15 which shows the titration curve for phosphoric acid (H_3PO_4).

Theoretical titration curves can be calculated from equilibrium constants and the total concentrations of acids and bases. An experimental curve, on the other hand, requires only a sample. Theoretical titration curves can be generated from the electroneutrality equation (also known as the charge balance equation). The electroneutrality equation takes into account all of

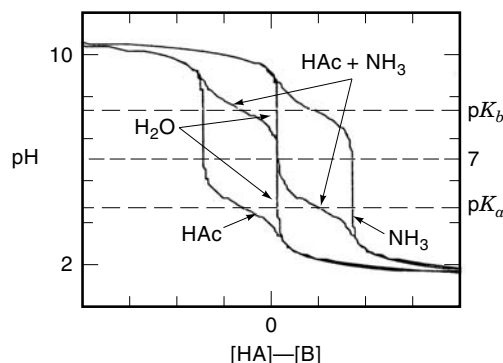


Figure 14. Typical monoprotic titration curves.

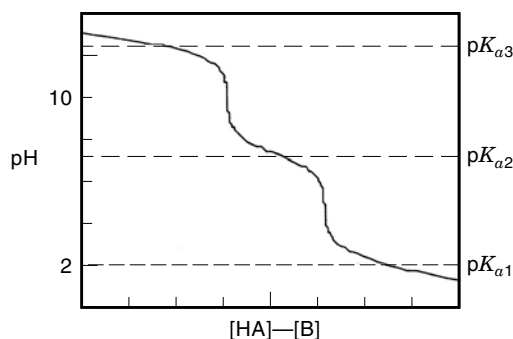


Figure 15. A polyprotic (H_3PO_4) titration curve.

the charged ions of the solution. The solution has no overall charge, and therefore the sum of all individual charges must be zero:

$$\sum_n \pm z_n \cdot [x z_n^\pm] = 0 \quad (45)$$

$z_n^\pm$ is the charge of an individual ion. Ions are usually divided into the components of water (oxonium and hydroxide ions), the components of strong acids (H_2A) and bases (B), and the components of weak acids (H_2a) and bases (b):

$$[\text{H}_3\text{O}^+] - [\text{OH}^-] + \sum_n z_n [\text{H}_{z_n} \text{B} z_n^+] - \sum_n z_n [\text{A} z_n^-] + \sum_n z_n [\text{H}_{z_n} \text{b} z_n^+] - \sum_n z_n [\text{a} z_n^-] = 0 \quad (46)$$

The titration curve shows the relationship between pH (a function of oxonium ion activity) and acid/base concentration. Strong acids and bases dissociate completely, so their ion concentrations are also total concentrations. Weak components dissociate only partially, so their ion concentrations have to be calculated from equilibrium constants. Each weak acid or base ion concentration can be replaced in Eq. (46) by a function of oxonium, hydroxide, and total ion concentrations calculated from equilibrium equations, such as Eqs. (13)–(15). Then the hydroxide ion concentration is replaced by a function of the oxonium ion concentration and the ionic product of water in accordance with Eq. (19).

After all this manipulation Eq. (46) contains only total acid and base concentrations and oxonium ion concentration (from which the pH is calculated by Eq. (20)). A more detailed modeling procedure is given in Ylén and Jutila (13).

Distribution Diagrams. Distribution diagrams give more information than titration curves. They are generated for specific compounds, and they describe the distribution of different forms of a compound as a function of pH. Figure 16 shows the distribution diagram of pure phosphoric acid (H_3PO_4) solution, and Fig. 17 shows the same diagram on a logarithmic scale.

Distribution diagrams are generated from the total balance of a compound. Consider phosphoric acid

$$[\text{P}]_{\text{TOT}} = [\text{H}_3\text{PO}_4] + [\text{H}_2\text{PO}_4^-] + [\text{HPO}_4^{2-}] + [\text{PO}_4^{3-}] \quad (47)$$

Each concentration in Eq. (47) is replaced by a function of equilibrium constants, oxonium ion concentrations, and total phos-

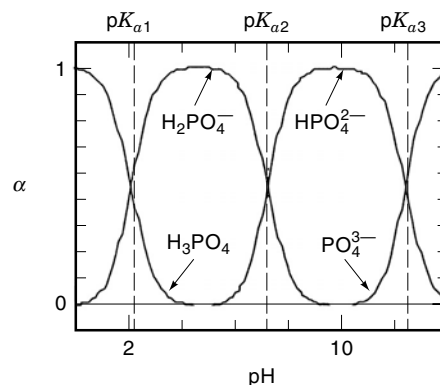


Figure 16. The distribution diagram of H_3PO_4 .

phoric concentrations. The distribution for component X is defined as follows:

$$\alpha_i = \frac{[\text{X}_i]}{[\text{X}]_{\text{TOT}}} \quad (48)$$

Distribution diagrams are useful for evaluating optimal pH ranges for specific reactions, such as separation processes. In separation processes a distribution diagram is often made for different phases of the compound. Figure 18 shows the distribution of calcium in solid and liquid phases of a calcium carbonate solution. A similar distribution diagram could be constructed for carbonate in the same solution. That would be a three-phase diagram (carbon dioxide in the gas phase, calcium carbonate in the solid phase and all carbonates in the liquid phase, consisting of dissolved carbon dioxide, carbonic acid, hydrocarbonate ions, calcium bicarbonate ions, and carbonate ions).

Distribution diagrams for reactions with different compounds do not usually reach a maximum value of one for each fraction. In Fig. 18 the maximum value of the CaCO_3 fraction is determined by the ratio of the total calcium concentration and the total carbonate concentration. There is more calcium than carbonate and thus the carbonate runs out and some of the calcium is left in the liquid phase.

Dynamic Modeling

Dynamic modeling is fundamentally more difficult than static modeling. Static modeling concentrates only on the equilibrium state where the system does not change with time. In dynamic

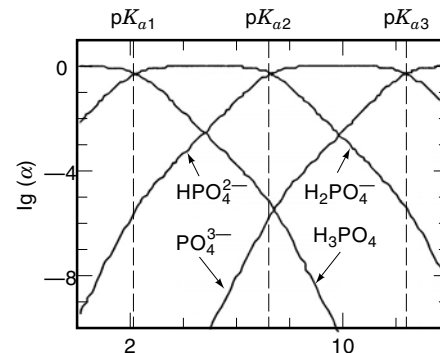


Figure 17. The logarithmic distribution diagram of H_3PO_4 .

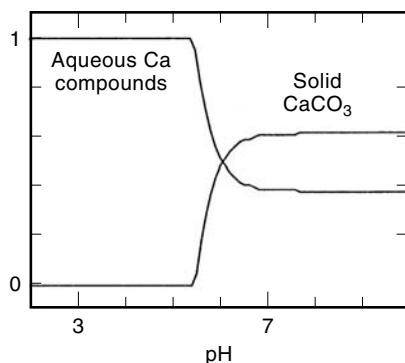


Figure 18. The solid–aqueous distribution diagram of CaCO_3 .

modeling the behavior of the pH and related phenomena are considered as functions of time.

Dynamic pH systems can be divided into two categories, systems where chemical phenomena are significantly faster than flow and mixing phenomena and systems where this is not the case. These categories are treated in more detail in the following sections.

Separate Dynamics for Chemical Reactions and Flow Phenomena. In this simple case the reaction kinetics are much faster than the flow dynamics and can be treated separately. At every instant the chemical reaction takes place instantaneously, and therefore it can be treated as a static nonlinearity. The static models presented in earlier chapters can easily be embedded into flow dynamics. The simplest models are obtained when the weak acid and base component concentrations are constant throughout the system and only the strong acid or base concentrations change. If this applies, the shape of the titration curve does not change. Only the position on the curve changes. In this case the acid–base concentration difference is driven by flow dynamics, and the output is filtered by the titration curve that is presented as pH versus acid–base concentration difference.

The dynamic model is somewhat more complicated if weak components also change. In this case the titration curve of the system can also change arbitrarily and no fixed nonlinearity can be used. The pH value has to be iterated at each step. Two widely used methods are iteration of the aqueous solution's electroneutrality equation or minimization of the Gibbs energy function. Both methods require that each participating chemical component is driven by flow dynamics before the iteration is performed to find the pH.

Interconnected Chemical and Flow Dynamics. Models become even more difficult when chemical kinetics and flow dynamics have approximately the same timescale. In these cases the two phenomena cannot be separated, and the models have to be treated simultaneously. Systems requiring this kind of model are typically multiphase processes or they have slow side reactions that bring the overall chemical reaction rate up to the same level as the flow dynamics. Very often there are slow phenomena in practical systems that slow the otherwise instantaneous chemical kinetics down. The equilibrium assumption has to be modified so that the slow reacting components and their slow kinetics are eliminated from the equilibrium

equations and taken into consideration in the dynamic model. This methodology is described in more detail in Ref. 32.

pH CONTROL APPLICATIONS

Difficulties With pH Control

Difficulties can arise in pH control from the quality of the process liquid or the reagent, the dimensioning of the reactor, pH electrode problems, or because the process liquid has rapid and wide variations in quality and quantity.

Calcium hydroxide [$\text{Ca}(\text{OH})_2$] is a cheap chemical, so it is widely used in neutralization as a base reagent. Because its solubility is quite low, 1.16 g/L at 25 °C, sufficiently large continuous flow reactors are needed to complete the neutralization reactions before the flow reaches the output point of the reactor. The dominating time constant in a stirred tank reactor must be at least 15 min. Calcium hydroxide is supplied to the reactor as a lime slurry, for example 10% by weight. The preparation of the lime slurry calls for large equipment investment. When very large amounts of base are needed, solid lime, quicklime CaO , or even limestone (CaCO_3) can be used as the reagent. In these cases the accuracy of the supplying devices limits the quality of pH control (14). If process liquids including sulfuric acid (H_2SO_4) are neutralized with lime, precipitation of calcium sulfate may cause problems. Sodium hydroxide (NaOH) is an easy-to-handle base, but it is four times as expensive as lime. Sulfuric acid is widely used as an acid reagent. Its use is inhibited by the maximum allowed concentration of sulfates (in wastewater, about 300 to 400 mg/L). The cheapest source of acid reagent, carbon dioxide (CO_2), is flue gas. Its use in the neutralizing alkaline wastes is limited by the high cost of the equipment needed.

When the process liquid is either acidic or alkaline, both acid and base reagents are needed for neutralization. This requires additional investment in supply devices. When acidic and alkaline waste amounts alternate, it is often beneficial to construct an assembly vessel before the neutralization reactor for maximal self-neutralization and to avoid using two reagents; see Fig. 19.

In many basic scientific papers, the large sensitivity of the pH-process is said to be the main problem in the pH-control (15). This is based on the simplest model of the pH-process—the strong acid–strong base model in dilute water solutions. In this theoretical case no buffering components exist, and the very high demand of the rangeability of the reagent supply devices is connected to the enormous sensitivity. When reactors with good mixing include a dead time of 5% to 10% of the dominant

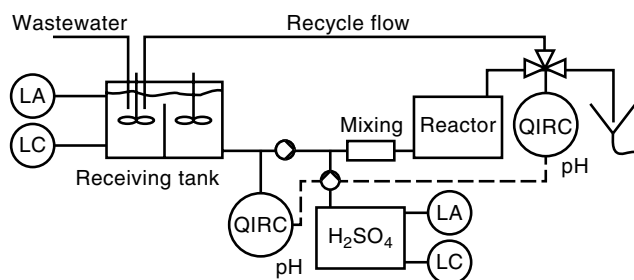


Figure 19. A wastewater neutralization plant in a dairy.

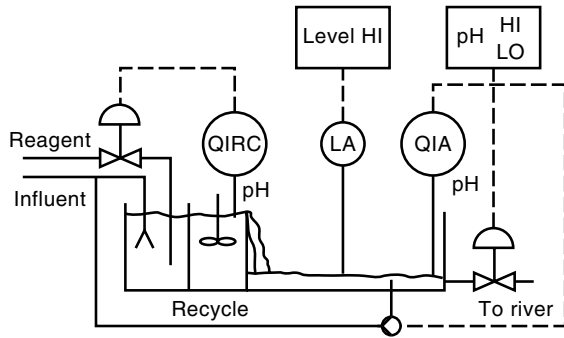


Figure 20. A neutralization reactor with a holding tank and recycle.

time constant and the accuracy of the reagent supply devices is limited while the pH-sensor has a small time constant, the whole task seems impossible. However, the situation is never quite so difficult.

For instance, the alkalisation of turbine water consists of adding strong base to demineralized water. There is always soluted carbon dioxide present, which decreases the demand of reagent addition accuracy considerably. In this case, the pH step is 3 units from ca. 5 to 8. Because of the buffering, the dosage accuracy demand is much less than three decades. In almost every practical control task many additional buffering salts are present, and many decades of the accuracy demand are left away.

In practical sensitive pH-processes, the S-shape of the titration curve is still quite strong, and the sensitivity is high near the neutral point. However, in most cases the high frequency oscillation around the setpoint (usually the neutral point) in the pH-control is less dangerous than long period offsets far from the setpoint. For this reason the authors have the opinion that the main difficulty from the algorithmic point of view is the nonlinear dependence of the pH-value on the reagent amount. In practice, two parallel reagent supply devices are enough to meet the largest demands of the rangeability in the control of the most sensitive pH-processes.

pH Control in Stirred Tank Reactors

Different Algorithm Types. The difficulty of controlling neutralization processes in continuous-flow, stirred, tank reactors

arises from two basic features: the nonlinearity of the process and the changes in this nonlinearity based on the buffering effects caused by varying species. Poor pH control can be improved by adding mixing capacity in series with the neutralization reactor, thus attenuating oscillations that exceed allowable limits. When the process flow is very large, this method is very expensive; see Fig. 20.

If the process sample titration curves are not steep and do not vary in shape (large but constant buffering), the usual PID-feedback can be used for pH control. Even on-off control or ratio control using the throughput flow may be adequate. If the titration curves are quite nonlinear but do not vary much in shape (small but constant buffering), a fixed nonlinear reshaping element in the feedback PID-control can be used. Feedforward control based on a fixed titration curve shape can also be combined with feedback control. When the titration curve shapes change significantly, adaptive controls strategies that take into consideration changes in the process gain on-line give the best results.

Constant Parameter Algorithms. Based on electroneutrality, it can be shown that when the concentrations of weak (buffering) components are constant and only concentrations of the strong component change, all of the titration curves (pH vs. strong reagent concentration) have the same shape. When the concentration of the strong reagent is very high compared to the other concentrations, the pH titration curves versus the reagent flow also have similar shapes. These kinds of titration curves can be used in the constructing the feedback or feedforward-feedback control. Construction of the feedforward control is presented in Fig. 21(a) and (b). The amount of acid reagent with different values of input pH, pH_1 , pH_2 , and pH_3 , can be read from the horizontal axis. The quantity pH_{ref} is the set point value of the pH. Now the reagent demand Q_C can be calculated as $Q_C/Q_P = f_1(pH_{ref}) - f_1(pH_{input})$. The process liquid flow Q_P can change and thus the feedforward control action Q_{CFF} obeys the following equation:

$$Q_{CFF} = [f_1(pH_{REF}) - f_1(pH_{INPUT})]Q_P \quad (49)$$

The function f_1 (the titration curve shape in Fig. 21(b)) is easy to realize with a microprocessor by defining a suitable

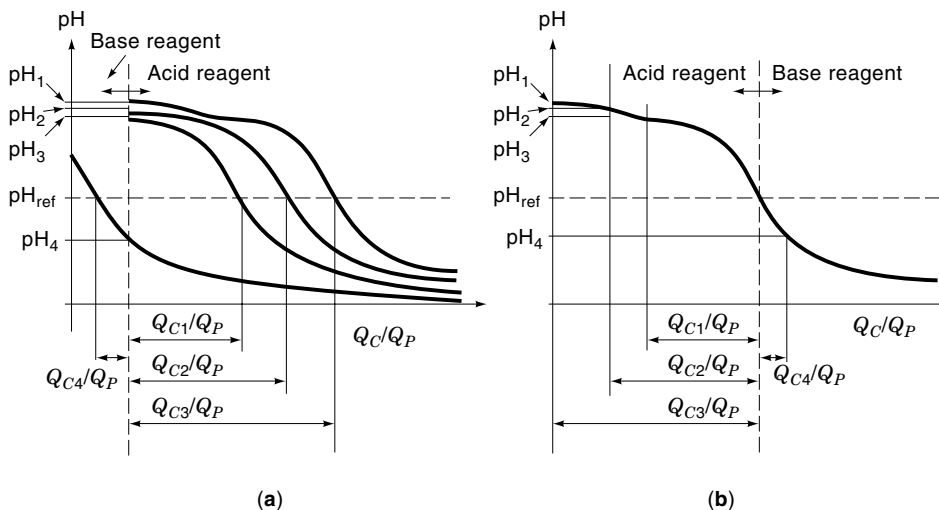


Figure 21. (a) Titration curves with similar shape. (b) Determination of the feedforward control from a titration curve.

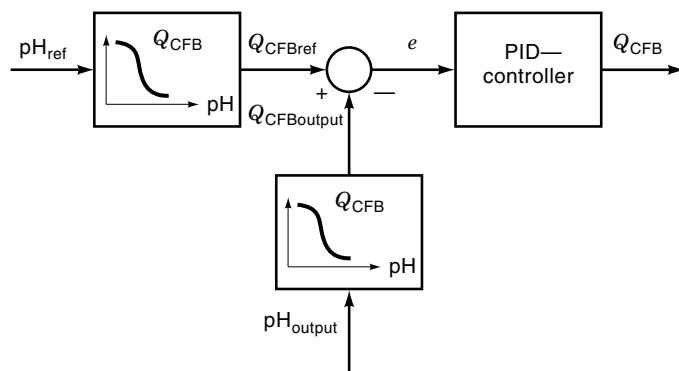


Figure 22. Linearization of the feedback control loop with a constant titration curve shape.

number of points and using interpolation. In constructing the feedback, the same titration curve shape can be used (Fig. 22). If the process flow Q_P changes markedly the feedback PID-algorithm can be flow-adaptive:

$$Q_{CFB} = \frac{Q_P(k)}{Q_{PNM}} K_P \left[e(k) + \frac{T_D Q_{PNM}}{Q_P(k)} \cdot \frac{e(k) - e(k-1)}{\Delta T} \right] + \sum_{i=1}^k \frac{K_P Q_P^2(i)}{T_I Q_{PNM}^2} e(i) \Delta T \quad (50)$$

where K_P , T_I , and T_D are the tuning parameters of the PID-controller, Q_{PNM} is the operating point value of the process flow (the point at which the controller has been tuned), and $Q_P(k)$ is the instantaneous value of the process flow. If the feedforward and feedback algorithms are connected, the final control is given by $Q_C = Q_{CFF} + Q_{CFB}$. If the input pH value cannot be measured, the operating point value can be added to this position PID-algorithm. The final control law is given by $Q_C = Q_{CFB} + Q_P(k) Q_{CFBNM} / Q_{PNM}$, where Q_{CFBNM} is the operating point value of the feedback control Q_{CFB} .

The suitable size for the time constant of the mixing reactor with soluble reagents is about 0.5 min to 3 min. Only in cases of constantly highly buffered process liquids can the neutralization reactor consist of the chamber of an efficient pump or static mixer.

Adaptive Algorithms. When shape of the titration curve changes significantly, feedforward control is not very effective, and consequently adaptive algorithms are needed. Analysis of the past trajectory is the basis for the adaptive feedback control method developed by Welfonder et al. (16). The method is realized with a minicomputer, and it has been tested in an actual wastewater treatment plant. Foxboro Inc. has presented a commercially available adaptive unit controller Exact (17) that also examines the past trajectory (during a waiting period) whether it is oscillating or overdamped. If needed, a new PID-tuning (using a specific tuning algorithm) is used after each waiting period. A narrow deadband around the reference pH value avoids adaptation to noise. This controller has been widely used in pH control, but it is suited only to systems with continuous pH measurement. Subsequently, other instrument manufacturers have made their own versions of this type of controller.

Another adaptive method that does not include any modeling knowledge was developed by Kurz (18). In this method

the adaptation mechanism tries to manipulate the gain of the PID-controller so that the measured pH approaches the reference value along a reference trajectory. This method has been applied in industrial wastewater treatment plants.

The commercially available ASEA Novatune controller has also been applied to pH control. Its function is based on identifying a nonphysical time-series model, and it uses the self-tuning control principle (19). Another method which does not include an exact model is the self-organizing fuzzy controller. It has two fuzzy logic hierarchies. The higher hierarchy adapts the ordinary lower level fuzzy controller to meet changing process conditions. This control method has been applied to the control of an ammonia gas scrubber (13). The self-organizing Kohonen map has also been successfully applied to pH control (20). On the other hand, neural back-propagation algorithms do not apply as well to adaptive pH control.

Kaesser has developed an adaptive feedback control algorithm based on an approximate process model. In his method the slope of the titration curve is estimated and used in tuning the feedback PI-control loop (21). This method has also been applied in neutralizing industrial wastewater. Gustafsson and Waller (22) and Jutila and Visala (23) used more complex physicochemical models closer to practical processes. In these papers model knowledge was also included in designing adaptive control algorithms. The variables to be identified were concentrations of reaction invariants or hypothetical species, and with this knowledge it was possible to construct an instantaneous titration curve and to linearize the PID-control loop. These algorithms were tested by simulation and laboratory-scale pilot plants. Later on Henson and Seborg used the reaction invariant concept in model reference adaptive control algorithms (33).

Lee et al. constructed an algorithm including the physicochemical model of one weak acid. Both the dissociation constant and the concentration of this weak acid were identified and used in the linearizing the control. Only simulation tests were made with a process model differing slightly from the identification model (24). The model reference approach of the pH control made by Pajunen includes a Wiener model (25). Simulation and pilot plant control results are presented in that paper. Many simulation articles have been published, in which advanced control concepts are developed for simple, precisely known processes. Problems may occur in real-life application of such methods, because an accurate model is included in the control algorithm. Model predictive algorithms may turn out to be an interesting choice for pH control, but there are still few actual applications.

Tracking Control

The control of batch fermenters in producing enzymes is a good example of tracking control of pH. Both the growth of the biomass and the production of the enzyme can be controlled with the pH and temperature. It has been shown that keeping these factors constant lowers production compared to the case where these factors change based on optimizing certain growth and production models.

BIBLIOGRAPHY

1. J. N. Butler, *Ionic Equilibrium—A Mathematical Approach*, Menlo Park, California: Addison-Wesley, 1964.

2. R. G. Bates, *Determination of pH, Theory and Practice*, 2nd ed., New York: Wiley, 1973.
3. M. Cremer, On the electromotoric properties of textures, *Z. Biol.*, **47**: 562–608, 1906 (in German).
4. F. Haber and Z. Klemensiewicz, On the electrical phase limit powers, *Z. Physik. Chem.*, **67**: 385–431, 1909 (in German).
5. TBI Bailey Inc., pH/ORP sensors for process monitoring, 1994.
6. Yokogawa Inc., pH in theory and practice, 1997.
7. Honeywell Inc., pH analysis and control solutions, 1997.
8. Hach Inc., Hach systems for electrochemical analysis, 1997.
9. Foxboro Inc., 870/TPH intelligent electrochemical transmitter for pH, ORP, and ion selective measurements, 1997.
10. TBI Bailey Inc., Model TB82 advantage series two-wire pH/ORP transmitter, 1997.
11. Great Lakes Inst. International Inc., What is pH and how is it measured, Technical Handbook for Industry, 1997.
12. Yokogawa Inc., Model pH18 differential pH sensor: general specifications, 1997.
13. J.-P. Ylén and P. Jutila, Fuzzy self-organising pH control of an ammonia scrubber, *Cont. Eng. Practice*, **5**: 1233–1242, 1997.
14. F. G. Shinsky, *pH and pION Control in Process and Waste Streams*, New York: Wiley, 1973.
15. G. K. McMillan, *pH Control*, Instrument Society of America, USA, 1984.
16. E. Welfonder, T. Beyerle, and M. Alt, Continuous neutralization of wastewater using a minicomputer, *Regelungstechnische Praxis*, **20**: 116–154, 1978, (in German).
17. Foxboro Inc., PID algorithm with self-tuning, 1984.
18. H. Kurz, Adaptive control of wastewater neutralization process, *Proc. 9th IFAC World Congr.*, New York: Pergamon, 1985, VI, pp. 3257–3261.
19. Asea Inc., Novatune control system with adaptive controllers, 1985.
20. H. Hyötyniemi and T. Ylöstalo, Modeling and control of a pH process using a self-organizing map, *11th IFAC Symp. Syst. Identification*, 1997, (SYSID'97), Vol. 2, pp. 559–564.
21. R. Kaesser, Adaptive continuous pH control monitored by on-line gain identification, *Proc. 4th IFAC Congs. Instrum. Automation Paper, Rubber, Plastics, Polymerisation Industries*, Oxford: Pergamon, 1980, pp. 285–293.
22. T. K. Gustafsson and K. V. Waller, Dynamic modelling and reaction invariant control of pH, *Chem. Eng. Sci.*, **38**: 389–398, 1983.
23. P. Jutila and A. Visala, Pilot plant testing of an adaptive pH control algorithm based on physicochemical modelling, *Math. Comput. Simulation*, **26**: 523–533, 1984.
24. S. D. Lee, J. Lee, and S. Park, Nonlinear self-tuning regulator for pH systems, *Automatica*, **30**: 1579–1586, 1994.
25. G. Pajunen, Adaptive control of Wiener type nonlinear systems, *Automatica*, **30**: 781–785, 1992.
26. Reagecon Inc., SOLPHET™ sensing technology from Reagecon, 1998.
27. Brylski, pH Measurement with ISFET Sensor, Theory, Comparison & Practical Applications, Sentron publication, 1999.
28. Denver Instruments Company, Electrochemistry Instruments, 1999.
29. Zellweger Analytical Company, Polymetron Compact Probe Model 8370, For online pH-measurements, 1999.
30. Satron Instruments Company, PASVE pH Mounting & Service Valve, August 1999.
31. V. -P. Jutila, P. Jutila, and J. -P. Ylén, The measurement and control of the pH of thick deinked pulp, *Proc. of the 3rd Asia-Pacific Conf. on Control and Measure*. APCCM 98, P. R. China, China Aviation Industry Press, 1998, pp. 326–330.
32. J.-P. Ylén and P. Jutila, Dynamic chemical state modeling in paper and pulp processes, *Proc. of the 4th Asia-Pacific Conf. on Control and Measure*. APCCM 2000, P. R. China, China Aviation Industry Press, 2000, pp. 27–34.
33. M. A. Henson and D. E. Seborg, Nonlinear adaptive control of a neutralization process, *Proc. of the American Control Conf.*, Chicago, 1992, IV, pp. 2586–2590.

JEAN-PETER YLÉN

PENTTI JUTILA

Helsinki University of Technology

PULSE HEIGHT ANALYZERS

The pulse height analyzer (PHA) was originally developed to measure the amplitude of pulses from nuclear radiation detectors. It is therefore an important part of most systems that are used for analyzing event energies from alpha, beta, gamma, and X rays, for analyzing the time between events (from a time-to-pulse-height converter), and for ensuring orderly storage of count information in sequential time periods. PHAs are ideally suitable in system configurations for experimental and production applications in nuclear, chemical, medical, and materials analyses. PHAs are an outgrowth of the single-channel analyzer, which counted amplified electrical signals of a given pulse height created by nuclear events occurring within a detector. The amplitude was accepted by a peak voltage within a “window” or channel defined by two discriminators—referred to as a differential discriminator. Some of the first PHAs stacked 10 or 20 of these discriminators to form a multichannel PHA (1). (Reference 1 is one of the first articles addressing the need for PHAs that process fast pulses required in nuclear spectroscopy.) These early PHAs were commonly referred to as kick-sorters. By 1956 Schumann and McMahon (2) had developed a 256-channel PHA by using a Wilkinson analog-to-digital converter (ADC) and a random access magnetic core memory. This PHA was very successful and led to the first commercial design of this type (manufactured by RIDL—Radiation Instrumentation Development Lab).

PULSE PROCESSING WITH THE PHA

The input signals to the PHA must be optimally processed to achieve the best signal-to-noise ratio (3). Pulse-processing electronics convert the burst of charge, produced when a radiation particle or photon is absorbed by the detector, to a shaped pulse (usually Gaussian). The quantity of charge produced is related to the energy of the incident radiation. A charge-sensitive preamplifier/voltage amplifier combination serves to amplify and shape the charge into a pulse with an amplitude (peak voltage) proportional to this charge, hence to the energy lost in the detector. The pulses, which are random in both time and amplitude, are fed to an ADC. The pulses are digitized and recorded in an address register and their number stored in a memory. The ADC produces this number, informs the memory, and the memory selects the address and tallies one count. This process occurs each time an event is recorded. The pulse from the detector, created by each random event, is digitized with one count added to the channel corresponding to the digitized value. This is the basis of the PHA and pulse height analysis. This process really lends itself to a computer, therefore,

special-purpose computer systems that provide a variety of functions for the acquisition, storage, display, and interpretation of information from random events are referred to as multichannel analyzers (MCAs). The terms PHA and MCA are often used interchangeably; however, MCA is a broader term signifying a computer system that can be configured in a number of ways (including operation in the PHA mode). The term spectrometer refers to a detector, pulse conditioning electronics, and the MCA that is capable of rendering a spectrum describing the detected source and having the capability of performing analyses.

In the early 1960s, ADCs (e.g., the 4096 channel successive approximation ADC) (4) were used with Digital Equipment's (DEC) minicomputer, the PDP-1 (5) and later the more familiar PDP-8 (6). Pulse-height analyzer technology and computer technology merged in a major way during the 1970s. With the advent of DEC's minicomputer, the PDP-11, and later DEC's microcomputer, the LSI-11, data acquisition and analysis systems using MCAs were well on their way. The earliest MCAs came shortly after the development of the ADC but did not become sophisticated until the 1970s. The major components of the MCA are the CPU, memory, control program, acquisition interface, display controller, keyboard, display, and input/output (I/O) interface. Many early designs and references are given in the text by Robert Chase (7).

PRINCIPLES OF OPERATION

A block diagram of a typical firmware-controlled MCA system is shown in Fig. 1. In the PHA mode the ADC provides a channel address proportional to the input pulse amplitude and transfers the digitized peak voltage information to the channel address register. The count information in that channel is placed in the data register and updated by one count. The updated data are then placed in memory. The CRT display shows that a pulse has been processed by flashing the channel (horizontal) and

count (vertical) information in the appropriate location. The time required by the ADC to digitize the pulse height is called the conversion time and the time required to store the data in memory is called the memory cycle time. A typical gamma-ray energy spectrum is shown in Fig. 2.

The ADC

The ADC must be capable of processing the pulses rapidly without erroneous measurements in the pulse height. The Wilkinson method (8) has been the most popular over the years because of its excellent combination of linearity, temperature stability, and speed. Successive approximation ADCs are fast (typical conversion time of about $1 \mu\text{s}$), however, steps must be taken to improve the differential nonlinearity (see discussion below). The fastest conversion time is accomplished today with the Flash ADC (typically 200 ps for a 256-channel unit). This design is similar to a stack of single-channel analyzers but also requires steps to improve the linearity. One of the fastest PHAs reported in the early 1980s used a Flash ADC with a postexperiment, differential nonlinearity correction scheme (9). This method used a bin correction table and had the capability of histogramming the data at 15 MHz.

One of the most important specifications of an ADC (and likewise the PHA) is its linearity. There are two terms used in describing linearity or, more precisely stated, its nonlinearity: (1) integral and (2) differential.

Integral nonlinearity can be described as the maximum deviation of any address from its nominal position described by a straight-line plot of address versus input pulse amplitude, and is shown graphically in Fig. 3. The address with the maximum deviation is identified by N ; the pulse amplitude that is converted to the address N is identified as V_N ; and the nominal voltage corresponding to N is V_{nom} . The integral nonlinearity is then given by the expression

$$\% \text{ integral nonlinearity} = (V_{\text{nom}} - V_N / V_{\text{max}}) \times 100 \quad (1)$$

where V_{max} is the maximum value of the input voltage.

Since the slope of the straight line has a zero intercept of voltage and address this definition may be more conservative than definitions that define the slope and the intercept by the best fit to actual values. Integral nonlinearities should be less than 0.05% in an ADC so that the centroid position of spectral peaks will not be significantly displaced.

Differential nonlinearity describes the nonuniformity of address widths over the entire number of addresses in the ADC. Therefore, differential nonlinearity is the percent deviation of the maximum and minimum address widths (W_{max} and W_{min}) to the average width (W_{avg}) of all addresses, as shown in Eq. (2).

$$\% \text{ differential nonlinearity} = \frac{1}{2} \times (W_{\text{max}} - W_{\text{min}}) / W_{\text{avg}} \times 100 \quad (2)$$

The factor $\frac{1}{2}$ is used in Eq. (2) so that the differential nonlinearity may be expressed as a plus or minus deviation. For good performance differential nonlinearity should be less than $\pm 1\%$ over the top 99% of full scale.

Steps to improve linearity began in the 1950s with improvements by a randomizing method in the 1960s and 1970s

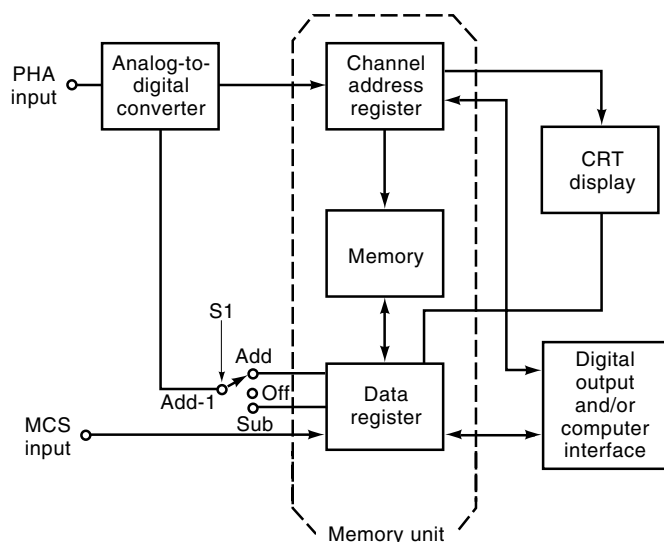


Figure 1. This simplified MCA block diagram illustrates the use of an analog-to-digital converter to digitize the analog signal from PHA to generate the spectrum channel number. The spectrum is displayed on the CRT or transferred to the computer.

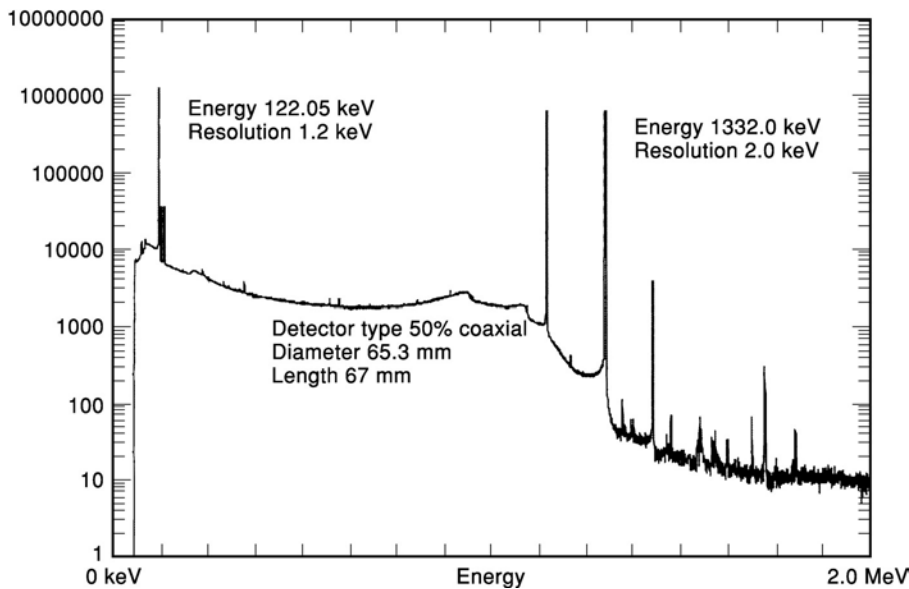


Figure 2. Cobalt 60 (^{60}Co) spectrum, including background radiation, taken with a 50% relative efficiency COAX high purity germanium (HPGe) detector. The high energy resolution typical of HPGe detectors allows for accurate identification of individual energy peaks.

(10,11). Today there are linearization and integrated circuitry techniques that allow performance, economy, and small size. Further discussion on improvement of linearity is presented by Knoll (12).

A typical ADC operation (Wilkinson) is briefly described using the timing diagram of Fig. 4. An input pulse, with an amplitude exceeding the lower-level discriminator threshold, charges a capacitor with a constant-current source. The capacitor voltage is directly proportional to the pulse height as shown in Fig. 4. When the peak of the pulse is sensed the capacitor is discharged by a constant-current. The time required to discharge the capacitor to the baseline, called the rundown time, is proportional to the pulse height. When the discharge is begun a gate is opened to allow the address scaler to count clock pulses from an oscillator (typically 100 MHz clock). The number of clock pulses counted during the rundown time represents the desired address or channel. At the end of the rundown time the address in the scaler is strobed into the memory address

register. An input gate prevents the acceptance of additional input pulses from the start of rundown until the address transfer is complete.

The resolution of the ADC is typically equivalent to the conversion gain which is the number of channels over which the full input voltage will be spread. For example, a 12 bit ADC is normally subdivided into 4096 channels for an input voltage range of 0 V to 10 V. In this example the resolution and conversion gain is 4096. The ADC busy time or dead time is shown in the timing diagram as the total processing time of the input pulse. These pulses can provide a percent dead-time indication and therefore a live-time determination of the MCA. The dead time of the MCA has two components: (1) the processing time of the ADC and (2) the memory storage time. A clock in the MCA that records elapsed time of an analysis is gated off by the busy signal so that the analysis time will not include the time that is spent processing pulses. Accurate dead-time correction (providing accurate live-time) is very important for the analysis of some data. This subject is given further detail by Knoll (13).

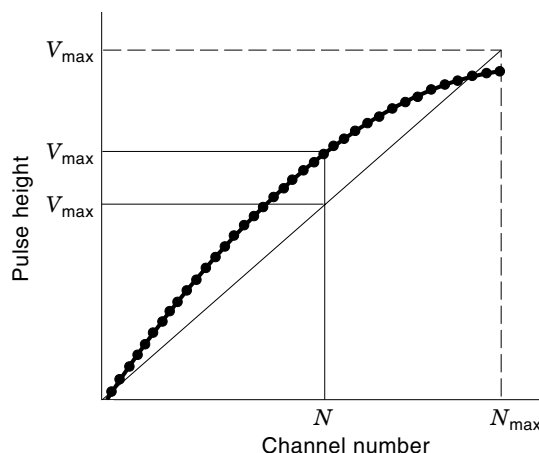


Figure 3. Integral nonlinearity spectral distortion function. Note the significant difference between the idea and actual performance. Integral nonlinearity results in spectral distortion and energy calibration and spectral analysis problems.

The Computer-Controlled MCA

The strong growth of the personal computer has had a major impact on the MCA. Many of the functions of a traditional MCA can be performed as well as or better by a PC. These include spectral display, data analysis, and control functions. Some features of an MCA are unique, including the need for high-speed conversion and excellent linearity. By coupling the unique features of an MCA to a PC, the power of MCAs has dramatically increased as the power of PCs has increased.

The block diagram of a typical PC-based MCA is shown in Fig. 5. A microprocessor controls the ADC functions and communicates with the PC. The microprocessor is controlled by an internal program memory stored in firmware. The output of the ADC is stored in the data memory, which can be accessed both by the ADC and the PC. A modern data memory can store as many as $2^{32}-1$ counts per channel, or over two billion counts per channel. Commands and data are sent from and to the PC using a variety of protocols including serial port, parallel port, dedicated link and, most recently, via Ethernet. Mass storage

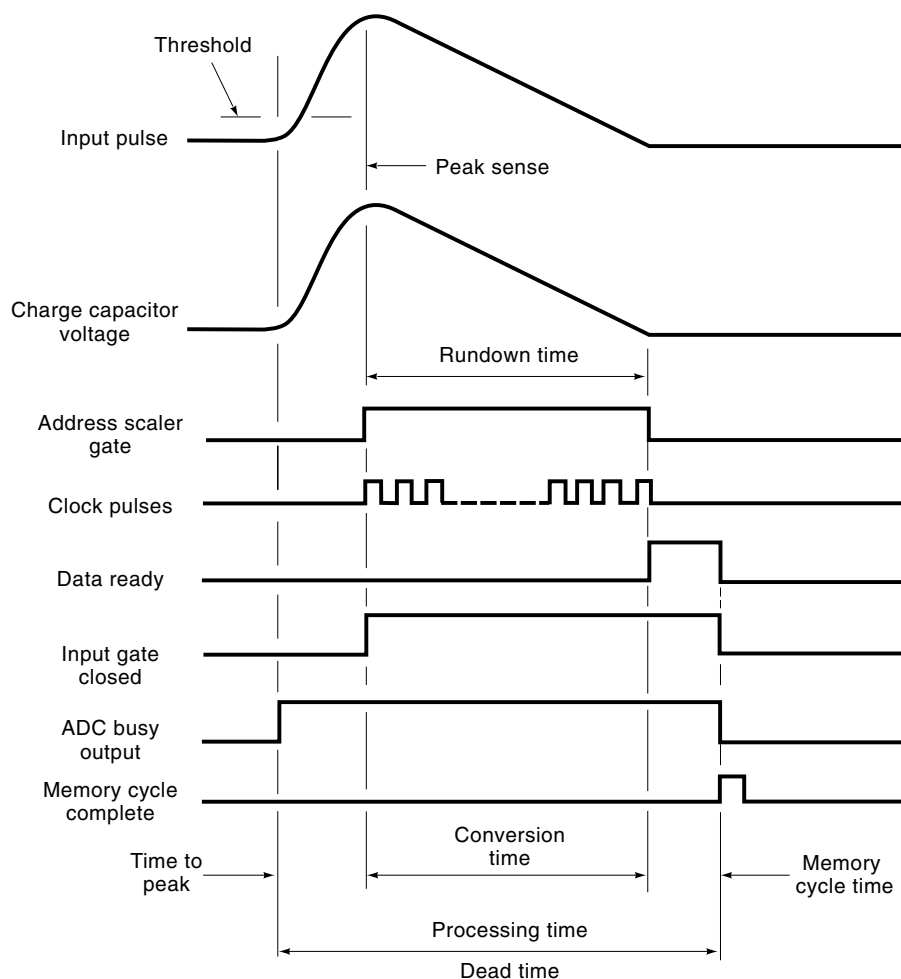


Figure 4. Typical ADC timing sequence. The charge capacitor follows the input pulse until the peak is sensed. The capacitor then discharges during a clocked period to determine the peak amplitude.

of the spectrum files is controlled from the PC using either its internal hard drive or floppy disks. One advantage of the PC-based MCA is that data can be collected and stored in the data memory while the PC is used to analyze a previously collected spectrum.

The hardware of a PC-based MCA takes many forms. One popular format is a plug-in card containing the ADC, microprocessor, data memory and control memory. Communication with the PC is via its internal bus. Other formats include NIM

modules, especially for very-high-performance systems. Stand-alone boxes are also available, which combine the MCA function with other parts of the spectroscopy chain including the spectroscopy amplifier and detector bias supply. Many modern PC-based MCAs are available with this general architecture. Other useful variations include the segmentation of the data memory. This allows spectra to be stored from multiple ADCs and is especially useful in alpha spectroscopy, where there are often many detectors requiring relatively low conversion gains (64 to 1024 channels each). Another variation is to precede the ADC with an analog FIFO (first-in first-out) memory to enhance the data conversion rate.

Preset controls are used during PHA operations to terminate an accumulation period automatically. This provides a convenient method of normalizing the data collection of individual accumulation periods. Additionally, one or more sections of the spectrum, each defined as a region of interest (ROI), can be analyzed for specific information. Some common presets include real time, live time, ROI peak count, ROI integral count, data overflow, and required statistical accuracy on a peak net area.

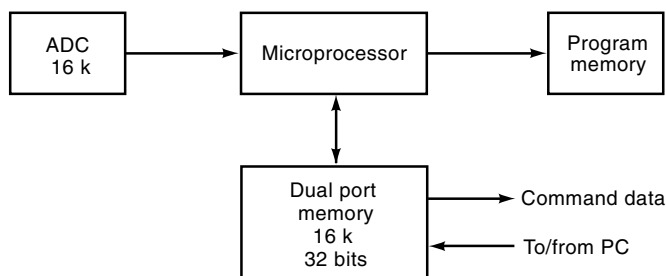


Figure 5. A typical PC-based MCA architecture requires an ADC of appropriate bit length (resolution dependent), a microprocessor, sufficient program memory to store the MCA program, and data memory to store the acquired spectrum.

Integrated Spectrometers

A recent trend has been the integration of all components of a spectrometer into a single package. In addition

to the ADC program, and data memory, these integrated spectrometers include a spectroscopy amplifier and detector bias supply. These integrated spectrometers can be general purpose and support a variety of detectors including germanium, silicon, and NaI(Tl). Alternately, they may be designed for a single detector such as NaI(Tl) to reduce cost and size. Some systems include a spectrum stabilizer and additional data memory to store multiple spectra. Spectrum stabilization is a useful feature when making long term, high-precision measurements with detectors that exhibit characteristic energy peak drift over time. Peak drift can be the result of various factors including temperature fluctuations and component aging. Stabilization is a process by which the system gain and a user-defined energy peak in the spectrum are continuously monitored for channel location. The system gain is adjusted if the energy peak varies from the user-defined channel location. Many implementations of spectrum stabilization include the permanent insertion of a weak, non-interfering radioactive source (typically ^{241}Am in the case of NaI) to provide a continuous, stable, and reliable calibration peak. All major functions of the spectrometer are set via the computer including the amplifier gain, shaping time, and pole-zero; bias supply control; and spectrum stabilizer parameters.

A special form of the integrated spectrometer is the portable spectrometer. These portable versions can be as powerful as their line-powered counterparts, except that they are battery powered. The world-wide need to control special nuclear material has been a strong driver in the development of portable systems. A high-end system will feature a 16 k channel ADC with $<10\ \mu\text{s}$ conversion time, an amplifier with multiple shaping times, bias supply to support both solid-state and scintillation detectors, data memory to store multiple spectra, multiple computer interfaces, and operate on internal batteries for over eight hours. The portable spectrometers are most often operated with the new portable PCs to form a fully portable spectrometer. Because of requirements for light weight, small size, and long battery life, some portable spectrometers are specialized and support a limited number of detectors and applications. Examples include systems developed for the International Atomic Energy Agency.

Future Trends

The future power of the MCA is linked to the future power of the PC. Indications are that PCs will continue to enhance their performance in speed, size, and computational capability, resulting in improved MCA performance. The uniquely nuclear portion of the MCA will also show improvement as integrated circuits continue to become more complex and dense, resulting in increased speed and reduced size. Portable spectrometers should benefit from advances in battery technology driven by other industries such as cellular phones.

Perhaps the most exciting prospect for the future of MCAs is the advent of digital signal processing, or DSP. In a DSP-based MCA, the signal from the detector is converted to a digital word early in the signal processing. Digital filters replace the traditional analog filters and are capable of optimum filter functions that are not realizable with analog circuits. DSP-based MCAs can better match the signal processing electronics to the charge-collection properties of the detector. System throughput and resolution are both enhanced as well as temperature

stability. The first DSP-based integrated spectrometers show this promise (14).

DSP technology offers additional promise. The shape of the signal from the detector contains information about the type and location of the interaction in the detector. As DSP-based systems improve, it may be possible to extract this shape information from the detector signal to improve the overall performance of the spectrometer.

ACKNOWLEDGEMENT

The authors would like to thank Gary T. Mattesich for his exceptional assistance and generous contributions of time and expertise.

BIBLIOGRAPHY

1. G. G. Kelley, *Nucleonics*, **10** (4): 34, 1952.
2. R. Schumann and J. McMahon, Argonne 256 channel pulse-height-analyzer, *Rev. Sci. Instr.*, **27**: 675, 1956.
3. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., New York: Wiley, 1989, pp. 607–610.
4. C. Cottini et al., *Nuc. Instrum. Methods*, **24**: 1963.
5. J. Leng, Instrumentation Techniques in Nuclear Pulse Analysis, Report No. 40, National Academy of Sciences, 1963, p. 23.
6. G. T. Mattesich, *CUBE Symp. Proc.*, CONF-741001, 1974, pp. 188–194.
7. R. Chase, *Nuclear Pulse Spectroscopy*, New York: McGraw-Hill, 1961, p. 74.
8. D. H. Wilkinson, *Proc. Cambridge Philos. Soc.*, **46**: Pt. 3, 1950, p. 508.
9. H. P. Spracklen, 7 MHz particle identifying PHA, *IEEE Trans. Nucl. Sci.*, **NS-29**: 896–899, 1982.
10. S. G. Gobbur, D. A. Landis, and F. S. Goulding, Fast successive approximation analog-to-digital converter, *Nucl. Instrum. Methods*, **140**: 405, 1977.
11. C. Cottini, E. Gatti, and V. Svelto, *Nucl. Instrum. Methods*, **24**: 241, 1963.
12. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., New York: Wiley, 1989, pp. 665–666.
13. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., New York: Wiley, 1989, pp. 669–670.
14. EG&G ORTEC Model DSPec Digital Gamma-Ray Spectrometer.

JAMES H. McQUAID
ANTHONY D. LAVIETES
WAYNE D. RUHTER
Lawrence Livermore National
Laboratory
Thomas J. Paulus
EG&G ORTEC

SCINTILLATION COUNTERS

Scintillation counters employ a phenomenon known as luminescence: a particle passing through a medium excites or ionizes atoms along its path, and, upon deexciting, these atoms emit photons. Such a medium, whether solid, liquid, or gaseous,

which has the property of emitting luminous light, is called a scintillator.

HISTORY OF INSTRUMENTS

Sir William Crookes built the first scintillation counter in 1903, employing a microscope for visual observation of light flashes from a zinc sulfide screen. The physiological limit of this tiring procedure is a counting rate of about one per second (1). Nevertheless, this setup was used in 1909 by Geiger and Marsden (2) to study the scattering of α particles and resulted in Rutherford's discovery of the atomic nucleus (3). Visual scintillation counters were used up to the 1930s. In 1941 Krebs (4) invented the first electrical counter for scintillation events in the form of a Geiger-Müller tube sensitive to luminous photons. This device had a response time of about 10^{-4} s, which is typical for gas counters. The application of the scintillation method greatly increased in the 1950s after the development of the photomultiplier in 1944 by Curran and Baker (5), which enabled the automation of the tedious manual counting and at the same time allowed much higher counting rates. Another advantage over the previously used zinc sulfide screen came in 1947 and 1948 with the discovery of luminescent materials that are transparent to their own scintillation light, such as naphthalene by Kallmann (6), anthracene by Bell (7), and thallium-activated crystalline sodium iodide by Hofstadter (8). It was found that, due to their transparency, these materials can be used as scintillators in large volumes to detect even the only weakly interacting γ rays. Moreover, naphthalene and anthracene exhibit pulse decay times on the order of 10^{-8} s and hence, in conjunction with a photomultiplier, allowed the construction of detectors that were much faster than any gas-filled counter.

APPLICATIONS

Today, scintillation counters are widely used for the detection of α and β particles, γ rays, and neutrons in many fields, such as in high-energy physics research, environmental radiation protection, dosimetry, and medicine. They can be operated (1) as pulse counters to measure the activity of a radioactive source or the dose received by food items or biological tissue, (2) as spectrometers to determine the energy of the incident particles mentioned, and (3) as timing devices for the investigation of cosmic rays or for lifetime measurements on unstable particles.

Examples of applications are monitoring the exhalation of the radioactive noble gas radon from soils and rocks and examining beam profiles in accelerators. Scintillating materials are also used as screen coatings—for example, in television picture tubes.

When compared to other counter types, scintillation counters show the highest counting efficiency, because of their large volumes and their high absorption. Scintillation counters are also much faster than gas counters and semiconductor counters and can handle counting rates of up to 10^8 s $^{-1}$. However, semiconductor counters show better energy resolution (below 1%) than scintillation counters (on the order of 10%). The performance characteristics of scintillation counters are discussed in detail later on.

A detailed treatment of the applications of scintillators as well as the physical processes therein is contained in the standard work by Birks (1). The usage of scintillation counters in high-energy physics experiments such as coincidence measurements is described in Ref. 9. A comprehensive introduction into the detection of nuclear radiation with various detector types can be found in Refs. 10 and 11.

STRUCTURE OF A COUNTING SYSTEM

A modern scintillation counter (as shown in Fig. 1) consists of three main parts: a scintillator, a photomultiplier, and signal-processing electronics.

In the scintillator the energy of the incident particle is converted into light pulses. This scintillation light travels through the scintillator volume and eventually hits the photocathode of the photomultiplier, where it releases photoelectrons. In order to increase the number of photons incident on the photocathode, the scintillator can be covered with a reflecting layer.

In the photomultiplier, a system of electrodes, called *dynodes*, accelerates the electrons towards the anode by means of a high voltage. Upon impact on a dynode, each electron can release additional electrons, a process called *secondary emission*. Hence, the number of electrons in the pulse is multiplied between subsequent dynodes, leading to a large current at the anode. Since the photomultiplier can easily be destroyed if this current is too large, both the scintillator and the photomultiplier have to be screened from ambient light (see Fig. 1). Good optical contact has to be made between them in order to minimize optical losses due to total reflection. A light-guiding structure can also be inserted if the geometry of the application requires the scintillator and the photomultiplier

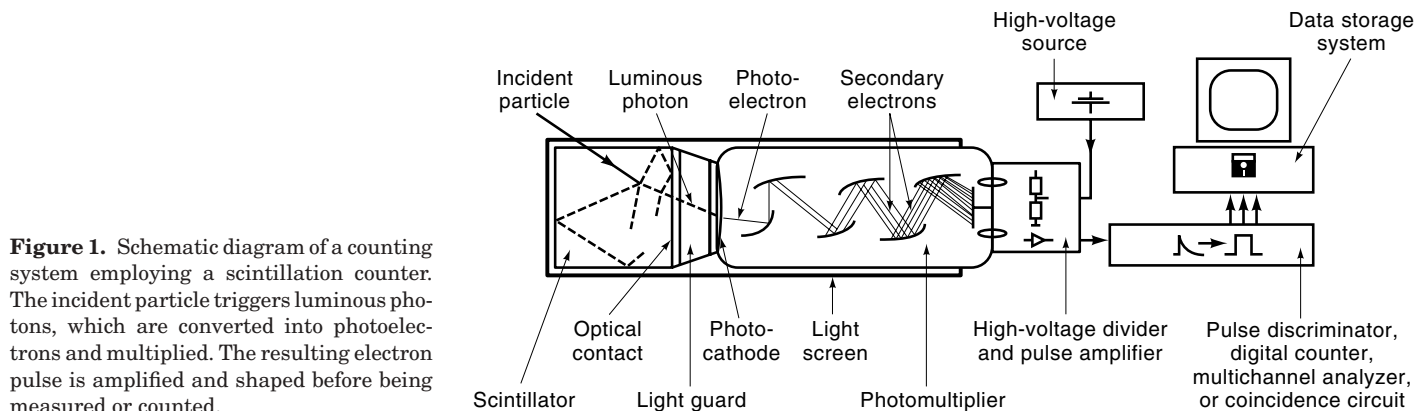


Figure 1. Schematic diagram of a counting system employing a scintillation counter. The incident particle triggers luminous photons, which are converted into photoelectrons and multiplied. The resulting electron pulse is amplified and shaped before being measured or counted.

to be at some distance from each other. The anode pulse of the photomultiplier is fed into a pulse amplifier. In order to minimize background noise, the connection between the photomultiplier and the amplifier has to be kept as short as possible. Therefore, the amplifier is usually contained in a housing, which is directly connected to the photomultiplier socket. In most cases, this housing also contains a voltage divider circuit, which distributes the high voltage to the dynodes of the photomultiplier.

The final constituent of a complete detection system is the signal-processing and data acquisition electronics. The design of the signal-processing electronics depends on the type of application. For measurements of the particle energy, the amplifier output signal is connected to a pulse discriminator/digital counter or a multichannel pulse height analyzer. In measurements requiring a low background, two or more photomultipliers are coupled to the same scintillator. Their pulse amplifiers are connected to a coincidence circuit, which passes a signal to a counter only if the pulses delivered by different photomultipliers arrive within the resolving time of the circuit. In this way, random, spurious pulses generated independently in the photomultipliers can be excluded. In many cases, a digital counter or a multichannel pulse height analyzer is connected to a digital data storage system, for example a computer, where the counts are stored to memory.

The physical processes leading to scintillation are dependent on the scintillator material. The nature of these processes also determines important performance characteristics of scintillation counters, such as the *response time*, the *sensitivity*, the *proportionality*, and the *energy resolution*. This is described in the following sections. A more detailed treatment of these aspects can be found in Refs. 1 and 12.

SCINTILLATOR MATERIALS

Scintillators can be classified into inorganic crystals, glasses and gases, and organic compounds.

The most commonly used inorganic materials are alkali halides such as NaI, CsI, or KI. These are doped with about 0.1% of thallium, which acts as a center for the emission of luminescent photons. Thallium-activated sodium iodide, NaI(Tl), is the most commonly used scintillator; it can be produced in large crystals. Zinc sulfide crystals doped with silver, ZnS(Ag), are used as a scintillating coating. Silicate glasses scintillate when doped with lithium or cerium.

Among the organic substances are anthracene and stilbene crystals, and terphenyl dissolved in liquids (for example xylene or toluene) or incorporated in plastics (for example polystyrene).

Finally, the noble gases helium, neon, argon, krypton, and xenon can be used as scintillators.

The wavelength of maximum intensity of the emitted scintillation light varies among different scintillator materials and is typically between 300 nm and 500 nm for inorganic crystals and organic compounds, and in the ultraviolet region for noble gases.

PROCESSES IN SCINTILLATORS

Producing a scintillation event in one of the materials mentioned involves two steps: the interaction of an incident particle and the subsequent emission of luminescent light.

In general, a part of the energy of the incident particle is absorbed, causing excitation or ionization of scintillator molecules. Luminescent light is emitted during transitions of excited electrons to lower energy levels. The interaction depends on the type of the particle, and the emission process varies between different scintillator materials. Incident particles can be either primary or secondary, examples of the latter being photoelectrons and Compton electrons produced by a γ ray, and secondaries of neutrons. In addition to luminescent light, erenkov radiation can be emitted by a charged particle traversing matter at a speed greater than the speed of light in the respective medium. However, this effect will not be treated in this article.

The energy loss of charged particles in matter, as given by the Bethe-Bloch formula, and the probability for photoelectric absorption, Compton scattering, and pair production by γ rays all increase with increasing density ρ and atomic number Z of the scintillator. Therefore, materials with high values of ρ and Z show high absorption of incident particles per unit volume. In this respect, inorganic crystals ($3.0 \text{ g cm}^{-3} < \rho < 7.5 \text{ g cm}^{-3}$; e.g. $Z = 53$ for iodine) have advantages over organic compounds ($0.8 \text{ g cm}^{-3} < \rho < 1.2 \text{ g cm}^{-3}$; $Z \leq 6$) and noble gases ($\rho \leq 10^{-3} \text{ g cm}^{-3}$ at atmospheric pressure and room temperature). However, because of their low absorption efficiency for γ rays, organic scintillators are predominantly used for the measurement of charged particles at low γ background counting rates.

Charged particles usually deposit all their energy, especially in solid or liquid scintillators, because the range of α particles and electrons in these scintillators is in the order of micrometers and millimeters, respectively. In contrast, γ rays attenuate exponentially by a factor of two over several centimeters and more, and it is therefore possible that they leave only a fraction of their energy in the scintillator. This circumstance is reflected in characteristic features in the spectra observed with scintillation counters suitable for detecting γ rays. This will be briefly explained in the following (with high Z).

If the detector is dense (with high Z) and large enough, some γ rays can be totally absorbed in the scintillator due to the photoelectric effect. It is also possible that they undergo a number of Compton scattering processes before leaving the scintillator, which implies that only the energy of the Compton electrons is transformed into scintillation light. Finally, a γ ray with an energy of more than two electron rest masses (1.022 MeV) can produce an electron-positron pair. The positron is likely to recombine with electrons in the scintillator, producing two 0.511 MeV photons, one or both of which can leave the scintillator.

Figure 2 shows a typical count spectrum of monoenergetic γ radiation measured with a scintillation counter. The peak denoted by E_γ contains events where γ rays deposit their total energy in the scintillator, which includes photoelectric absorption and positron annihilation without the escape of an annihilation photon. The broad spectrum at energies below E_γ is the Compton distribution, resulting from scattering processes involving variable amounts of energy transfer. Within this distribution, there are two distinct peaks at $E_\gamma - 0.511 \text{ MeV}$ and $E_\gamma - 1.022 \text{ MeV}$, which are due to the escape of one or two annihilation photons. It should be pointed out that both the Compton distribution and the single and double escape peaks decrease in height with increasing absorption (that is, with increasing atomic number, density and volume of the scintillator). Hence, a γ spectrum of many different energies is easier

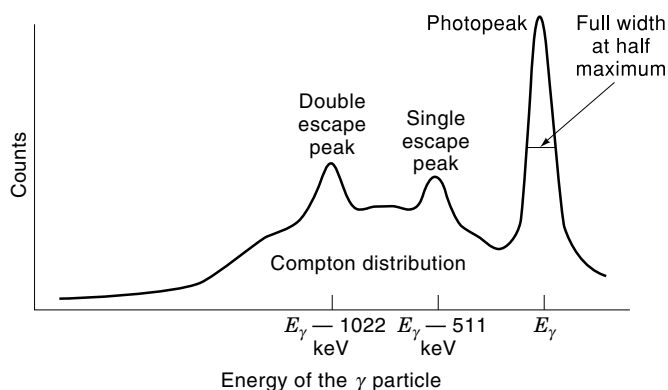


Figure 2. Energy spectrum of γ rays obtained with a scintillation counter.

to interpret when measured with a large and highly absorbing detector.

Neutrons can be detected with scintillation counters by means of secondary reaction products. In (hydrogen-containing) organic scintillators, for example, fast neutrons can produce recoil protons in a collision with hydrogen nuclei (*hydrogen knock-on* reaction). Boron trifluoride (BF_3) counters can detect slow neutrons via nuclear capture according to $^{10}\text{B} + n \rightarrow ^7\text{Li} + \alpha + 2.78 \text{ MeV}$ (11). Europium-activated lithium iodide [$\text{LiI}(\text{Eu})$] crystals have also been successfully used for the $^6\text{Li}(n, \alpha)^3\text{H}$ reaction.

It is possible to distinguish between two types of luminescent emission, depending on how energy is transferred from the excited particle to the luminous photon. A process is called *phosphorescence* if it involves deexcitation of scintillator molecules from relatively long-lived, metastable energy levels. If this is not the case, the process is termed *fluorescence*. Historically, these two variants were distinguished by the typical *relaxation time* τ_r of the pulses, which characterizes the decrease of the scintillation intensity $I(t)$ (number of luminous photons per unit of time t). This decrease can be approximated by

$$I(t) = I(t=0) \exp(-t/\tau_r) \quad (1)$$

These relaxation times are on the order of 10^{-6} s to 1 s for phosphorescence and 10^{-7} s to 10^{-9} s for fluorescence. For further details about processes in scintillators, see Ref. 1.

PERFORMANCE CHARACTERISTICS

The type of emission process depends on the scintillator material. Since a typical photomultiplier spreads a pulse by about 10^{-9} s, the choice of scintillator determines the response time of the whole detector. Organic scintillators and noble gases, for example, show fluorescence, and are therefore predominantly used for applications requiring a fast response. In inorganic crystals, the energy deposited by the incident particle has to be transferred thermally to the luminescent centers (thallium or silver atoms). This leads to longer relaxation times, or, if these luminescent centers have metastable energy levels, to phosphorescence. The scintillation processes in inorganic and organic scintillators and their influence on the response time are described in detail later on.

The *sensitivity* S of a scintillation counter can be written (see Refs. 12 and 13) as the product

$$S = \eta_s T_s f_1 s_\lambda f_2 \eta_c G \quad (2)$$

of (1) the *quantum yield* or *conversion efficiency* η_s (the ratio E_s/E_p of the energy E_s of the emitted luminescent photons to the energy E_p deposited in the scintillator by the incident particle), (2) the *transparency* T_s of the scintillator to its own scintillation light (dependent on the distance traveled through the scintillator volume), (3) a factor f_1 representing the *optical loss* due to reflection and absorption in the windows of the scintillator and the photomultiplier as well as the optical contact between them, (4) the maximum *spectral sensitivity* s_λ of the photocathode (the ratio of the photoelectron current to the power of luminescent photons), (5) a factor f_2 for the *spectral match* of the photocathode material to the wavelength of the scintillation light, (6) the *collection efficiency* η_c of the first photomultiplier dynode for photoelectrons, and finally (7) the *photomultiplier gain* G . All quantities mentioned above are dimensionless, except for the spectral sensitivity s_λ , which has units of milliamperes per watt. The overall sensitivity S can thus be expressed as the current at the photomultiplier output per unit energy flux of the incident particle current. It has units of $\text{mA}/(\text{MeVs}^{-1})$.

The output pulse of a scintillation counter shows *proportionality* to the energy E_p of the incident particle only if this energy is completely deposited in the scintillator and if the conversion efficiency η_s is constant. In this case, the pulse charge released in the photomultiplier from the last dynode to the anode is determined by the sensitivity S of the counter according to

$$Q = SE_p \quad (3)$$

It is this charge Q that is measured with the signal-processing electronics and then multiplied by the counter sensitivity S (usually obtained from calibration) in order to obtain E_p . However, in some circumstances [as in measuring γ rays with a small $\text{NaI}(\text{Tl})$ crystal or α particles with a scintillation chamber, as will be discussed later], only a variable fraction of E_p is deposited in the scintillator. In this case, the charge collected at the photomultiplier anode does not reflect E_p , but only the fraction deposited.

The *energy resolution* describes the uncertainty occurring in the energy measurement described above, and hence is a measure of the ability of the counter to distinguish two different particles even if their energies E_p are very close together. It can be obtained from a measured pulse height or particle energy spectrum as the ratio of the full width of the corresponding peak at half maximum height to the energy E_p . The spread of a pulse height or energy peak is due to two factors: (1) statistical fluctuations in the numbers N_γ , N_p , and N_s of released luminous photons, photoelectrons, and secondary electrons, and (2) variations in the counter sensitivity S . Since the relative statistical fluctuations are proportional to $1/\sqrt{N}$, and since $N_p < N_\gamma$ and $N_p < N_s$, the largest fluctuations are introduced during the release of photoelectrons. Furthermore, the light collection (and hence the transparency T_s) is not uniform over the scintillator volume, and the photomultiplier gain G fluctuates due to instabilities in the high voltage (14), both of which affect the counter sensitivity. For the detection of 1 MeV γ -rays with a $\text{NaI}(\text{Tl})$ crystal, for example, the energy resolution is in the order of 7%.

Table 1. Physical Properties of Some Common Inorganic Scintillators

Scintillator	Light Output [% of NaI(Tl) for Bialkali PMT]	Wavelength of Maximum Emission (nm)	Decay Time (ns)	Density (g · cm ⁻³)
NaI(Tl)	100	415	230	3.67
CsI(Na)	80–85	430	630	4.51
CsI(Tl)	45–50	530	1000	4.51
CsI(pure)	4–8	310	~8	4.51
BGO	10–20	505	300	7.13
BaF ₂	20	310	630	4.88
CaF ₂ (Eu)	50	435	940	3.19
CsF	5	390	5	4.11
ZnS(Ag)	130	450	70	4.09
LiI(Eu)	30–35	470	1400	4.08
GSO(Ce)	20	440	60	6.71
Ar	4	250	20	1.8×10^{-3}
Xe	~20	330	20	5.9×10^{-3}

DETECTOR TYPES AND ACCESSORIES

This section will deal with the most common types of scintillation counters, such as inorganic and organic scintillation counters and crystal, liquid, and gas scintillation counters, as well as scintillation chambers. In addition, details on photomultipliers and optical contacts are presented. Since the early visual and Geiger–Müller scintillation counters have no relevant applications anymore, they will not be described here. A detailed treatment of these types can be found in Refs. 1 and 15.

Inorganic Scintillators

Inorganic scintillators can be classified into three main groups: crystals, glasses, and gases (including liquefied noble gases). The most commonly used are the crystal scintillators, which are mainly alkali halides. Inorganic scintillators are generally characterized by high quantum yield (and therefore good resolution and sensitivity), high density, and high Z (and therefore high efficiencies). See Table 1 for the physical properties of some common organic scintillators. They are therefore ideal as high-energy-electron and γ -ray detectors, and they are used as such in medical imaging applications, γ -ray spectrometry, and radiation dosimetry and as general-purpose radiation detectors for nuclear and particle physics research.

Scintillation Mechanism. Details of the scintillation mechanism for inorganic crystal scintillators vary from one material to another, but the general principle is similar. The scintillation mechanism depends on excitation and deexcitation of electrons between energy states determined by the crystal lattice of the material. The large number of atoms in the crystal leads to energy *bands* (consisting of large numbers of very closely spaced energy levels). Most electrons are in the *valence band* (ground state). A charged particle passing through the material will excite some of the electrons to the *conduction band* (excited state). The electron may emit a scintillation photon and return to the valence band or lose its energy to nonradiative thermal motion. In a pure crystal the scintillation photon can be readily absorbed, since it corresponds to the energy difference between the valence and conduction bands and there are many electrons in the valence band. Absorption leads again to an excited electron that can either emit a photon or lose its energy to non-

radiative thermal motion. At each stage further energy is lost to the nonradiative process, so generally pure crystals are not efficient scintillators.

This problem is solved by adding small amounts of impurities, called *activators*, to the crystals. The activators modify the energy band structure of the crystal by introducing activator ground states that are at a higher energy than the crystal valence band and excited states that are at a lower energy than the conduction band; see Fig. 3. When an electron is excited from the valence band to the conduction band, a hole (vacancy) is formed in the valence band. The electron can move freely through the crystal, whilst the hole can do the same. The hole will quickly drift to a site in the crystal lattice occupied by an activator atom. An electron from the ground state of the activator atom will drop down to the valence band of the crystal, releasing thermal energy. The excited electron (in the conduction band) can now drop down to the excited activator state, also releasing thermal energy. It then emits a scintillation photon as it deexcites to the activator ground state. This has two important consequences. First, the energy of the scintillation photon is less than that required to excite a valence electron (most electrons) to the conduction band, so it is not readily absorbed. The scintillator is therefore transparent to the photons it emits. Second, the energy difference between the ground and

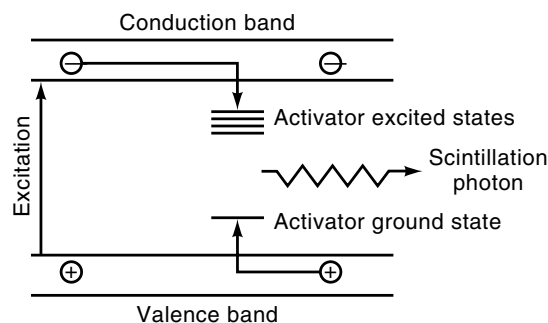


Figure 3. Energy band structure of an activated inorganic scintillator. The presence of activator sites introduces new energy levels. Transitions between these levels lead to photons that are not readily absorbed by the scintillator.

excited activator states corresponds to photons in the visible range that are better matched to the spectral response of photomultiplier tubes. The energy difference between the valence and conduction bands is higher, corresponding to ultraviolet (UV) photons.

Under certain conditions pure (or intrinsic) crystals can, however, be used as efficient scintillators. The difference in the emission and absorption spectra for scintillation photons is attributed to the Franck–Condon principle, which states that any transition between electron energy states happens very quickly compared to atomic or ionic vibrations. This is important because the minimum energy for the conduction band occurs at a larger interionic separation than for the valence band. When an electron is excited to the conduction band, the interionic separation initially remains the same. This means the electron is left with excess energy. It loses it as thermal energy to the crystal lattice while the interionic separation increases. It then emits a scintillation photon and deexcites to the upper levels of the valence band. The energy of this scintillation photon is slightly less than the energy required to excite an electron from the valence band to the conduction band. Pure crystals are efficient only if the number of electrons in the higher valence states is small, and this is usually achieved by cooling the crystals. At room temperature electrons can be found in the higher valence state due to thermal motion and cause reabsorption of the scintillation photons. As long as the pure crystals are cooled, their quantum yield is high and pulse decay time short (16).

In some cases the excited electron is left in a state that cannot decay directly to the ground state (forbidden transition). The electron must obtain some extra energy (usually from thermal excitation) and move to a higher state from which it can decay to the ground state. This process is slow, and the emission is known as phosphorescence.

Crystal Scintillators. By far the most common general-purpose γ -ray detectors are the alkali halide crystals. Sodium iodide crystals with 0.1% thallium as an activator, NaI(Tl), are readily available commercially and are the standard detector used for comparisons with other scintillators. The notable features of NaI(Tl) are its high quantum yield, good energy resolution, and linear response over a large energy range of incident radiation. Unfortunately, it is also hygroscopic, which requires the crystal to be hermetically sealed, usually together with the photomultiplier tube. If not sealed, the crystal quickly deteriorates due to the absorption of moisture in the air. It also has a relatively long scintillation decay time of 230 ns, which may be a problem for very high counting rates, and a phosphorescence component with decay time of 0.15 s. In most applications the phosphorescence is not important, since each phosphorescence photon is detected as a single event giving rise to a single photoelectron. Such a small signal can be ignored by using a discriminator set to the appropriate level. NaI is also fragile and can be easily damaged by mechanical or thermal shock.

For low-energy γ rays the efficiency of a scintillation detector can be increased by using a *well crystal*. This is a circular cylinder with a well, or cavity, along the axis of the cylinder. The source or sample to be measured is placed inside the well, and the scintillator almost completely surrounds it. The counting efficiency is therefore close to 100%.

Cesium iodide is also a popular, readily available scintillator, activated with either sodium or thallium, CsI(Na) and

CsI(Tl). Both types are slightly denser than NaI(Tl), resulting in a larger absorption coefficient. They are only slightly hygroscopic and not as susceptible to shock and vibration. A useful property of CsI(Tl) is its variable decay time for different incident particles (17). This allows *pulse shape discrimination* between heavy charged particles such as α particles and protons on the one hand and electrons and γ rays on the other. However, the emission spectrum is at a much longer wavelength than NaI(Tl) and is poorly matched to the spectral response of most photocathodes. The response is greatly improved by using photodiodes (semiconductor devices) that are sensitive to these longer wavelengths.

Lithium iodide activated with europium, LiI(Eu), and enriched with ^6Li is often used for the detection of slow neutrons. A slow neutron interacts with ^6Li in the crystal and produces tritium and an α particle. The total energy of the reaction products is fixed, and so the response is characterized by a well-defined peak. However, this response also corresponds to that for electrons or γ -rays of energy 4.1 MeV. Therefore, any electron and γ -ray background at these energies is important when making neutron measurements. Lithium iodide is also highly hygroscopic and must be sealed.

A high-density pure crystal scintillator that is commonly used is bismuth germanate (BGO), $\text{Bi}_4\text{Ge}_3\text{O}_{12}$. Its high density is an advantage in applications where a compact high-efficiency counter is required. It is also mechanically and chemically robust and therefore easy to handle. The scintillation emission is characterized by a fast component with a decay time of 60 ns and a dominant slower component of 300 ns making its timing characteristics similar to NaI(Tl). The quantum yield, however, is only about 10% to 20% of that of NaI(Tl), and so the resolution is much poorer. This means it is appropriate for counting applications, such as PET positron emission tomography) scanners, but unsuitable for spectrometry.

One of the first scintillators used was zinc sulfide activated by silver, ZnS(Ag). Unlike other crystal scintillators, it can only be produced as a polycrystalline powder rather than a single crystal. It is opaque to its own scintillation light, so it is used as a thin layer or screen between 20 μm and 60 μm thick. It is primarily used for counting very low-energy electrons, α particles, or other highly ionizing charged particles. Its quantum yield is comparable to that of NaI(Tl), making it very efficient. It is sometimes used to coat layers of plastic, or some other material high in hydrogen, in order to detect recoil protons due to incident fast neutrons (16). If the plastic contains ^6Li or ^{10}B , it can also detect slow neutrons.

Glass Scintillators. Most glass scintillators are made from silicate glasses and activated by cerium. The decay time for the scintillation emission is typically between 50 ns and 100 ns. They can be produced cheaply and easily in many different shapes and sizes, and they are mechanically and chemically very robust, making them suitable for extreme environments such as high temperatures or the presence of corrosive chemicals. Their quantum yield is only about 10% of that of NaI(Tl), so they are not suitable for spectrometry. Glass scintillators containing lithium are widely used as fast-neutron detectors. Their short response time means they are suitable in neutron time-of-flight spectroscopy.

Gas Scintillators. Gases can also be made to scintillate and serve as useful detectors for certain applications. The most

Table 2. Physical Properties of Some Common Organic Scintillators

Scintillator (Solvent–Primary Solute– Secondary Solute)	Light Output [% of NaI(Tl) for Bialkali PMT]	Wavelength of Maximum Emission (nm)	Decay Time (ns)	Density (g · cm ⁻³)
<i>Crystals</i>				
Anthracene	40	440	30	1.25
<i>trans</i> -Stilbene	20	410	6	1.16
<i>Plastics</i>				
Polystyrene-tetraphenylbutadiene	14	450	5	~1.0
Polyvinyltoluene- <i>p</i> -terphenyl- <i>p</i> , <i>p'</i> -diphenylstilbene	19	380	3	~1.0
Polyvinyltoluene- <i>p</i> -terphenyl-tetraphenylbutadiene	18	445	4	~1.0
<i>Liquids</i>				
Toluene-PPO	20	365	3.8	0.88
Toluene- <i>p</i> -terphenyl-POPOP	24	425	3.5	0.88
Toluene-PPO-POPOP	30	425	3.7	0.88

efficient scintillating gases are the noble gases and nitrogen (18). Several different processes give rise to the scintillation light, of which the most efficient is the formation of excited molecules that radiate when returning directly to the ground state. For the noble gases the emission spectrum is usually in the UV and forms a continuum. To detect such photons requires the use of special UV-sensitive photocathodes, photodiodes, or wavelength shifters, such as a small amount of nitrogen gas. The quantum yield is quite low compared to that of NaI(Tl), but the decay time for the emission is generally short, in the order of 20 ns or less. The quantum yield also depends on the purity of the gas, as any impurities can reduce the light output. This is because the impurities absorb the excitation and ionization energy without emitting scintillation photons. Such a process is termed *quenching*. The response of gas scintillators is also very linear with energy over a large range. At low pressures (<10⁵ Pa = 1 atm) they have very low efficiencies for γ rays, so that they are used almost exclusively as heavy-ion detectors, most notably for fission fragments. In most practical applications the gas is used at high pressure, increasing the efficiency of the detector.

Gas scintillators are also used in a hybrid detector called the *scintillating proportional detector* (19). A large electric field between two electrodes accelerates any electrons produced from ionizing interactions in the gas. As these electrons move towards the anode, they collide with neutral gas molecules, exciting them. The gas molecules deexcite, emitting scintillation photons. This is a source of secondary scintillation. The quantum yield is greatly increased, approaching or even exceeding that of NaI(Tl), but the drift of the electrons is slow and the secondary scintillation photons are emitted a few microseconds after the prompt scintillation photons and last for a few microseconds, resulting in a relatively slow response.

The noble gases also scintillate when condensed cryogenically to a liquid or solid. In fact, their quantum yield is then comparable to that of NaI(Tl). Liquid argon (LAr) ionization chambers have been used for many years, but only recently has the need for fast detectors prompted their use as scintillators because of their fast scintillation response. Liquid xenon and krypton detectors, which have high *Z*, can also be used as γ -ray detectors. Being liquid, they can be produced in any shape and size, which is also advantageous.

Organic Scintillators

Organic scintillators can be classified into three types: crystal, liquid, and plastic. Compared to the inorganic scintillators, they have lower densities and *Z* values, resulting in lower detection efficiencies, especially for γ rays. The quantum yield is also lower, resulting in poorer energy resolution. (See Table 2 for the physical properties of some common organic scintillators.) They do, however, have several advantages. They are very cheap, can be produced in many shapes and sizes, and are mechanically robust. They are also characterized by extremely short response times, whence their use in fast timing applications and at high event rates. Addition of boron or gadolinium produces scintillators that can be used for neutron detection and spectrometry.

Scintillation Mechanism. The luminescence process in organic scintillators is due to energy transitions in individual molecules, rather than the lattice structure of the material as in inorganic crystals. Any particular molecule can therefore be observed to scintillate, regardless of its physical state (solid, liquid, or gas). Almost all organic scintillators contain at least one benzene ring (or benzene-ring derivative) and it is this structure that is responsible for the scintillation.

Benzene is an organic molecule that contains six carbon atoms arranged in a closed ring. Each carbon atom forms four bonds with its four valence electrons. Three of these bonds are σ bonds that are close to the carbon atom and hence strong bonds. The fourth electron forms a π bond that is further away and only weakly bound. This electron is therefore easily excited or ionized and is responsible for the scintillation mechanism. The π electron is usually found in the ground state, but it can be excited into any of the higher energy states. The structure of these electronic states is shown in Fig. 4. It consists of the basic electronic states S_0 (the ground state), S_1 (first excited state), S_2 (second excited state), and so on. The typical difference in energy between S_0 and S_1 is about 3 eV or 4 eV. These states are further split into substates with slightly different energies due to vibration of the molecule, the difference between adjacent states being about 0.15 eV. They are denoted by S_{00} , S_{01} , S_{10} , S_{11} , and so on. These states are actually split even further by different molecular rotational energy substates, but

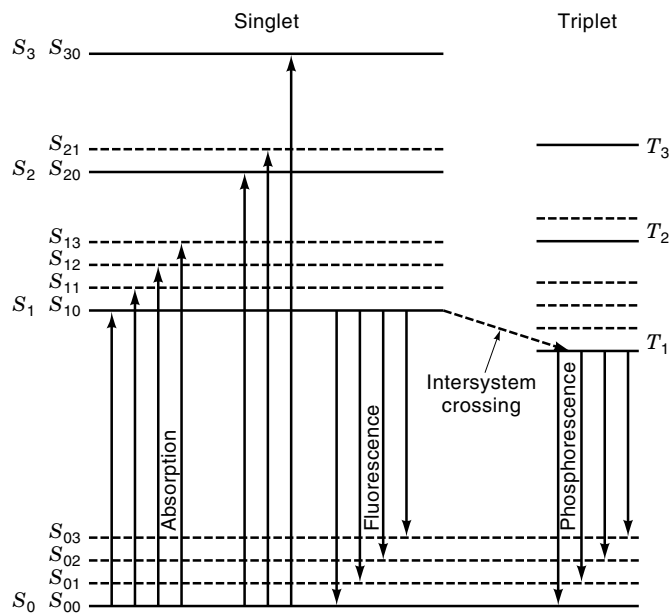


Figure 4. Energy levels of an organic molecule with π electron structure. Electron transitions are indicated for fluorescence and phosphorescence.

the difference of such energy levels is much smaller and can be ignored when considering scintillation. Together these energy levels form the singlet states (spin 0) for the π electron. There is a similar set of states called the triplet states (spin 1) that the π electron can also occupy. These are indicated as T_1 , T_2 , T_3 , and their splitting has not been included in Fig. 4 for simplicity.

Three types of luminescence have been identified for organic scintillators that differ in either their emission spectrum or their decay time. They are fluorescence, phosphorescence, and delayed fluorescence. The most important process that is responsible for most of the scintillation light emission is fluorescence.

The spacing of the vibrational and rotational substates means that at room temperature nearly all the molecules of a scintillator are in the S_{00} state. A charged particle incident on the scintillator can either excite such a molecule to any of the higher energy level states or ionize it. An excited molecule quickly ($\sim 10^{-12}$ s) loses energy through internal nonradiative processes until it reaches one of the S_1 substates and then thermally deexcites to the S_{10} state. The molecule then emits scintillation light (fluorescence) as it deexcites from the S_{10} state to one of the S_0 substates. This process occurs on a time scale of nanoseconds and determines the decay time of fluorescence. Since the higher vibrational substates of the ground state are not occupied, the emission and absorption spectra overlap only slightly and the scintillator is transparent to its own emission. An ionized molecule, on the other hand, usually recombines with the electron and ends up in one of the excited states. In most cases this is a triplet state.

In some cases the excited S_{10} state can undergo a nonradiative transition to the T_1 triplet state. This is called *inter-system crossing*. The lifetime of the T_1 state is characteristically much longer (10^{-4} s) than that of the S_1 state. The molecule deexcites to an S_0 state by emitting light that is usually of slightly longer wavelength than that emitted during fluorescence, due to the

difference in energy between the T_1 and S_1 states. This emission is also much slower and is called phosphorescence. Whilst in the T_1 state, a molecule may absorb some energy and be excited back to the S_1 state. This molecule may then deexcite to a S_0 state, just as in normal fluorescence. The emission spectrum is the same as in fluorescence, but the process is delayed ($\sim 10^{-6}$ s) due to the time spent in the T_1 state. Such emission is called *delayed fluorescence*.

Organic Crystals. There are many organic crystal scintillators, but only three are widely used: anthracene, trans-stilbene, and quaterphenyl. Anthracene has the highest scintillation efficiency of all organic scintillators, yet still only about half that of NaI(Tl). The fluorescence of both trans-stilbene and quaterphenyl has a decay time that depends on the incident particle. Fast fluorescence is produced by γ rays, slower fluorescence by neutrons, and even slower fluorescence by α particles. Pulse shape discrimination can therefore be used to identify the incident radiation and perhaps reduce or eliminate any background. Since organic scintillators contain a large amount of hydrogen, they are efficient fast-neutron detectors when used in pulse shape discrimination mode. Organic crystal scintillators are, however, fragile and difficult to obtain in large sizes. In addition, the quantum yield is known to depend on the direction of the incident particle with respect to the crystal axis (20). The difference in response with direction can be as much as 20% to 30%, which severely degrades the energy resolution of such detectors.

Liquid Organic Scintillators. Liquid organic scintillators form a very versatile and useful category of detectors. The liquid state allows easy addition of other materials to improve and optimize the efficiency of such detectors. Addition of boron or gadolinium allows the detection of slow neutrons. To enhance the probability of photoelectric conversion of γ rays, lead or tin is added, which results in a very efficient and fast scintillator (21). Low-energy γ rays or β particles can be easily detected by dissolving or incorporating the sample in the liquid organic scintillator. This reduces losses due to absorption by the scintillator outer covering or the source itself, resulting in very high detection efficiencies. In addition, further fluorescent compounds can be added to shift the scintillation light spectrum to longer wavelengths in order to match better the spectral response of most photocathodes.

The simplest type of liquid organic scintillator consists of a fluorescent material, called the *primary solute*, dissolved in the solvent that makes up the bulk of the scintillator. Common solvents include toluene, xylene, and 1,4-dioxan. These compounds fluoresce when exposed to ionizing radiation; however, the quantum yield is low, and self-absorption can be a problem for large detectors. The addition of a primary solute has several effects that increase the efficiency of the scintillator. Firstly, the primary solute is a much more efficient scintillator, so that energy is transferred from molecules in the solvent (most energy loss occurs in the solvent, since it constitutes the bulk of the material) to molecules of the solute. This can occur by one of the following means: radiative transfer (scintillation light emission by solvent and absorption by solute), nonradiative dipole-dipole exchange, or collision exchange via diffusion of excited solvent molecules. Secondly, scintillation emission from most solvents is usually at short wavelengths that are not easily detected by most common photocathodes. The solute,

therefore, also acts as a wavelength shifter with an absorption spectrum well matched to the emission of the solvent and an emission spectrum well matched to the spectral response of most photocathodes. This wavelength shift also eliminates the problem of self-absorption by the solvent. In some cases it is necessary or desirable to add a secondary solute to act purely as a wavelength shifter, absorbing the scintillation light of the primary solute and emitting at slightly longer wavelengths. This achieves an optimum match with the spectral response of the photocathode and reduces self-absorption by the primary solute if the dimensions of the detector are large. Popular primary solutes include PPO (2,5-diphenyloxazole), PBD (2-phenyl-5-(4-biphenyl)-1,3,4-oxadiazole), and PTP (paraterphenyl), a common secondary solute is POPOP [1,4-di-(2-5-phenyloxazolyl)benzene].

The quantum yield of liquid organic scintillators is affected by the presence of impurities. Such impurities, especially dissolved oxygen, can serve as quenching agents significantly reducing the quantum yield of the scintillator. Dissolved oxygen is generally removed by bubbling nitrogen through the liquid and kept out by sealing the liquid in an airtight container.

A popular use of liquid scintillators is as anticoincidence shields to reduce unwanted background events. Such shields are effective at detecting high-energy cosmic rays or high-energy γ rays that may mimic a real event. When an event is detected by the anticoincidence shields, any signal in the main detector is vetoed. Liquid scintillators can be of any shape or size and hence provide complete coverage at a very low cost.

Liquid organic scintillators are also routinely used to measure the activity of biological or chemical samples containing ^{14}C or ^3H . The low-energy electrons from such sources are most easily measured by dissolving the sample in the liquid scintillator. This may introduce problems such as low solubility of the sample, quenching by the sample, and reactions with the solvent or solute. However, stable solutions for most cases can be obtained by the appropriate choice of solvent and solute (22). Once the solution of liquid scintillator and sample has been prepared, it is placed in a light-tight enclosure and counted by a photomultiplier tube.

Plastic Scintillators. Organic plastic scintillators can be thought of as solid solutions, and hence they are very similar to organic liquid scintillators. They are prepared by using a solvent that can be polymerized (such as polystyrene or polyvinyltoluene), a primary solute, and perhaps a secondary solute. The solutes that make efficient scintillators for liquid solutions are also efficient scintillators in solid (plastic) form, so that PPO, PBD, and PTP are used. POPOP is a common secondary solute used as a wavelength shifter. For plastics the quantum yield of the scintillator can also depend on the method of preparation as well as its constituents (23). In addition, transfer of energy from the solvent to the solute is not as efficient as for solutions (solvent molecules cannot diffuse), and therefore greater concentrations of solute must be used to obtain comparable quantum yields.

Due to their mechanical strength, plastic scintillators are easy to handle and can be cut to any shape and size. They are relatively cheap and therefore practical in large-volume coincidence or anticoincidence counters. They are used extensively in nuclear physics, for lifetime and time-of-flight measurements, because of their fast response. Although, like the other organic



Model 110A

Figure 5. Lucas cell. This type of counter is used for measuring α radiation in gases.

scintillators, they can be used to detect both fast (scattering from protons) and slow (loading with boron or gadolinium) neutrons they do not have such good pulse shape discrimination characteristics.

Plastic detectors can be produced as very thin film scintillators with thicknesses down to $10\ \mu\text{m}$ (24). This allows the detection of low-energy highly ionizing particles such as heavy ions without complete loss of energy. Such detectors are known as *transmission detectors*.

Scintillation Chambers or Lucas Cells

A *Lucas cell*, named after its inventor (25), is a scintillation chamber used for the measurement of α radiation in gases. The cell is a cylinder of volume 0.1 L to 2 L, with a $\text{ZnS}(\text{Ag})$ coating on the internal surfaces. One end of the cylinder is a quartz window, while at the opposite end two connectors are mounted for the exchange of gas samples (see Fig. 5).

Because of their high counting efficiency (about 75% for the measurement of 6 MeV α particles in a standard 150 mL cell), Lucas cells allow the measurement of gas samples with small volumes and activities as low as $1\ \text{Bq m}^{-3}$. The $\text{ZnS}(\text{Ag})$ scintillator material has its maximum emission at a wavelength of $\lambda = 450\ \text{nm}$, with a bandwidth of about 55 nm. Because of its high density and atomic number ($\rho = 4.1\ \text{g cm}^{-3}$, $Z = 23$), it has a high conversion efficiency of $150\ \text{eV/keV} \leq \eta_s \leq 200\ \text{eV/keV}$. In order to ensure a maximum yield of luminous photons, the coating thickness has to be larger than the range of the α particle, which, for ZnS , is about $15\ \mu\text{m}$ at energies around 6 MeV. The only disadvantage of Lucas cells is that $\text{ZnS}(\text{Ag})$, as a phosphorescent material, has a long relaxation time $\tau_r \approx 20\ \mu\text{s}$.

The efficiency of Lucas cells depends on their geometry and size. This is because the distance which the α particle has to travel to reach a cylinder wall depends on the location of the α decay within the cell and the direction of emission of the α particle. The efficiency of the cell is equal to the fraction of decay events where the distance between the decay location and the cylinder wall for a given direction is less than the range of the α particle in the gas (26). This circumstance also implies that, even for monoenergetic α particles, the amount of energy

deposited in the scintillator ranges from zero (the particle does not reach the cylinder wall) to E_p (the particle is emitted at a location directly adjacent to a wall). Therefore, Lucas cells are unsuitable for energy measurements.

Lucas cells are extensively used to detect radon, a naturally occurring radioactive noble gas. They can monitor the radon concentration in air either as a passive device with radon diffusing into the sensitive volume (27), or actively with air being continuously forced through the cell (28) or periodically exchanged (29). Radon is produced in the decay of radium in soils, rocks, and building materials. From there it enters residential buildings or workplaces such as mines, and, together with its progeny, forms a significant source of the human radioactive dose. In addition to surveys of the radon concentrations in the human environment, Lucas cells have also been employed to investigate the migration of radon in the earth's crust. Radon is released into the pore space of the rock preferentially under mechanical stress, and is then transported to the surface along fractures and faults. Thus, field studies have demonstrated a correlation of the radon exhalation from the ground with periodic stresses caused by tidal forces as well as with transient stresses before and during seismic events. Due to the latter fact, a sudden increase in the radon exhalation can be utilized as an earthquake precursor. In contrast, a constant regional anomaly of the radon exhalation can indicate the presence of an underground uranium ore body.

Photomultipliers

The photomultiplier is an essential constituent of every modern scintillation counter, because it enables the scintillation light pulse to be converted into an electrical signal, which can be electronically processed for further purposes.

A photomultiplier, as shown in Fig. 6, is a vacuum tube with a window for the entry of light at the top and a socket with electrical connectors at the base. The inside of the window is coated with an either opaque or semitransparent photoemissive layer,

the photocathode. Photons incident on the photocathode release up to five photoelectrons, which are collected by an electron—optical system and focused onto the first of several electrodes. In the case of a photomultiplier, these electrodes are called *dynodes*. Since a high voltage (≈ 100 V to 150 V typically) exists between each pair of subsequent dynodes, the electrons are accelerated from one dynode to the next. Upon impact on each dynode, an electron causes the emission of further secondary electrons, thus amplifying the pulse from one stage to another. The secondary electron yield depends on the impact energy of the primary electrons and therefore on the accelerating high voltage. Electrons emitted from the last dynode are collected on the anode, which forms the output stage of the photomultiplier. The transit time of a pulse from the photocathode to the anode is in the order of 50 ns, but the pulse is spread by less than 1 ns during its passage. During operation the photomultiplier has to be shielded from ambient light in order to avoid its destruction due to an excessive anode current. Furthermore, the electrostatic focusing of the dynodes and hence the pulse shape can be adversely affected by external magnetic fields. Therefore, the light shield can also serve as a magnetic shield when made of Mumetal (9).

Photomultiplier tubes vary mainly with regard to the number of dynodes and the photocathode material. Tubes with ten or more dynode stages are common. The wavelength λ_m of maximum photocathode sensitivity should be within the spectral range of the scintillation light. It depends on the photocathode material and ranges from 200 nm to about 450 nm. At wavelengths above 600 nm, the photon energy is below 2 eV and is hence lower than the energy E_a that is needed to release a photoelectron from the surface of a typical photoemissive material ($2 \text{ eV} \leq E_a \leq 4 \text{ eV}$). The sensitivity at short wavelengths is limited by the absorption of photons in the window glass. The cutoff wavelength is about 350 nm for normal glass, but can be extended to 160 nm by using a quartz window (14). The maximum sensitivity s_λ is typically 80 mA/W. With h being Planck's constant, c the speed of light, and e the electron charge, this corresponds to photocathode quantum efficiency

$$\eta_c = s_\lambda \frac{hc}{\lambda_m e} \quad (4)$$

of about 25%.

Two further important characteristics of photomultipliers are the *gain* and the *dark current*. The gain is the ratio of the amplified anode current to the initial photoelectron current. It depends on the number and design of the dynodes and also on the high voltage applied. In order to ensure a constant gain, the high voltage should be well stabilized. The gain is typically very high (between 10^5 and 10^8), so that single photons incident on the photocathode can be registered. The dark current is mainly the anode current at zero photocathode illumination. It consists of very short pulses (≈ 1 ns, the transit time spread), which are due to the thermally activated emission of electrons from the photocathode. The dark current is therefore dependent on the operating temperature. Typical values range from 1 nA to 100 nA, depending on the high voltage.

Since the gain is determined by the high voltage, particular attention has to be paid in designing a suitable high-voltage divider circuit. The voltage distribution along the divider can vary widely, depending on the desired pulse shape. Figure 7

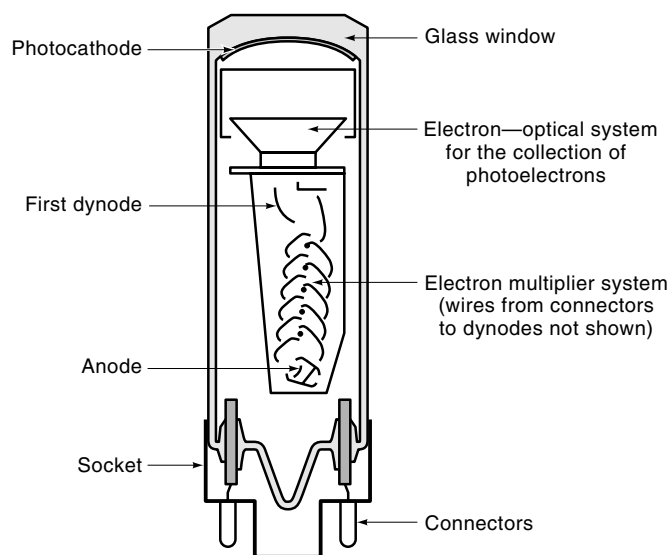


Figure 6. Internal structure of a photomultiplier. (Courtesy of Photonis, France)

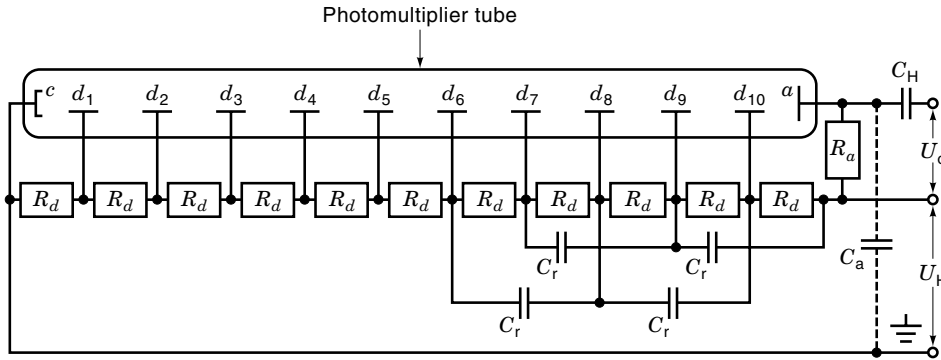


Figure 7. A simple high voltage divider circuit used with a photomultiplier. Photoelectrons generated at the photocathode c are accelerated toward the anode by a high voltage U_H .

shows a simple divider circuit supplying a high voltage to the dynodes d_1 to d_{10} as well as to the cathode c and anode a of a photomultiplier tube.

The anode has to be positive with regard to the photocathode in order to accelerate the photoelectrons. This can be achieved by either (1) connecting the high voltage to the anode and grounding the photocathode, or (2) vice versa. The second polarity option has the disadvantage of a high voltage across the glass housing, which can induce spurious pulses due to electroluminescence in the glass (14). However, the anode can be connected directly to the input of the pulse amplifier. This is not the case in the first option (as shown in Fig. 7), where a high-voltage capacitor C_{HV} has to be inserted in order to protect the pulse amplifier from the high dc voltage while still allowing the signal pulse to pass.

The gain and therefore the signal current I_a at the anode of the photomultiplier is adjusted via the high voltage U_H . If A is the maximum count rate and Q is the charge per pulse [see Eq. (3)], then

$$I_a = QA \quad (5)$$

I_a should be much higher than the dark current.

The electrons transiting between the dynodes cause a mean current across the divider resistors R_d in the opposite direction. Hence, the divider voltage is reduced during a pulse passage, giving rise to variations in the photomultiplier gain. To minimize the effect of a variable mean anode current on the gain, these resistors are chosen so that the divider current $I_d \geq 100 I_a$.

Furthermore, a pulse current across the last dynodes with a decay time equal to the relaxation time τ_r of the scintillation pulse causes a sudden reduction of the voltage U_d across the corresponding divider resistor R_d of about $\Delta U_d = QR_d/\tau_r$, leading to gain fluctuations. These fluctuations can be attenuated by using *reservoir capacitors* (C_r in Fig. 7). During a pulse these capacitors instantaneously supply enough charge to keep the interdynodal voltage almost constant. They are subsequently recharged with a current through the corresponding divider resistor R_d with a time constant $\tau_d = R_d C_r$. If $\tau_d \gg \tau_r$, the pulse current across R_d is spread out in time and has a much lower peak amplitude of about $\Delta U_d = Q/C_r$ (9).

Similar considerations apply for the shape of the signal pulse at the photomultiplier output. This stage consists of an effective anode resistance R_a and a capacitance C_a , forming an RC circuit with a time constant $\tau_a = R_a C_a$. Here C_a is usually not

a physical capacitor, but the stray capacitance of the anode to ground (and hence drawn in dashed lines in Fig. 7). Note that R_a and C_a include resistances and capacitances at both the photomultiplier output and the input of the following pulse amplifier (22). Two cases can be distinguished. If (1) $\tau_a \ll \tau_r$, the pulse charge is quickly passed to ground through R_a and the duration τ_o and height ΔU_o of the output pulse are determined by the relaxation time τ_r , with $\tau_o = \tau_r$ and $\Delta U_o = QR_a/\tau_r$. This mode of operation is called the *current mode*. It ensures a minimum response time and is therefore used in applications requiring fast timing. In this mode, the thermally activated pulses of the dark current can produce quite large output pulses because of their short duration. This background can be effectively eliminated in *coincidence measurements*, where two photomultipliers are coupled to the same scintillator, and a pulse is counted only if it occurs in both photomultipliers within a very small time window. If (2) $\tau_a \gg \tau_r$, the whole pulse charge is integrated on C_a before it can flow to ground through R_a . Then $\tau_o = \tau_a$ and the pulse height $\Delta U_o = Q/C_a$ is proportional to the charge Q and independent of τ_r . Hence, if the sensitivity of the counter is a constant, the energy E_p of the incident particle can be calculated from ΔU_o [see Eq. (3)]. This mode is called the *voltage mode* and is suitable for the measurement of particle energy spectra (30).

Further details about pulse amplifier and high-voltage divider circuits can be found in Ref. 31.

Optical Contact

When transiting from the scintillator to the photocathode of the photomultiplier, the luminous photons can be lost from the light pulse due to reflection from the window surfaces (see Fig. 1). These optical losses can be avoided by coupling the scintillator and the photomultiplier windows with a transparent optical grease or oil, which should have a refractive index that is in between that of the scintillator (1.45 to 1.85) and that of the glass of the photomultiplier window (≈ 1.5).

Some applications require the photomultiplier to be mounted at some distance from the scintillator, for example, because of geometrical constraints in the experiment or because of the presence of strong magnetic fields near the scintillator. In these cases, the scintillator and the photomultiplier are connected with a light guide made of Perspex or Plexiglas or a similar transparent material. Light guides are also used in order to couple scintillators and photocathodes of different sizes and shapes.

PRESENT STATE OF TECHNOLOGY AND OUTLOOK

The widespread use of scintillators has ensured that advances and improvements in scintillator technology continue. Research into new scintillating material with higher quantum yield, better-matched spectral response, improved radiation hardness (deterioration of response with exposure to radiation), faster response, or a combination of these properties is actively pursued. In addition, new methods for using scintillating materials are being developed.

The new generation of nuclear and particle physics experiments require beams of even higher energy and higher fluxes. This means the new detectors must be radiation-hard, very fast, and dense (to provide compact detectors). A very likely candidate for such a scintillator material is barium fluoride BaF_2 (32). It has a fast (subnanosecond, 0.6 ns to 0.79 ns) decay component and a slower (100 ns to 1000 ns) component. These components have different emission spectra (195 nm and 220 nm for the fast component and 310 nm for the slow component) and a large difference in decay constants, so that it is very simple to select the fast component only. A disadvantage is that the fast emission is in the UV, requiring special photomultiplier windows and photocathodes that are sensitive at these lower wavelengths but insensitive to higher (>300 nm) wavelengths. Attempts to decrease the relative contribution of the slow component include doping the BaF_2 with lanthanum. This has proved to be successful.

Other possible candidates for fast, dense, radiation-hard scintillators include cerium fluoride (CeF_3) (33) and gadolinium orthosilicate doped with cerium [$\text{GSO}(\text{Ce})$] (34). They both show excellent radiation hardness, good quantum yield, and a good spectral match with most existing photocathodes. Their main disadvantage is their relatively long decay times: 20 ns to 35 ns for CeF_3 , and even longer for $\text{GSO}(\text{Ce})$, depending on the doping concentration of Ce. Along with lutetium orthosilicate doped with cerium [$\text{LSO}(\text{Ce})$], these new scintillating materials have already been used in positron emission tomography (PET).

Advances have also been made with organic scintillators. Scintillators that are radiation-hard and have a higher quantum yield and better-matched spectral response have been developed. Radiation hardness depends on how the optical transmission properties of the base are affected by radiation and also on damage to fluor molecules. Radiation-tolerant bases have been developed by some manufacturers (35), and the effect of radiation damage to the fluor molecules can be minimized by increasing their concentration. New polystyrene bases have also been produced and tested regarding radiation hardness, quantum yield, and stability of response with time (36). To produce scintillators with better spectral response and longer attenuation length, a new class of fluors with a large Stokes shift (difference in the absorption and emission spectra) have been developed, such as 1-phenyl-3-mesityl-2-pyrazoline (PMP) and 3-hydroxyflavone (3HF). They enable the construction of larger detectors and longer scintillating fibers.

A recent application of scintillators is in *scintillating fiber trackers* (37). Such detectors consist of a large number (10^5 to 10^6) of scintillating fibers that can be either glass, plastic, or capillaries filled with liquid scintillator. The fibers can be square, circular, or hexagonal in cross section and are typically 20 μm to 50 μm across and up to several meters in length. The

scintillating fibers provide fast and efficient charged-particle tracking with position resolution down to 20 μm combined with the flexibility of instrumenting large and unusually shaped volumes. Such trackers have been used in many particle and nuclear physics experiments. Glass fibers doped with cerium oxide were the first to be implemented and produced fast and efficient trackers. However, the short attenuation length of such fibers limits their use in larger detectors. Plastic scintillators and capillaries with liquid scintillator show more promise for such applications, although a disadvantage of plastic scintillators is crosstalk between nearby fibers. The light output of such huge arrays of fibers can be detected using several different types of photosensors. If the event rate is low, image intensifiers and charge-coupled devices (CCDs) can be used. Otherwise, multianode photomultipliers, avalanche photodiodes, or visible-light photon counters (VLPCs) must be used. The last are the preferred choice, due to their spectral response, high quantum yield (70% to 80%), high gain, gain stability, and low noise level. Although for the present scintillating fiber trackers are used only in research detectors, it is only a matter of time before their excellent resolution, speed, and efficiency are made available in medical imaging (PET detectors) and other areas.

An alternative to scintillating fibers is *scintillating pads* (38). Such pads are rectangular pieces of organic plastic scintillator of area 1 cm $\times$ 2 cm and about 3 mm thick. They are coupled to photodetectors by wavelength-shifting fibers. They do not have the position resolution of scintillating fibers, but can be effectively used for triggering and tracking over large volumes.

A further method of using scintillators is in *scintillation proximity assay* (PSA). This is a technique used to trace biological molecules and analyze biomolecular interactions. Small beads of scintillator about 5 μm across are chemically coated to allow molecules such as antibodies or proteins to become attached to the bead. These molecules are labeled with a radioactive tracer such as tritium. The beads and molecules are placed in aqueous solution. When the molecule becomes attached to the scintillating bead, the radiation from the tracer causes the scintillator to emit light that can be detected by a photomultiplier or other light-sensitive device. The tracer is chosen to emit low-energy radiation, so that a bead will only emit scintillation light if the tracer molecule is very close. The scintillating beads are usually made from cerium-doped glass or some organic plastic scintillator.

Current research and development of new scintillating materials and scintillation techniques will ensure that scintillators will continue to be an important part of radiation and particle detection. New devices based on semiconductor technology have not replaced but rather enhanced the use of scintillators by providing alternative and more sensitive photosensors.

BIBLIOGRAPHY

1. J. B. Birks, *The Theory and Practice of Scintillation Counting*, Oxford: Pergamon Press, 1964.
2. H. Geiger and E. Marsden, On a diffuse reflection of the α -particles, *Proc. Roy. Soc.*, **A82**: 495–500, 1909.
3. E. Rutherford, The scattering of α and β particles by matter and the structure of the atom, *Phil. Mag.* VI, **21**: 669–688, 1911.

4. A. Krebs, Ein Demonstrationsversuch zur Emanationsdiffusion, *Ann. Physik*, **39**: 330–332, 1941.
5. S. C. Curran and W. R. Baker, Photoelectric alpha-particle-detector, *Rev. Sci. Instrum.*, **19**: 116, 1948.
6. H. Kallmann, *Natur und Technik*, July 1947.
7. P. R. Bell, The use of anthracene as a scintillation counter, *Phys. Rev.*, **73**: 1405–1406, 1948.
8. R. Hofstadter, Alkali halide scintillation counters, *Phys. Rev.*, **74**: 100–101, 1948.
9. Yu. K. Akimov, *Scintillation Counters in High Energy Physics*, New York: Academic Press, 1965.
10. H. Frauenfelder and E. M. Henley, *Subatomic Physics*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1991.
11. S. E. Hunt, *Nuclear Physics for Engineers and Scientists*, New York: Wiley, 1987.
12. E. Fenyes and O. Haiman, *The Physical Principles of Nuclear Radiation Measurements*, New York: Academic Press, 1969.
13. W. J. Price, *Nuclear Radiation Detection*, New York: McGraw-Hill, 1964.
14. J. Chariton and C. A. Lea, Some experiments concerning the counting of scintillation produced by alpha particles, *Proc. Roy. Soc.*, **A122**: 304–352, 1929.
15. G. F. Knoll, *Radiation Detection and Measurement*, 2nd ed., New York: Wiley, 1979.
16. W. H. Tait, *Radiation Detection*, London: Butterworths, 1980.
17. C. J. Crannell, R. J. Kurz, and W. Viehmann, *Nucl. Instrum. and Meth.*, **A115**: 253–261, 1974.
18. M. Mutterer, *Nucl. Instrum. Meth.*, **A196**: 73–81, 1982.
19. A. J. P. L. Policarpo, *Space Sci. Instrum.*, **3**: 77, 1977.
20. D. B. Oliver and G. F. Knoll, *IEEE Trans. Nucl. Sci.*, **NS-15**: 122–126, 1968.
21. Z. H. Cho, I. Ahn, and C. M. Tsai, *IEEE Trans. Nucl. Sci.*, **NS-21**: 218–226, 1974.
22. B. W. Fox, *Techniques of Sample Preparation for Liquid Scintillation Counting*, Amsterdam: North-Holland, 1976.
23. M. G. Schorr and F. L. Torney, *Phys. Rev.*, **80**: 474, 1950.
24. G. Bendiscioli et al., *Nucl. Instrum. Meth.*, **A206**: 471–476, 1983.
25. M. F. Lucas, Improved low-level alpha scintillation counter for radon, *Rev. Sci. Instrum.*, **28**: 680–684, 1957.
26. M. Lenzen and H. J. Neugebauer, A theoretical investigation in the Lucas cell, *Nucl. Instrum. Meth. Phys. Res. A*, **368**: 479–483, 1996.
27. H. Woith et al., On the feasibility of monitoring radon in soil gas and groundwater as a precursor to earthquakes, in J. Zschau and O. Ergünay (eds.), *Turkish-German Earthquake Research Project*, Kiel, Germany: Kiel University, 1989, pp. 112–129.
28. J. W. Thomas and R. J. Countess, Continuous radon monitor, *Health Phys.*, **36**: 734–738, 1979.
29. M. Lenzen and H. J. Neugebauer, An automatic radon sensor for borehole measurements, *Rev. Sci. Instrum.*, **68** (7): 2898–2903, 1997.
30. W. E. Burcham and M. Jobes, *Nuclear and Particle Physics*, Harlow, Essex: Longman Scientific & Technical, 1994.
31. E. Kowalski, *Nuclear Electronics*, Berlin: Springer-Verlag, 1970.
32. F. Sauli (ed.), *Instrumentation in High Energy Physics*, Singapore: World Scientific, 1992.
33. D. F. Anderson, *Nucl. Instrum. Meth.*, **A287**: 606–612, 1990.
34. M. Kobayashi et al., A beam test on a fast EM-calorimeter of gadolinium silicate, GSO(CE), *Nucl. Instrum. Meth.*, **A306**: 139–144, 1991.
35. C. Zorn, Progress in the design of a radiation-hard plastic scintillator, *IEEE Trans. Nucl. Sci.*, **NS-38**: 194–199, 1991.
36. T. Hasegawa et al., Tests of new polystyrene-based scintillators, *Nucl. Instrum. Meth.*, **A311**: 498–504, 1992.
37. R. C. Ruchti, The use of scintillating fibers for charged-particle tracking, *Annu. Rev. Nucl. Part. Sci.*, **46**: 281–319, 1996.
38. D. Adams et al., Scintillating pad detectors, *IEEE Trans. Nucl. Sci.*, **NS-44** (3): 455–459, 1997.

MANFRED LENZEN

The University of Sydney

GEORGE BRAOUDAKIS

Australian Nuclear Science and
Technology Organisation

THIN FILM ANALYZERS

The term *thin films* has been applied to coatings whose thicknesses range from less than an atomic monolayer to films on the order of 1 mm thick. In addition to their well-known use as decorative or protective finishes, thin films of various types have found extensive use in the fields of electronics and optics.

This article describes a number of different thin film analyzers. Stated properly, since each film analyzer uses a number of different instruments, this article treats a number of different techniques of film *analysis* (i.e., methods and instruments used to determine experimentally such parameters as film thickness, structure, optical behavior, and material content of various types of films).

A very important application of thin films is in the field of semiconductor electronics. Although semiconductor circuits are generally fabricated on chips that are on the order of 0.5 mm thick, it is a combination of thin layers at the top of the chip that performs the electronic functions. Though structurally integral with the rest of the chip, these active layers can therefore be considered to be a set of thin films. Consequently, experimental methods of determining the structural and material content of such top layers are also those used in the analysis of thin films.

OPTICAL METHODS OF ANALYSIS

Reflectivity

The partial reflection of light at the interface between two materials of different index of refraction permits the measurement of the thickness of a transparent film. Specifically, when light of a range of wavelengths from a source in air strikes the front surface of a film of index of refraction n_1 and thickness t_1 , some of this light is reflected, and the rest is transmitted. Similarly, if the slab is either free standing or mounted on a medium of refractive index different from n_1 , some of the transmitted light is reflected from the back surface, reaches the front surface, and is again partially reflected and partially transmitted toward the source. This process is repeated ad infinitum, so that the total amount of light reflected toward the source is the sum of these individual waves, each with a phase determined by the index of refraction of the film, its thickness, and the frequency of the light. The result of this interference effect is a reflectance (i.e., the ratio of reflected to incident light), with peaks at frequencies governed by n_1 , t_1 , and the angle of incidence of the

incoming light. The film thickness can be extracted by determination of several of these peak frequencies (1,2). By using a white light source and dispersing the reflected light with a spectrometer unto a diode array, the technique can be set up for automatic data acquisition (3,4).

Alternative methods exist, such as the use of interference of light reflected from a variable thickness reference with that from the film under test, the use of a Fourier transform spectrometer, and the use of a monochromatic light source with variable angle of incidence (5–7).

For visible light, the interference technique is effective for films of thickness greater than 2000 Å.

Because of the relation of the color of light to its wavelength, the interference technique is well known in the semiconductor industry, where the thickness of a film of SiO₂ or Si₃N₄ on a silicon wafer can be easily determined by observation of the color of the light reflected from the wafer.

Transmission

Because of wave interference effects, the transmittance of a free-standing transparent film oscillates with wavelength λ . Maximum transmission for normal incidence of light occurs when the thickness of the film is an integral number of half-wavelengths (in the film). As a result, for transparent films it is found that the film thickness t_1 is given by $t_1 = (1/2n_1)(1/\lambda_a - 1/\lambda_b)^{-1}$, where λ_a and λ_b are free-space wavelengths corresponding to two adjacent maxima of transmission in the plot of transmittance versus λ (1).

If the film contains materials that absorb energy at certain frequencies within the spectrum of electromagnetic waves, these absorption peaks manifest themselves as attenuation peaks of the transmitted signal. *Transmission spectroscopy*, a technique that is very popular in the chemical industry, is the method of detecting the presence and quantity of such materials by finding the presence and magnitude of such absorption lines. Methods for converting the attenuation to the density of the absorbing material are given in the literature.

While absorption lines can be found by scanning through the spectrum with a broad spectrum source and a monochromator, a preferred method, with a great improvement in signal-to-noise ratio, is Fourier transform infrared spectroscopy (FTIR), which uses a broad spectrum source (8–10). The spectral distribution of the transmitted signal is found by taking the fast Fourier transform of the interferogram that is obtained when the transmitted signal is used as the source of a Michelson interferometer. (The interferogram is the signal output versus location of the movable mirror of the interferometer.)

Ellipsometry

The technique of ellipsometry is suitable to measurement with great precision of the thickness and index of refraction of thin, nonabsorbing films on highly absorbing substrates.

In basic ellipsometry, a beam of light that is linearly polarized at an angle with respect to the plane of incidence impinges on the film/substrate system. The components that are vertically and horizontally polarized with respect to the plane of incidence are reflected with different intensities and phases, thereby combining into an elliptically polarized beam. If the complex index of refraction of the substance is known, the reflection coefficients (i.e., the complex ratio of reflected to incident electric field for each of the two polarizations) can be used to compute the index of refraction and thickness of the film.

Films as thin as 10 Å, as well as others thousands of angstroms thick, have been measured by this ellipsometric method (11,12).

A convenient method of ellipsometry uses calibrated computer-controlled stepping motors to vary polarizations until a null output is found. Another system, the rotating analyzer ellipsometer, has high speed and high accuracy but makes more stringent demands on the optical system (13,14).

In a dynamic application of the technique, ellipsometry carried out at several frequencies has been used to monitor and control growth rate, layer thickness, substrate temperature, and ternary alloy composition in a III–V semiconductor molecular beam epitaxy system (15).

STRUCTURE

Profilometry

For some applications of thin films it is important to know the degree of flatness or the contours of a surface. An instrument used to carry out such measurements is called a profilometer.

Profilometer. A mechanical instrument that has been in use for a number of years plots a contour map by running a stylus along a surface with constant force. As an example of the performance of such an instrument, following are specifications given for the Dektak IIA Surface Profile Measuring System, distributed by Sloan Technology Corp., Santa Barbara, California:

Vertical resolution: 5 Å
Horizontal resolution: 500 Å
Stylus force: 50 mg (adjustable)
Scan length: 25 μ m to 30 mm

Scanning Tunneling and Atomic Force Microscopes. A great advance in resolution occurred with the invention of the scanning tunneling microscope (STM) (16) and atomic force microscope (AFM) (17). (The latter is also known as scanning force microscope.)

In the STM a sharp metal tip is scanned a few angstroms above an electron-conducting surface. With a few volts applied between tip and surface, an electron tunneling current flows, of magnitude exponentially dependent on the separation. As the tip is scanned over the surface, a feedback circuit maintains constant tunneling current by adjusting the tip height via a piezoelectric transducer on which the tip is mounted. A plot of the voltage applied to the transducer results in a contour image that has atomic resolution.

Alternatively, with the feedback turned off and scanning performed at constant height, variations in the tunneling current can result in an atomic resolution "current image."

Differences in the contour heights can be revealed to better than 1/100 of an atomic diameter. The lateral resolution is determined by the radius of protrusions at the tip. (A tip terminating in a single atom will have atomic lateral resolution.)

Since an STM yields a very detailed image of a surface, it can be used for surface metrology, such as measuring the line widths of lithographic patterns.

An atomic force microscope is an ultrasensitive profilometer that can produce contour images of both conducting and insulating surfaces with atomic resolution. A tip of a material such as silicon or silicon nitride that has been fastened to

the end of a cantilever of a light material, such as aluminum foil, is placed in contact with the surface to be scanned. As the surface is moved under the tip, the cantilever deflects with the topography. In one configuration, a laser beam reflecting off the cantilever falls on a two-element photodetector, which provides a feedback signal to a piezoelectric transducer that controls the tip height so as to maintain a constant (small) force between tip and surface. A plot of the piezoelectric voltage necessary to maintain this constant force is a contour map of the surface, with atomic resolution.

Extensions of the STM and AFM have resulted in a magnetic force microscope that is capable of mapping the domain pattern of thin magnetic films (18) and in an electrostatic force microscope that can measure potential or doping variations in semiconductor circuits (19–21).

Low-Energy Electron Diffraction

Low-energy electron diffraction (LEED) uses the wave nature of electrons to reveal the crystal structure of the very surface of a film (22). In LEED an electron beam of energy range 20 eV to 300 eV and of narrow energy spread is shot normally at the film. A fraction of the incident electrons is backscattered (reflected). Those that are backscattered elastically from the top few atomic layers form an electron diffraction pattern that reveals the geometry of the surface. To display this pattern in a LEED instrument, all of the backscattered electrons are first caused to pass through a set of retarding grids, to remove those that have lost some energy. The remaining ones are accelerated, to strike a fluorescent screen and display the desired diffraction pattern. Since changes in the surface, such as by adsorption of foreign molecules, often introduce new periodicities in the surface placement of atoms, these changes can be detected by observation of the displayed pattern.

Since even at a relatively good vacuum of 10^{-6} torr a surface will become contaminated in seconds, LEED measurements must be carried out with the surface that is being probed in an ultrahigh vacuum (10^{-10} torr or better).

Reflection High-Energy Electron Diffraction

In reflection high-energy electron diffraction (RHEED), also called HEED, electrons of energy of around 10 keV to 100 keV are shot at a surface at a glancing angle. The resulting diffraction pattern of these forward-scattered electrons provides information about the smoothness of a surface. Since electrons are scattered forward, it is possible in some cases to use RHEED for continuous monitoring during controlled growth of a film, such as in the deposition of thin films by molecular beam epitaxy (23).

Scanning Electron Microscopy

In scanning electron microscopy (SEM) an electron beam of approximately 10 Å to 50 Å diameter is scanned across the surface of the film that is to be analyzed. The incident electrons cause secondary electrons to be emitted. By collecting these secondaries or the backscattered incident primary electrons and displaying their intensity on a monitor, an image that relates to the film's surface structure is obtained (24–26). Since an electron's DeBroglie wavelength can be much smaller than wavelengths of visible light, the so-called far field resolution of an electron microscope is much finer than that of an optical microscope. (The wavelength of a 10 keV electron is $0.000012\ \mu\text{m}$,

while the wavelength of visible light is in the range $0.4\ \mu\text{m}$ to $0.7\ \mu\text{m}$.)

Transmission Electron Microscopy

In transmission electron microscopy (TEM), electrons are accelerated to energies of 100 keV to 500 keV, for transmission through and imaging of structures of films of thickness from a single monolayer up to a few thousand Ångströms (27). TEM is therefore analogous to optical transmission microscopy (OTM), but with a higher degree of resolution. However, unlike in OTM, where images are formed principally because of absorption, in TEM image formation is due to electron scattering or diffraction by the film.

The electrons that pass through a film in TEM form a diffraction pattern in the so-called focal plane and an image in another plane, the corresponding image plane of the electron lens system. Both can be displayed on a fluorescent screen with additional electron lenses, to provide information on crystalline phases, amorphous regions, and crystalline imperfections.

To obtain detailed local information, electrons are focused to a small spot and scanned across the film. This method is thus the transmission equivalent of SEM.

MATERIAL ANALYSIS

The preceding text has discussed what can be called the mechanical structure of a film or its surface. The following deals with determining the material composition or chemical structure of films.

Photoluminescence

Photoluminescence (PL) is a nondestructive technique that can detect the presence of certain materials, such as impurities in semiconducting films (28). To carry out PL, the specimen of interest is cooled to near liquid helium temperature and illuminated with light whose photon energy exceeds the semiconductor's bandgap, to create electron-hole pairs. When recombination takes place, the determination of the wavelengths of the recombination radiation produced identifies the impurities.

Raman Spectroscopy

In the analysis technique based on the Raman effect, a film is illuminated with a laser, the "pump," and the radiation emitted by the sample is examined for the presence of light whose frequencies differ slightly from the pump frequency (29). The amounts of the frequency shifts between the laser light and the emitted light, which are due to the parametric conversion of some of the pump energy into energy of phonons, reveal information about content and structure of the film. Spectroscopic examination of the emitted light and comparison to known shifts allow the determination of composition, degree of crystallinity, strain, alloy fluctuations, and other imperfections, such as damage due to ion implantation. This technique has been found most successful in detecting organic contaminants and in detecting stress in materials.

Electron Beam Analyzers

Auger Electron Spectroscopy. In Auger electron spectroscopy (AES), electrons of 1 keV to 25 keV are injected into the surface of a film, and the kinetic energies of electrons emitted from the

surface are measured, to form a plot of $N(E)$ versus E . Here E is the electron energy and $N(E)$ is the number of electrons emitted per unit energy increment per second. The display of the derivative of this plot, $d[N(E)]/dE$, will generally have one or more outstanding features, such as nearly vertical sharp lines. These are due to the so-called Auger electrons that are characteristic of and reveal the presence of elemental species (identified by the value of energy at the position of the sharp line) lying within the first 50 Å of the surface (30,31). In addition, slight shifts in the positions from those expected for elemental atoms or changes in line shape indicate chemical binding information.

To determine concentrations of the various species, peak-to-peak heights of the features in the derivative spectrum are used, corrected for elemental sensitivity factors.

Several ways of measuring the energy of collected electrons can be used. A preferred instrument is the cylindrical mirror analyzer spectroscopy (CMA). Its basic structure consists of segments of two coaxial cylinders, across which a variable dc voltage is applied, to select electrons of the desired energy that are to be collected (32).

Depth profiling can be carried out by alternating Auger analysis with removal of layers of material by sputter etching with an inert ion beam. Scanning AES has been carried out with electron beams focused into spots as small as 130 Å in diameter. All elements other than hydrogen and helium can be detected.

Electron Microprobe/Electron Probe Microanalysis. In electron microprobe analysis (EMP), electrons of energy 5 keV to 20 keV are injected into the surface, and identification of the elemental species is by measurement of the spectrum of emitted X-rays (33,34). In addition to a continuum in this spectrum one finds one or more sharp X-ray "lines." The latter are characteristic of the elements emitting them. For example, the presence of copper is revealed by its K-alpha line of energy 8.048 keV, corresponding to a wavelength of 1.54 Å.

Two types of X-ray spectrometers are in use: (1) energy dispersive (EDS) and (2) wavelength dispersive (WDS). Energy analysis in EDS is by a cooled lithium-drifted silicon diode that generates electron-hole pairs (ehp's). These are converted into voltage pulses, which are stored in the "bins" of a multichannel analyzer. Since the number of ehp's generated per X-ray pulse is a function of its energy, the output of the multichannel analyzer is the desired spectrum.

In WDS one or more crystals diffract(s) the X-rays to angles determined by their wavelengths. Detection is by a Geiger-counter-type of gas-filled tube.

EDS is more rapid than WDS and is therefore used more frequently for quick surveys or spatial maps. However, it has poorer energy resolution (e.g., 150 eV versus 5 eV).

Spatial resolution of EMP is on the order of 1 µm to 10 µm. While therefore not used for detecting surface contaminants, it is well suited for locating impurities within a film and determining alloy composition.

Electron Energy Loss Spectroscopy. The transmitted electrons in scanning transmission electron microscopy (STEM), in which a fine electron beam is raster-scanned across a film, can be analyzed for energy loss by sensing the current collected for different retarding voltages applied to the electrode that collects the transmitted electrons. This technique, electron energy loss spectroscopy (EELS), has been used to locate precipitates con-

taining elements that are not readily located by other techniques such as EMP (35,36).

Secondary Ion Mass Spectrometry/Sputtered Neutral Mass Spectrometry

The term *secondary* refers to particles emitted from a surface because this surface has been struck by other, or "primary," particles.

In contrast to most other methods of analysis in use, secondary ion mass spectrometry (SIMS) and secondary neutral mass spectrometry (SNMS) are destructive methods of analysis, for in these techniques ions of energy in the 1 keV to 20 keV range bombard and erode the surface of a film. Most of the products of this erosion are electrically neutral particles; a small fraction is ionized.

In SIMS these ions are sent through a mass spectrometer that reveals their charge-to-mass (Q/M) ratio and therefore identifies the composition of the emitted material. SNMS makes use of the neutral particles by ionizing them either with a strong laser beam or by sending them through a plasma chamber prior to directing them to the mass spectrometer. Progressive erosion and analysis provide information of relative density of the elements and isotopes versus depth of the film (37-43).

In an instrument called an *ion microprobe*, spatial analysis is carried out by focusing the bombarding beam to a small spot and scanning the sample, while simultaneously displaying on a screen a signal indicative of the spatial distribution of the relative density of a particular ion. To carry out the ion analysis, electrostatic quadrupole ion spectrometers, magnetic sector spectrometers, and time of flight tubes are in use. The quadrupole spectrometer comprises four parallel rods to which both a dc voltage V_{dc} and an ac voltage V_{rf} of a certain frequency are applied. Ions of a desired Q/M that will pass through the spectrometer are selected by adjustment of V_{dc} , V_{rf} , and the frequency.

In the *ion microscope*, a primary ion beam is used to flood a surface. The emitted secondary ions are passed through a "magnetic sector spectrometer." This unit preserves information of their spatial distribution while also permitting only ions of the Q/M selected by the intensity of an applied magnetic field to reach a so-called microchannel plate detector. The electric output signal of this distributed detector is used to generate a display on a fluorescent screen that is a map of the spatial distribution of secondary ions of the selected charge-to-mass ratio emitted by the surface.

To carry out depth profiling, in *static SIMS* a sputtering rate of about 1 Å/h will generally record a complete mass spectrum of the top 5 Å of a film. In *dynamic SIMS* a depth profile of the relative content of a particular ion is carried out at the rate of around 10 µm/h.

Lateral resolution of SIMS is generally around 100 µm, though it can be made as small as 1 µm. Depth resolution is on the order of 50 Å to 100 Å.

Rutherford Backscattering Spectrometry

Another technique for determining atomic composition and depth information is Rutherford backscattering spectrometry (RBS), also called high-energy (back)scattering ion spectrometry (HEIS). In RBS a film is bombarded with a beam of 1 MeV to 3 MeV He^+ or He^{++} ions, generally accelerated to this energy by

a tandem accelerator, and an energy analysis is carried out on the small fraction that undergoes elastic collisions with atomic nuclei within the film and are returned with a velocity that has a component in direction opposite to the primary beam (37–44).

The energy of a backscattered ion can be measured by a silicon surface barrier detector, which generates voltage pulses of amplitude proportional to its energy. These pulses are fed to a multichannel analyzer, which stores them in different “bins,” depending on their height. The number of pulses in a particular bin is an indication of the relative density of the particular species of atom near the surface of the film and permits calculation of the actual area density.

For atoms on the surface of the film, the ratio of the energy of a backscattered ion to the energy it had when incident on the film at a specified angle depends only on the ratio M_1/M_2 . Here M_1 is the mass of the bombarding ion, and M_2 is that of the nucleus of the atom in the film that caused it to be backscattered. Ions backscattered by atoms below the surface generally lose some energy during their passage to the surface. When these ions leave the surface, the curve of their yield versus energy is a series of quasi-rectangles, whose positions and widths provide information of the depth below the surface at which the detected atoms that backscattered the incident ions are located. It has been found that for film thicknesses less than 2000 Å the typical depth resolution is around 100 Å to 200 Å.

Applications of RBS are measurements of film thickness, uniformity, stoichiometry, and distribution of impurities. RBS has been found particularly suitable for detecting heavy elements in an environment of light elements. RBS can also be useful in determining crystallinity, through the phenomenon of *channeling*. For when a crystal is aligned so that an incident ion passes through it without colliding, the backscattering yield can decrease by as much as two orders of magnitude.

X-Ray Fluorescence

X-ray fluorescence (XRF), also known as X-ray fluorescence spectroscopy (XRFS), X-ray fluorescence analysis (XRFA), and X-ray secondary emission spectroscopy, is a nondestructive technique that can rapidly give average sample composition, including impurities. It can also be used for measuring film thickness (as thin as 100 Å) by comparing the intensity of X-ray fluorescence with that from a known standard (45,46).

In XRF a primary X-ray source illuminates the specimen, causing secondary X rays of lower energy to be emitted. As in EMP, the energy of the emitted X rays can be measured by either EDS or WDS systems. The energies identify the emitting substances; the relative intensities of emitted X-ray lines provide the information for calculating the concentrations of the emitting species. XRF can analyze areas as small as 10^{-6} cm².

X-Ray Photoelectron Spectroscopy

X-ray photoelectron spectroscopy (XPS), also known as electron spectroscopy for chemical analysis (ESCA), is a method for identifying chemical (not just elemental) species within a film (47–49). It is based on the well-known photoelectric effect, except that the incident radiation used is in the X ray, not the visible range of the electromagnetic spectrum.

In XPS a source of 1 keV to 2 keV X rays of narrow line width, obtained by using light elements (Al, Mg) as targets in the X-ray tube, illuminates a film, causing electrons to be

ejected. Since the kinetic energy of an ejected electron depends on the energy that bound the electron to the atom whence it originated, measurement of this kinetic energy reveals the type of atom or compound of origin. The resulting curve of collected electron current versus energy provides the information of the density of the various chemical species present.

Since the “escape depth” of electrons is around 50 Å, XPS is a surface analysis technique. By combining XPS with sputter erosion of a surface, a film may be depth profiled.

An advantage of XPS over methods based on the use of electron beams is that there is less danger of charging a surface with X rays, so that XPS is often a preferred choice for analysis of insulators.

X-Ray Topography

In X-ray topography (XRT) X rays are used to detect faults in the structure of single-crystal wafers or films. Basically, the film is placed at an angle with respect to the X-ray beam, to form a diffraction pattern on a photographic plate. Both film and plate are now moved perpendicular to the direction of the X-ray beam. A change in the diffraction pattern indicates a change in the crystal structure (50,51).

Neutron Activation

Neutron activation analysis is a method that can be used to find certain trace elements in a film. The method consists of placing the film into a nuclear reactor for several hours, to allow the nuclei of the trace elements to reach an excited state. After removal from the reactor, these excited nuclei “relax” by giving off gamma rays. Measurement of the energy and intensity of these with a germanium detector provides information on the presence and density of the atoms of these trace elements.

For this method of analysis to be successful, it is necessary that the host elements of the film have a half-life that is much shorter than that of the trace elements (52). An example of the use of the neutron activation technique is the detection and measurement of gold in silicon.

MACROSCOPIC FILMS

Although the term *thin film* generally refers to films whose thickness is a few microns or less, it has also been used to describe films of thickness on the order of 1 mm. Such films can be analyzed by the use of some of the techniques discussed in the previous sections, as well as by common microscopic and macroscopic techniques. Very rapid measurement of the height of a macroscopic film of liquid or solid, on the other hand, may require some special techniques, such as those discussed next.

Ultrasonic Systems

The methods of using mechanical waves of ultrasonic frequencies to measure the thickness of thin sheets or thick films are based on either “resonance” or “time of flight” measurements (53).

Techniques based on resonance utilize the fact that when an acoustic (ultrasonic) transducer is coupled to the sheet whose thickness is to be measured and the electrical input signal supplied to the transducer is swept through a frequency range, the electrical input impedance “seen” by the transducer exhibits resonance effects. If, for example, frequencies f_n and f_{n+1} are

two adjacent resonances of the same type, the thickness of the sheet is found to be $c/2\Delta f$, where c is the velocity of the ultrasonic wave in the material and Δf is the difference in frequency of the two adjacent resonances.

In time-of-flight-based instruments an ultrasonic transducer is mounted to send ultrasonic pulses through the film. If the velocity of propagation in the film is known, a measurement of the time interval between echo pulses received from the discontinuities of ultrasonic characteristic impedance at the bottom of the film and at the top of the film reveals the film thickness.

Microwave Microstrip Method

A noncontacting system for measuring the height of a film on a metal plate is based on the fact that v , the velocity of propagation of microwaves along a transmission line, depends on the dielectric constant of the medium in the line. The effective dielectric constant of a modified microstrip transmission line that has the metal plate as one conductor and a metal strip some distance above the film as the second conductor depends on the height of the film. Consequently, the determination of v by measurement of the wavelength at a fixed frequency or of the resonant frequency of a section of this line, after calibration with a "standard," provides information on the film height. In a refinement that eliminates the need for an external source of microwave energy, a section of this modified microstrip line is used to control the frequency of a transistor oscillator, thereby creating a compact, inexpensive instrument (54).

Fluorescent Dye System

To measure the instantaneous local height of a film of water, a fluorescent dye is mixed into the water. The dye is excited with a light source, causing it to fluoresce. The height of the film is determined by measuring the relative amount of light emitted from the film (55).

BIBLIOGRAPHY

1. H. Anders, *Thin Films in Optics*, London: Focal Press, 1967.
2. W. R. Runyan, *Semiconductor Measurements and Instrumentation*, New York: McGraw-Hill, 1975.
3. P. Burggraaf, How thick are your thin films?, *Semicond. Int.*, **11** (10): 96–103, 1988.
4. J. R. Sandercock, Film thickness monitor based on white light interference, *J. Phys. E.: Sci. Instrum.*, **16**: 866–870, 1983.
5. W. E. Beadle, J. C. C. Tsai, and R. D. Plummer, *Quick Reference Manual for Silicon Integrated Circuit Technology*, New York: Wiley-Interscience, 1985, pp. 4–23.
6. W. A. Pliskin and E. E. Conrad, Nondestructive determination of thickness and refractive index of transparent films, *IBM J. Res. Develop.*, **8**: 43–51, 1964.
7. W. A. Pliskin and R. P. Resch, Refractive index of SiO_2 films grown on silicon, *J. Appl. Phys.*, **36**: 2011–2013, 1965.
8. G. Horlick, Introduction to Fourier transform spectroscopy, *Appl. Spectrosc.*, **22**: 617–626, 1968.
9. R. J. Bell, *Introductory Fourier Transform Spectroscopy*, New York: Academic Press, 1972.
10. W. D. Perkins, Fourier transform-infrared spectroscopy, *J. Chem. Educ.*, **63**: A5–A10, 1986.
11. R. H. Muller, Principles of ellipsometry, *Adv. Electrochem. Electrochem. Eng.*, **9**: 167–226, 1973.
12. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, Amsterdam: North-Holland, 1987.
13. K. Riedling, *Ellipsometry for Industrial Applications*, Vienna: Springer, 1988.
14. D. E. Apnes and A. A. Studna, High precision scanning ellipsometer, *Appl. Opt.*, **14**: 220–228, 1975.
15. G. N. Maracas et al., In situ spectroscopic ellipsometry in molecular beam epitaxy for photonic devices, *Appl. Surf. Sci.*, **63** (1–4): 1–8, 1993.
16. G. Binnig et al., Surface studies by scanning tunneling microscopy, *Phys. Rev. Lett.*, **49**: 57–61, 1982.
17. G. Binnig, C. F. Quate, and C. Gerber, Atomic force microscope, *Phys. Rev. Lett.*, **56**: 930–933, 1986.
18. Y. Martin and H. K. Wickramasinghe, Magnetic imaging by "force microscopy" with 1000 Å resolution, *Appl. Phys. Lett.*, **50**: 1455–1457, 1987.
19. B. D. Terris et al., Localized charge force microscopy, *J. Vac. Sci. Technol. A*, **8**: 374–377, 1990.
20. S. A. C. Gould et al., From atoms to integrated circuit chips, blood cells, and bacteria with the atomic force microscope, *J. Vac. Sci. Technol. A*, **8**: 369–373, 1990.
21. J. Nogami, Sang-il-Park, and C. F. Quate, Behavior of Ga on Si (100) as studied by scanning tunnel microscopy, *Appl. Phys. Lett.*, **53**: 2086–2088, 1988.
22. K. Heinz, Structural analysis of surfaces by LEED, *Prog. Surf. Sci.*, **27**: 239–326, 1988.
23. B. F. Lewis et al., Reflection high energy electron diffraction intensity behavior during homoepitaxial molecular beam epitaxy growth of GaAs and implications for growth kinetics, *J. Vac. Sci. Technol. B*, **3**: 1317–1322, 1985.
24. J. I. Goldstein et al., *Scanning Electron Microscopy and X-Ray Microanalysis*, New York: Plenum, 1984.
25. M. von Heimendahl, *Electron Microscopy of Materials*, New York: Academic Press, 1980.
26. J. M. Cowley, *Diffraction Physics*, Amsterdam: North-Holland, 1975.
27. A. D. Romig, Analytical transmission electron microscopy, in R. E. Whan (coord. ed.), *Metals Handbook*, 9th ed., Vol. 10, Metals Park, OH: American Society of Metals, 1986, pp. 429–489.
28. P. J. Dean, Photoluminescence as a diagnostic of semiconductors, *Prog. Cryst. Growth Charact.*, **5** (1–2): 89–174, 1982.
29. S. Nakashima and M. Hangyo, Characterization of semiconductor materials by Raman spectroscopy, *IEEE J. Quantum Electron.*, **25**: 965–975, 1989.
30. P. Auger, On the compound photoelectric effect, *J. Phys. Radium*, **6**: 205–208, 1925 (in French).
31. L. A. Harris, Analysis of materials by electron-excited auger electrons, *J. Appl. Phys.*, **39**: 1419–1427, 1968.
32. H. Hapner, J. A. Simpson, and C. E. Kuyatt, Comparison of the spherical deflector and the cylindrical mirror analyzers, *Rev. Sci. Instrum.*, **39**: 33–35, 1968.
33. R. Castaing, Electron probe microanalysis, *Adv. Electron. Electron Phys.*, **13**: 317–386, 1960.
34. K. F. J. Heinrich, *Electron Beam X-Ray Microanalysis*, New York: Van Nostrand-Rheinhold, 1981.
35. C. Colliex, Electron energy loss spectroscopy in the electron microscope, *Adv. Opt. Electron Microsc.*, **9**: 65–177, 1986.
36. R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, New York: Plenum, 1986.
37. W. K. Chu et al., Principles and applications of ion beam techniques

- for the analysis of solids and thin films, *Thin Solid Films*, **17**: 1–41, 1973.
38. A. Benninghoven, Surface analysis by means of ion beams, *Crit. Rev. Solid State Sci.*, **6**: 291–316, 1976.
 39. C. W. Magee, Secondary ion mass spectrometry and its relation to high-energy ion beam analysis techniques, *Nucl. Instrum. Methods*, **191**: 297–307, 1981.
 40. C. C. Pantano, Secondary ion mass spectroscopy, in R. E. Whan (coord. ed.), *Metals Handbook*, 9th ed., Vol. 10, Metals Park, OH: American Society of Metals, 1986, pp. 610–627.
 41. M. T. Bernius and G. H. Morrison, Mass analyzed secondary ion microscopy, *Rev. Sci. Instrum.*, **58**: 1789–1804, 1987.
 42. A. Benninghoven, F. G. Rüdenauer, and H. W. Werner, *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications and Trends*. New York: Wiley, 1987.
 43. W. K. Chu, J. W. Mayer, and M.-A. Nicolet, *Backscattering Spectroscopy*, New York: Academic Press, 1978.
 44. W. K. Chu, Rutherford backscattering spectrometry, in R. E. Whan (coord. ed.), *Metals Handbook*, 9th ed., Vol. 10, Metals Park, OH: American Society of Metals, 1986, pp. 628–636.
 45. E. P. Bertin, X-ray secondary emission (fluorescence) spectrometry, General Introduction, Chapter 3, in *Principles and Practice of X-Ray Spectrometric Analysis*, New York: Plenum, 1972.
 46. J. V. Gilfrich, X-ray fluorescence analysis, in P. F. Kane and G. B. Larrabee (eds.), *Characterization of Solid Surfaces*, New York: Plenum, 1974, Chap. 12, pp. 275–306.
 47. C. Nordling, S. Hagström, and K. Siegbahn, Application of electron spectroscopy to chemical analysis, *Z. Phys.*, **178**: 433–438, 1964.
 48. C. D. Wagner et al., *Handbook of X-Ray Photoelectron Spectroscopy*, Eden Prairie, MN: Perkin Elmer, 1979.
 49. K. L. Smith and J. S. Hammond, Destructive and nondestructive depth profiling using ESCA, *Appl. Surf. Sci.*, **22/23** Part 1: 288–299, 1985.
 50. B. K. Tanner, *X-Ray Diffraction Topography*, Oxford, UK: Pergamon, 1976.
 51. R. N. Pangborn, X-ray topography, in R. E. Whan (coord. ed.), *Metals Handbook*, 9th ed., Vol. 10, Metals Park, OH: American Society of Metals, 1986, pp. 365–379.
 52. R. M. Lindstrom, Neutron activation analysis in electronic technology, in T. J. Shaffner and D. K. Schroder (eds.), *Diagnostic Techniques for Semiconductor Materials and Devices*, Pennington, NJ: Electrochemical Society, 1988, pp. 3–14.
 53. J. Szilard (ed.), *Ultrasonic Testing*, New York: Wiley, 1982.
 54. L. F. Root and I. Kaufman, Noncontacting low-cost instrument for film thickness measurement, *IEEE Trans. Instrum. Meas.*, **41**: 1014–1019, 1992.
 55. A. A. Boiarski, J. S. K. Liu, and R. P. Collier, Three-dimensional characterization of falling liquid film using laser induced fluorescence, *Symp. Polyphase Flow Transp. Technol.*, San Francisco, CA: 1980.
 - C. R. Bundle, C. A. Evan, Jr., and S. Wilson, *Encyclopedia of Materials Characterization: Surfaces, Interfaces, Thin Films*, Boston: Butterworth-Heinemann/Greenwich, CT: Manning, 1992.
 - L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and Thin Film Analysis*, New York: North-Holland, 1986.
 - I. Herman, *Optical Diagnostics for Thin Film Processing*, San Diego, CA: Academic Press, 1996.
 - D. R. Larson, *A Measurement for Determining the Optical and Electro-optical Properties of a Thin Film*, Boulder, CO: U.S. Dept. of Commerce, National Bureau of Standards, 1982.
 - L. I. Maissel and R. Glang (eds.), *Handbook of Thin Film Technology*, New York: McGraw-Hill, 1970.
 - E. M. Murt and W. G. Guldner (eds.), *Progress in Analytical Chemistry, Physical Measurements and Analysis of Thin Films*, New York: Plenum, 1969.
 - H. Oechsner (ed.), with contributions by H. W. Etzkorn et al., *Thin-film and Depth Profile Analysis*, Berlin and New York: Springer-Verlag, 1984.
 - R. A. Schmidt, *Growth and Applications of Thin Films*, Prague: Prometheus, 1994.
 - SIA (Periodical), *Surface and Interface Analysis*, Philadelphia, PA: Heyden, 1979.
 - H.-M. Tong and L. T. Nguyen (eds.), *New Characterizations for Thin Polymer Films*, New York: Wiley, 1990.

IRVING KAUFMAN
Arizona State University

RADIOMETRY

Radiometry is concerned with the measurement, analysis, and interpretation of radiant electromagnetic energy. The simplest and most common measurement, made using an instrument called a radiometer, is of received electromagnetic power in a given frequency band. Some instruments, however, analyze the received energy's spectral distribution (spectroradiometers), its polarization (polarimeters), or its angular distribution in space (imaging radiometers). The usual objective of radiometric measurements is the inference from them of the physical and/or chemical state of the material from, or through, which the received electromagnetic energy has been emitted, reflected, or transmitted. Radiometry is therefore a branch of remote sensing and could be considered to include radar. In practice, however, the term *radiometry* is usually used to describe the measurement and characterisation of naturally occurring radiation, or (unlike radar) scattered radiation originating as a narrowband signal. The former, called passive radiometry, deals principally with self-emitted thermal radiation and reflected radiation of solar origin. The latter, called active radiometry, relies on artificial illumination. Figure 1 illustrates these different types of radiometer. Active radiometers are sometimes called scatterometers.

Since all radiometers depend for their operation on the measurement of one or more parameters of electromagnetic radiation, a brief review of such radiation follows.

ELECTROMAGNETIC RADIATION

Whenever charges accelerate (other than those carried by electrons in stable orbits around the nuclei of atoms), or a subatomic particle in an atom makes a transition to a lower energy level,

Reading List

To prepare this article, extensive use was made of the material of Chapters 9 and 10 of the text by D. K. Schroder entitled *Semiconductor Material and Device Characterization*, New York: Wiley, 1990. The writer gratefully acknowledges the support and initial review of the manuscript by Dr. Schroder.

In addition to this text, as well as texts dealing with various specialized topics as listed in the Bibliography, material dealing with the analysis of thin films will be found in the following:

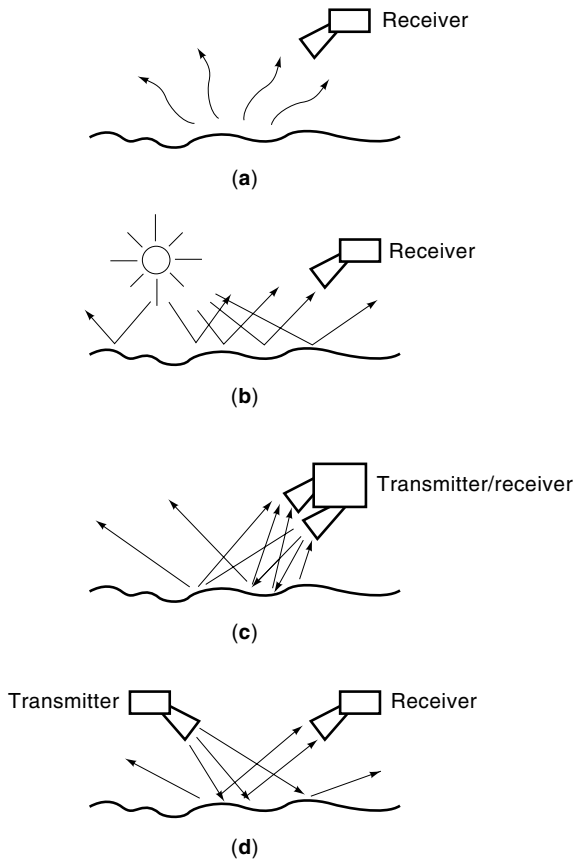


Figure 1. Passive and active radiometry. (a) Passive measurements of self-emitted thermal radiation. (b) Passive measurements of scattered solar radiation. (c) Active (scatterometer) measurements employing monostatic geometry. (d) Active (scatterometer) measurements employing bistatic geometry.

electromagnetic energy is radiated. This energy propagates as a disturbance of the electric and magnetic fields existing in space. An electromagnetic wave propagating in an isotropic, homogeneous, medium has, sufficiently far from its source, an electric field vector, $\mathbf{E}$, magnetic field vector, $\mathbf{B}$, and unit vector pointing in the direction of wave propagation, $\mathbf{u}$, which are mutually perpendicular. Furthermore, the wave front has a curvature that is sufficiently small for it to be considered locally plane. (In a truly plane wave $\mathbf{E}$ and $\mathbf{B}$ would be constant with respect to position in any plane normal to the direction of propagation.)

For radiation of a single frequency (said to be monochromatic), $\mathbf{E}$ and $\mathbf{B}$ will vary sinusoidally with time and space. The amplitude (V/m) of such a wave is the peak value of its electric field strength, and its cyclical frequency, f (Hz), is the number of oscillations (or cycles) that the electric field completes, at a given point in space, in one second. [The radian frequency, ω (rad/s), is the phase advance in radians, at a given point in space, that the wave makes in one second and is related to cyclical frequency by $\omega = 2\pi f$.] The phase constant, β (rad/m), is the phase advance in radians that the wave makes, at a given instant of time, over one meter of space in the direction of propagation and is related to wavelength, λ , by $\beta = 2\pi/\lambda$. ($1/\lambda$ and β are therefore the spatial equivalent of cyclical and radian frequency, respectively.)

Radiation of nonzero bandwidth has an amplitude at a point in space that fluctuates over time as the phase relationships between the constituent (and elemental) monochromatic waves change. The time scale over which this amplitude fluctuation occurs is related to the bandwidth, Δf , of the radiation and is characterized by a coherence time, Δt . Coherence time is defined as the maximum time shift for which the field of the original and time-shifted waves (at a given point in space) are correlated. An equivalent quantity, coherence length, Δl , is defined (for a given instant in time) as the maximum spatial shift for which the fields of original and shifted waves are correlated. For quantitative work the term *correlated* must be interpreted as a correlation value greater than some specified fraction of its maximum. For consistent definitions of bandwidth, coherence length, and coherence time, these quantities are related by

$$\Delta l = c \Delta t = \frac{c}{\Delta f} \quad (1)$$

If the tip of the $\mathbf{E}$ vector, representing a wave at a single point in space, traces out an elliptical curve, the wave is said to be (fully) polarized. If the curve traced out is random, the wave is said to be unpolarized. A wave with an $\mathbf{E}$ vector locus that can be resolved into a deterministic component and a random component is said to be partially polarized. A fully polarized wave having a polarization ellipse with an axial ratio of 1.0, or ∞ is said to be circularly, or linearly, polarized, respectively. A linearly polarized wave therefore has an $\mathbf{E}$ vector that traces out a straight line and that, in an isotropic medium, remains in a fixed plane as the wave propagates forward.

The concepts of coherence and polarization are not entirely unrelated since a wave observed on a time scale that is short with respect to its coherence time is always fully polarized. As the observation time increases, the polarization of the wave may gradually change; however, the changes appear random on a time scale large with respect to the wave's coherence time. A wave will remain fully polarized even when observed on a time scale much longer than its coherence time if energy is radiated in one polarization only.

In passive radiometry the received electromagnetic radiation is usually incoherent, spread across a wide band of frequencies, and unpolarized. The useful measurable quantities in this case are power density (W/m^2), radiation intensity (W/sr), power spectral density (W/Hz), and angular/spatial radiation distribution. In radiometry the first three of these quantities are generally referred to as irradiance (E), radiant intensity (I), and radiant spectral flux (Φ_f), respectively. In some applications, however (microwave radiometry, for example), different terminologies are also often used (see Table 1).

DEFINITIONS OF RADIOMETRIC QUANTITIES

The commonly used radiometric quantities are as follows:

Radiant energy, Q (J)—The quantity of electromagnetic energy in a specified region of space or associated with a specified process

Radiant flux, Φ (W)—Radiant energy per unit time crossing a specified surface

Radiant density, w (J/m^3)—Radiant energy per unit volume

Radiant flux density, E or M (W/m^2)—radiant flux per unit area normal to a specified direction

Table 1. Radiometric and Photometric Quantities

Quantity	Symbol	SI Unit	Other Common Units	Defining Equation	Comments
Radiant energy	Q_e	J	erg, kWh	E	Electromagnetic energy
Radiant flux	Φ_e	W	erg/s	$\Phi_e = \frac{dQ_e}{dt}$	Electromagnetic power
Radiant density	w_e	J/m ³	erg/cm ³	$w_e = \frac{dQ_e}{dV}$	Energy density
Irradiance	E_e	W/m ²	W/cm ²	$E_e = \frac{d\Phi_e}{dA}$	Power density (also called radiant flux density)
Radiant excitance	M_e	W/m ²	W/cm ²	$M_e = \frac{d\Phi_e}{dA}$	Power density (also called emittance and radiant flux density)
Radiant intensity	I_e	W/sr		$I_e = \frac{d\Phi_e}{d\Omega}$	Angular concentration of power (also called radiation intensity)
Radiance	L_e	W sr ⁻¹ m ⁻²	W sr ⁻¹ cm ⁻²	$L_e = \frac{dI_e}{(dA \cos \psi)}$	Angular concentration of power per unit projected area of source (also called radiometric brightness, especially in microwave applications)
Luminous energy	Q_v	lm s (talbot)	lm h	$\Phi_v = K_m \int_{0.36 \mu\text{m}}^{0.83 \mu\text{m}} \Phi_{e,\lambda}(\lambda) V(\lambda) d\lambda$	Photopic definition. Scotopic definition uses primed versions of K_m and $V(\lambda)$.
Luminous flux	Φ_v	lm (lumen)		$\Phi_v = \frac{dQ_v}{dt}$	
Luminous density	w_v	lm s m ⁻³		$w_v = \frac{dQ_v}{dV}$	
Illuminance	E_v	lx (lux)	fc, ph	$E_v = \frac{d\Phi_v}{dA}$	1 lux (lx) = 1 lm/m ² 1 foot-candle (fc) = 1 lm/ft ² 1 phot (ph) = 1 lm/cm ² Also called illumination
Luminous excitance	M_v	lx		$M_v = \frac{d\Phi_v}{dA}$	Also called luminous emittance
Luminous intensity	I_v	cd (candela)		$I_v = \frac{d\Phi_v}{d\Omega}$	1 candela (cd) = 1 lm/sr
Luminance	L_v	cd/m ² (nit, nt)	sb, L, fL, asb	$L_v = \frac{dI_v}{(dA \cos \psi)}$	1 stilb (sb) = 1 cd/cm ² 1 lambert (L) = 1 cd/(π cm ²) 1 foot lambert (fL) = 1 cd/(π ft ²) 1 apostilb (asb) = 1 cd/(π m ²) Also called photometric brightness

Irradiance, E (W/m²)—radiant flux density arriving at a specified surface

Radiant excitance (or emittance), M (W/m²)—Radiant flux density leaving a specified surface

Radiant intensity, I (W/sr)—Radiant flux per unit solid angle in a specified direction

Radiance, L (W sr⁻¹ m⁻²)—Radiant intensity in a specified direction per unit projected area of the source normal to the specified direction

Radiation consisting of electromagnetic power distributed over a continuum of frequencies has infinitesimally small

power at any single frequency. Each of the preceding quantities can be characterized at a single frequency, however, by considering the quantity to be measured within a small bandwidth Δf (centered on frequency f), dividing by Δf , and taking the limit as $\Delta f \rightarrow 0$. The result is a spectral density having those units of the original quantity per hertz. (An obvious variation on this definition uses radian frequency instead of cyclical frequency.) At optical frequencies spectral densities are usually expressed as quantities per meter of wavelength rather than per hertz of frequency. A spectral density is typically indicated by preceding the quantity with the word *spectral* and adding a subscript f , ω , or λ to the appropriate symbol. Spectral irradiance, for example, can be expressed in any of the following

ways:

$$E_f(f) = \lim_{\Delta f \rightarrow 0} \left[\frac{\text{power density contained in frequency band } f - \frac{\Delta f}{2} \text{ to } f + \frac{\Delta f}{2}}{\Delta f} \right] \quad (2)$$

$$E_\omega(\omega) = \lim_{\Delta \omega \rightarrow 0} \left[\frac{\text{power density contained in frequency band } \omega - \frac{\Delta \omega}{2} \text{ to } \omega + \frac{\Delta \omega}{2}}{\Delta \omega} \right] \quad (3)$$

$$E_\lambda(\lambda) = \lim_{\Delta \lambda \rightarrow 0} \left[\frac{\text{power density contained in wave band } \lambda - \frac{\Delta \lambda}{2} \text{ to } \lambda + \frac{\Delta \lambda}{2}}{\Delta \lambda} \right] \quad (4)$$

The relationship among these various expressions is summarized by

$$E_f(f) = E_\omega\left(\frac{\omega}{2\pi}\right) 2\pi = E_\lambda\left(\frac{c}{\lambda}\right) \frac{c}{f^2} \quad (5)$$

Spectral densities may be a function of frequency and, if so, may be measured by a spectroradiometer to yield information about the source of the radiation, material through which it has passed, or a surface from which it has been reflected. Often, however, spectral densities are constant with frequency (at least over the frequency band of interest). In this case they are referred to as white and are characterized by a single value.

Related Photometric/Luminous Quantities

Radiometric quantities relate to objective measures of energy and power carried by electromagnetic radiation of any wavelength. Photometric (or luminous) quantities relate to the subjective visual impact that these quantities have on an observer (1). In practice, photometric quantities are usually measured objectively but in such a way as to reflect the visual sensation experienced by a normal human observer. This is achieved by using detectors that mimic the response of the human eye/brain system and is called physical photometry. The definition of photometric quantities therefore involves the visual response of a (hypothetical) standard observer (2). In fact two standard responses, $V(\lambda)$ and $V'(\lambda)$, called the photopic and scotopic spectral luminous efficiencies, are defined (Fig. 2). These responses correspond closely to those of human observers under conditions of high (nominally daytime) and low (nominally nighttime) illumination, respectively. (The principal reason for the difference in these two responses is that different types of retinal receptors dominate under the two conditions; cones in the former case and rods in the latter case.)

Spectral luminous efficiencies are used to define a photometric/luminous quantity corresponding to each radiometric quantity, the same symbol usually being used for both. When con-

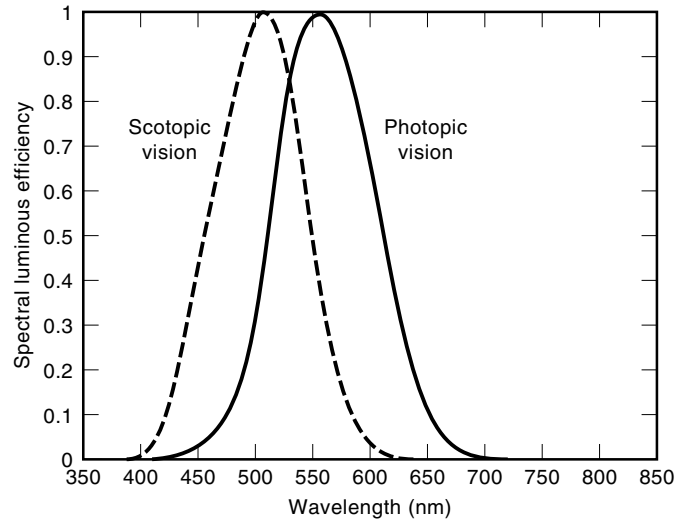


Figure 2. Spectral luminous efficiencies: photopic, $V(\lambda)$, and scotopic, $V'(\lambda)$.

fusion between the two is possible, the radiometric version is identified using a subscript e and the luminous version using a subscript v . For photopic vision, luminous flux, Φ_v , for example, is related to the radiometric quantity, radiant spectral flux $\Phi_{e,\lambda}(\lambda)$, by

$$\Phi_v = K_m \int_0^\infty \Phi_{e,\lambda}(\lambda) V(\lambda) d\lambda \quad (6)$$

[Note that since $V(\lambda)$ is zero outside the range $0.36 \mu\text{m}$ to $0.83 \mu\text{m}$, the limits of the integral in Eq. (6) can be replaced with these values.] The coefficient $K_m (= 683.002 \text{ lm/W})$ in Eq. (6) is a conversion factor that changes radiometric units of watts (represented by the integral) into photometric units of lumens (lm). The scotopic version of Eq. (6) has identical form but is distinguished by primes added to Φ_v , $V(\lambda)$, and K_m . The value of K'_m is 1700.06 lm/W (rods, which are more sensitive than cones, dominate vision under conditions of poor illumination). The curves $K(\lambda) = K_m V(\lambda)$ and $K'(\lambda) = K'_m V'(\lambda)$, called the photopic and scotopic spectral luminous efficiencies, (Fig. 3), intersect at a wavelength of $0.555 \mu\text{m}$ ($540 \times 10^{12} \text{ Hz}$), for which $V(\lambda)$ is a maximum (unity). At this frequency, therefore, the luminous efficiency is 683 lm/W under all illumination conditions. At other frequencies the luminous efficiency is illumination dependent. The wavelength of maximum visual response for scotopic vision is $0.507 \mu\text{m}$.

Table 1 summarizes the radiometric and the corresponding photometric quantities.

Photon Flux

At optical and quasi-optical frequencies, the interaction of electromagnetic fields with matter cannot always be explained in terms of smoothly varying field quantities, and recourse must be made to quantum theory. In this description electromagnetic fields are made up of discrete, quantized packets of energy called photons. The energy, Q (joules), carried by each packet, or quanta, is related to the radiation's frequency by $Q = hf$ (J), where $h = 6.626 \times 10^{-34} \text{ (J}\cdot\text{s)}$, is Planck's constant.

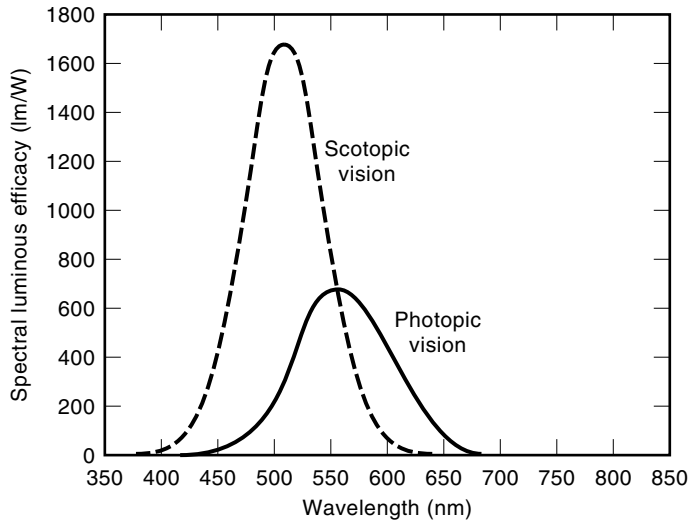


Figure 3. Spectral luminous efficacies: photopic, $K(\lambda)$, and scotopic, $K'(\lambda)$.

The probability that the energy carried by a photon will be transferred to a material particle (in a detector, for example) at some specific time and location is proportional to the irradiance (power density) at that time and location. In applications where individual photons may be detected, electromagnetic power is often referred to as a photon flux.

BLACKBODY RADIATION AND EMISSIVITY

A body that absorbs all electromagnetic radiation falling on it (converting the incident radiant energy to thermal energy) and that would therefore appear black at optical frequencies is called a blackbody. Such a body, in thermal equilibrium with its surroundings, must supply precisely as much energy back to, as it receives from, these surroundings. In the absence of conduction and convection the energy transfer mechanism is limited to thermal radiation. The thermal radiant energy emitted by a body can never, therefore, be greater than that emitted by a blackbody of the same size, shape, and physical temperature. The degree of blackness of a body is quantified by its surface emissivity, defined as follows.

Emissivity, ϵ —Ratio of spectral radiant exitance at the surface of a body to the spectral radiant exitance at the surface of a blackbody with the same physical temperature

The maximum value of emissivity, corresponding to that of a blackbody, is therefore unity and the minimum value, corresponding to a surface that reflects all radiant energy falling on it, is zero. A body with an emissivity that is independent of frequency, but less than unity, is called a gray body. Many natural materials have emissivities that are frequency dependent, however, and this dependence, if measured radiometrically, can yield useful information about the material's composition and physical state.

The reflecting, transmitting, and absorbing properties of a material's surface, which can also be inferred from radiometric measurements, are quantified as follows:

Reflectance, ρ —ratio of reflected to incident spectral radiant flux at a surface

Transmittance, τ —ratio of transmitted to incident spectral radiant flux at a surface

Absorptance, α —Ratio of absorbed to incident spectral radiant flux at a surface

It is clear that

$$\rho + \tau + \alpha = 1 \quad (7)$$

The surface of a blackbody has an absorptance of unity and, therefore, a reflectance and transmittance of zero. More generally, for any opaque ($\tau = 0$) material, $\rho = 1 - \alpha$ and emissivity is numerically equal to absorptance (i.e., $\epsilon = \alpha$).

A good model of a blackbody is the surface represented by a small hole drilled in a hollow cavity. Provided that the reflectance of the interior surface of the cavity is not too close to unity, then any radiant energy falling on to the hole is reflected a sufficient number of times inside the cavity for it to be fully absorbed before it can reemerge.

The spectral radiant exitance, $M_f(f)$, of a blackbody (Fig. 4), is given by Planck's law (3); that is,

$$M_f(f) = \frac{2\pi h}{c^2} \frac{f^3}{\exp(hf/kT) - 1} \quad (8)$$

For optical frequencies, where it is traditional to express spectral quantities on a per meter of wavelength, rather than a per hertz of frequency, basis Eq. (8) can be written as

$$M_\lambda(\lambda) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{\exp(hc/\lambda kT) - 1} \quad (9)$$

For $hf \ll kT$, which is appropriate for microwave radiometry at all commonly encountered temperatures, Eq. (8) reduces to the Rayleigh–Jeans approximation (Fig. 5). That is,

$$M_f(f) = \frac{2\pi}{c^2} kT f^2 \quad (10)$$

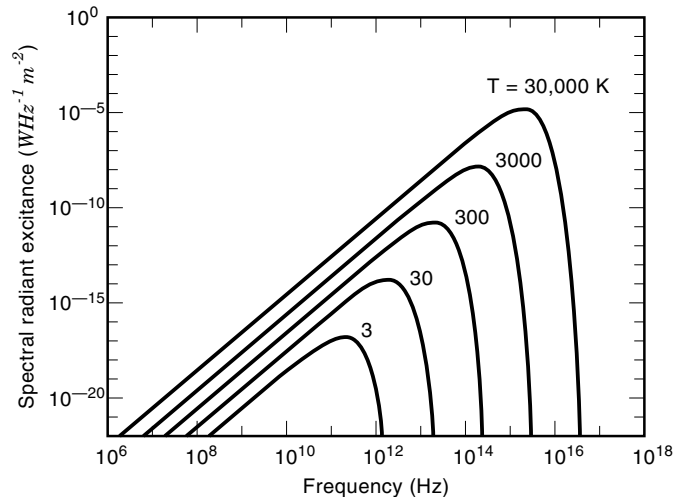


Figure 4. Planck's law for blackbody radiation.

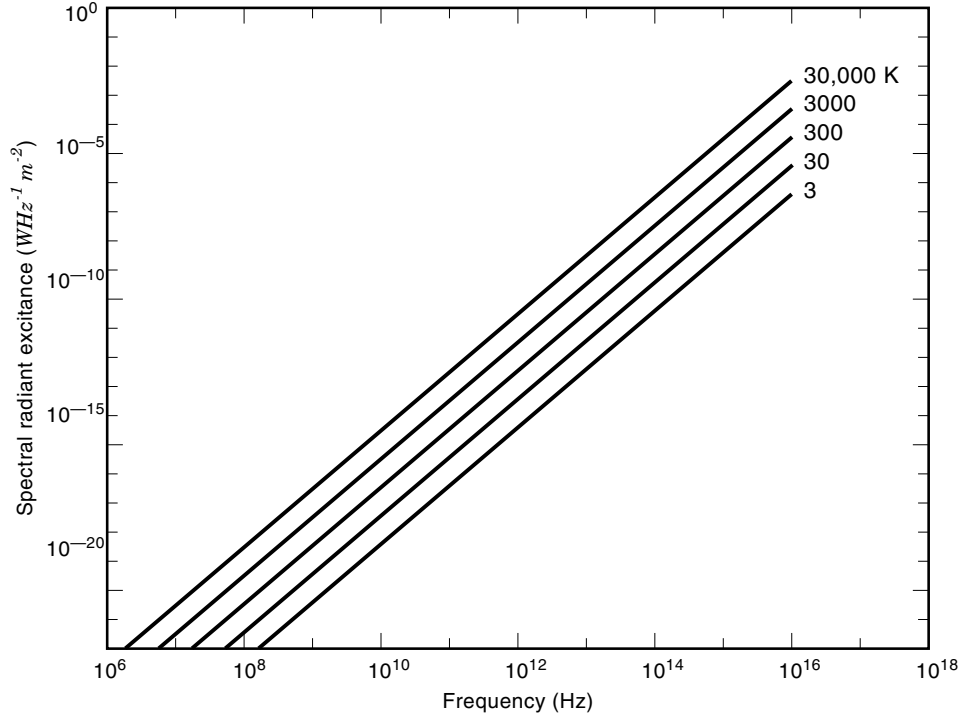


Figure 5. Rayleigh-Jeans approximation for blackbody radiation.

[The Rayleigh-Jeans version of Eq. (9) is $M_\lambda(\lambda) = 2\pi ckT/\lambda^4$.] The spectral radiant exitance of a body with emissivity < 1.0 is therefore given by

$$M_f(f) = \epsilon(f) \frac{2\pi}{c^2} kT f^2 \quad (11)$$

Lambert's Law and Lambertian Radiators

Lambert's cosine law is that the radiant intensity from a blackbody surface is proportional to $\cos \psi$ where ψ is the angle between the direction of observation and the surface normal. Any surface (irrespective of its emissivity) that has this distribution of radiant intensity is said to be Lambertian. The radiance of such a surface (being radiant intensity per unit projected area) is independent of ψ . The spectral radiant exitance of an arbitrary surface is related to its spectral radiance by

$$M_f(f) = \int_0^{2\pi} \int_0^{\pi/2} L_f(\psi, f) \cos \psi \sin \psi d\psi d\gamma \quad (12)$$

which, for a Lambertian surface, reduces to

$$M_f(f) = L_f(f) \pi \quad (13)$$

The spectral radiance of a blackbody surface is therefore given by

$$L_f(f) = \frac{2h}{c^2} \frac{f^3}{\exp(hf/kT) - 1} \quad (14)$$

which for $kT \gg hf$ gives the Rayleigh-Jeans approximation

$$L_f(f) = \frac{2kT}{c^2} f^2 = \frac{2kT}{\lambda^2} \quad (15)$$

Radiometric Brightness and Antenna Temperatures

Consider a lossless antenna (i.e., one having an ohmic efficiency of 100%) with effective aperture area $A_e(\theta, \phi)$ receiving radiant energy from a surface S (Fig. 6). If a surface element of area dS (m²) lies at a distance R (m) from the antenna in the direction (θ, ϕ) , then the spectral radiant flux density arriving at the

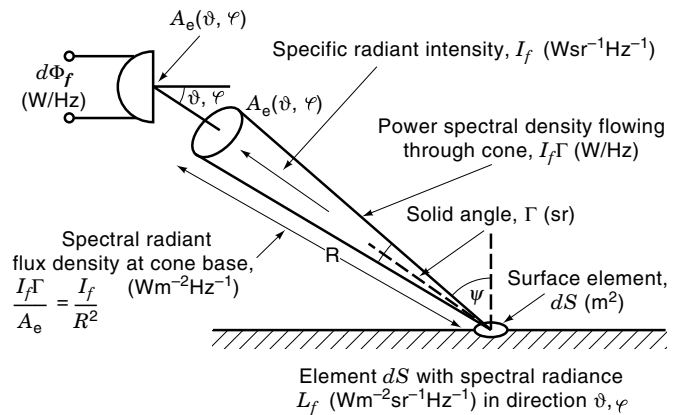


Figure 6. Spectral irradiance at radiometer antenna aperture due to radiating surface element dS .

antenna due to this element is

$$dE_f = \frac{I_f(\theta, \phi)}{R^2} = \frac{L_f(\theta, \phi) dS \cos \psi}{R^2} \quad (16)$$

The power spectral density (or spectral radiant flux) received by the antenna is therefore

$$d\Phi_f = 0.5 A_e(\theta, \phi) dE_f = \frac{0.5 A_e(\theta, \phi) L_f(\theta, \phi) dS \cos \psi}{R^2} \quad (17)$$

(The factor of 0.5 accounts for the fact that the antenna will be polarization matched to only half of the randomly polarized incident energy.) The quantity $dS (\cos \psi)/R^2$ in Eq. (17) can be identified as the solid angle, $d\Omega$, subtended by the surface element at the antenna (Fig. 7). The power spectral density available at the antenna output terminals can therefore be expressed as either a surface, or a solid angle, integral; that is,

$$\begin{aligned} \Phi_f &= 0.5 \iint_S \frac{A_e(\theta, \phi) L_f(\theta, \phi) \cos \psi dS}{R^2} \\ &= 0.5 \iint_{\Omega} A_e(\theta, \phi) L_f(\theta, \phi) d\Omega \end{aligned} \quad (18)$$

The power, or radiant flux, received within a band of frequencies $f_{\min}$ to $f_{\max}$ is then given by

$$\Phi = \int_{f_{\min}}^{f_{\max}} \Phi_f(f) df \quad (19)$$

Since Eq. (18) is a power spectral density, an antenna aperture temperature, T_a , may be defined such that the noise power measured by a noiseless radiometer (i.e., one that has a noise figure of 0 dB), with a lossless antenna, is $kT_a B_N$ (where B_N is the radiometer noise bandwidth and k is Boltzmann's constant). For small B_N , such that $\Phi_f(f)$ can be considered white,

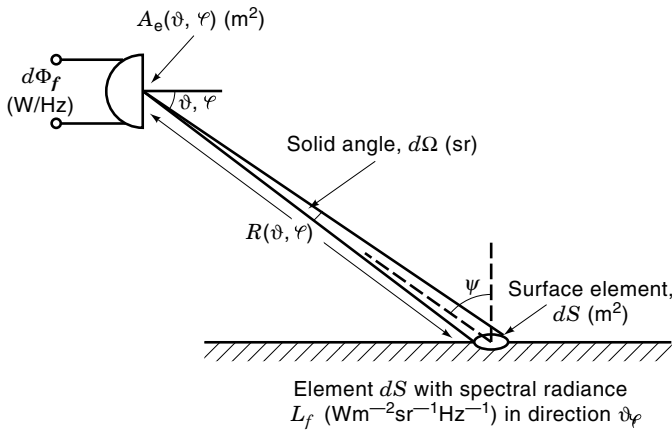


Figure 7. Solid angle subtended at radiometer antenna aperture by radiating surface element dS .

$T_a = \Phi/kB_N = \Phi_f/k$ and

$$T_a = \frac{0.5}{k} \iint_{\Omega} A_e(\theta, \phi) L_f(\theta, \phi) d\Omega \quad (20)$$

In a lossless antenna, effective area is related to antenna directivity by $A_e(\theta, \phi) = D(\theta, \phi) \lambda^2 / 4\pi$ (4), where $D(\theta, \phi)$ is the antenna directivity pattern. Equation (20) may therefore be written as

$$T_a = \frac{0.5}{k} \frac{\lambda^2}{4\pi} \iint_{\Omega} D(\theta, \phi) L_f(\theta, \phi) d\Omega \quad (21)$$

The quantity $(0.5 \lambda^2/k) L_f(\theta, \phi)$ is usually referred to as a radiometric brightness temperature, $T_B(\theta, \phi)$, allowing Eq. (21) to be written as

$$T_a = \frac{1}{4\pi} \iint_{\Omega} D(\theta, \phi) T_B(\theta, \phi) d\Omega \quad (22)$$

(In microwave radiometry the term *brightness* is often used as a synonym for spectral radiance.) The radiometric brightness temperature of a body is related to the body's physical temperature, T , by the body's emissivity; that is,

$$T_B(\theta, \phi) = \epsilon(\theta, \phi) T(\theta, \phi) \quad (23)$$

[The emissivity's dependence on ψ is not shown explicitly in Eq. (23) but is implied in its θ, ϕ dependence. The frequency dependence is suppressed only for clarity.] The first step in conventional radiometer data analysis usually requires that the radiometric brightness temperature of a particular region be estimated from a measurement of an antenna noise temperature, T_A , defined by

$$T_A = \frac{\Phi_R}{kB_N} \quad (24)$$

where Φ_R is the power supplied to the radiometer by the antenna. If the antenna has nonzero ohmic loss, the measured antenna noise temperature will comprise an attenuated aperture temperature (the aperture temperature being the temperature measured by a lossless antenna) and a thermal contribution from the physical antenna structure itself. For an antenna with ohmic (or radiation) efficiency, η_{Ω} (expressed as a per unit, not percent, quantity), and physical structure temperature, T_s , the measured antenna noise temperature will be

$$T_A = \eta_{\Omega} T_a + (1 - \eta_{\Omega}) T_s \quad (25)$$

If the antenna loss is sufficiently small, then $T_a \cong T_A$. Equation (22) shows, however, that the antenna aperture temperature is a weighted average of the observed brightness temperature in which the weighting function is the antenna directivity pattern. If the antenna has a radiation pattern with low side lobes and a beamwidth that is narrow with respect to the angular rate of change of brightness temperature, then T_a gives the required brightness temperature, in the direction of the main lobe, directly. For an imaging radiometer (which scans a beam over

some solid angle, for example, to form an image), with a beam that is not sufficiently narrow, deconvolution of the antenna pattern from the resulting image may be necessary.

Since antenna directivity and gain are related (4) by $G(\theta, \phi) = \eta_\Omega D(\theta, \phi)$, then Eqs. (22) and (25) can be combined and written as

$$T_A = \frac{1}{4\pi} \iint_{\Omega} G(\theta, \phi) T_B(\theta, \phi) d\Omega + (1 - \eta_\Omega) T_s \quad (26)$$

In the absence of significant atmospheric effects (losses and scattering), Eq. (26) shows that an adequate knowledge of antenna ohmic efficiency and antenna gain pattern will allow the brightness temperature of a surface (with a spatial resolution determined by the antenna beamwidth) to be found from a measurement of antenna noise temperature. [The noise figure of the radiometer must be considered, of course, when making this measurement (5).]

Atmospheric Loss and Scattering

If atmospheric effects cannot be neglected, then the brightness temperature inferred from Eq. (26) must be interpreted as an apparent brightness temperature since it will comprise a contribution from the target surface plus a contribution from the thermal emission of the atmosphere itself. Furthermore, the thermal radiation from each source (i.e., each surface element of the target and each volume element of the atmosphere) will be attenuated by the loss associated with the intervening atmosphere. Since atmospheric thermal emission is essentially isotropic, half of the thermal energy from each element is radiated upward (called upwelling radiation) and half is radiated downward (called downwelling radiation). For a downward-pointing radiometer (such as might be used in planetary-surface remote sensing applications), downwelling radiation may be reflected by the ground back into the antenna's main lobe. The contribution of this ground-scattered radiation will depend on the ground-reflectance and the integrated specific attenuation along the ground-reflected path to the radiometer. An upward-pointing radiometer (such as might be used in terrestrial atmospheric remote sensing applications) observes downwelling radiation from the atmosphere directly. The contribution from ground-scattered downwelling radiation, and ground thermal emission, will now be coupled into the antenna through its downward-pointing side or back lobes. The significance of this radiation will depend on the gain of these lower hemisphere lobes and the brightness temperature of the region within the antenna's main lobe. (If a microwave radiometer with high antenna gain is pointed at high elevation angle into a clear sky, well away from any discrete sources of radiation, then the brightness temperature observed by its main lobe may be as low as a few degrees kelvin. Radiation emitted by, or reflected from, the ground may, under these conditions, be a very significant source of measurement error.)

The preceding description assumes that the atmosphere is an absorbing, but nonscattering, medium. For an atmosphere filled with scattering particles, or more general scattering inhomogeneities, this simplified description is inadequate. [The operating wavelength, and the physical size and composition of inhomogeneities, will determine whether they act as significant scatterers or not (6).] In this case more rigorous theories

must be used to account for, and interpret, observed radiometric measurements (7).

APPLICATIONS

The most important application of radiometry lies in remote sensing of the earth's surface and its atmosphere (8). Polar ice mapping, for example, is possible using passive microwave radiometers located on earth-orbiting platforms. Microwave radiometry is especially useful for ice mapping because the contrast in emissivity between ice and water is particularly large at microwave frequencies. Furthermore, since passive microwave radiometry depends on self-emitted thermal radiation, and since the earth's atmosphere is essentially transparent at the microwave frequencies used, then data collection does not rely on solar illumination or the absence of cloud cover. (A requirement for solar illumination is particularly problematic in the polar regions, of course, due to the absence of the sun in winter.) Knowledge of annual, and longer term, changes in the extent of the polar ice shelves is important in the field of climatology generally and to recent concerns about possible climate change in particular.

Other surface materials (e.g., water/rock/sand) can be discriminated using microwave radiometers, and soil moisture mapping is also possible (9). Radiometer measurements applied in this area are of obvious interest in climatology, agriculture, and environmental science. Processes of desertification, like changes in the polar ice cover, are of concern in the context of global warming.

A rather different application of microwave radiometry concerns the development of the worldwide satellite communications network. Since a material's thermal emission and absorption are directly related, a brightness temperature measured at a given frequency can be used to infer the attenuation that an electromagnetic wave of that frequency would experience when propagating along the path observed. Ground-based, upward-pointing microwave radiometers can therefore be used to estimate the expected statistics of earth-space signal attenuation for use in the design of microwave satellite communication systems (10). The dominant mechanism resulting in large attenuation is the presence of precipitation along the path. Measurements of this type can, therefore, also be used in conjunction with appropriate signal attenuation/rain rate models to estimate rainfall intensity statistics for hydrological applications. (In recent years, however, the use of meteorological radar has probably become the preferred rainfall remote sensing tool.)

Active microwave radiometry can, like its passive variant, operate independent of solar illumination and is immune to cloud cover (and, depending on frequency, other meteorological phenomena that may make the atmosphere opaque). Downward-pointing scatterometers, for example, can be used to map surface wind velocities by measuring the effective reflectance of the ocean surface. The technique used in this application depends on the fact that the received signal for a nadir pointing, monostatic scatterometer decreases as the wind speed increases and the water surface becomes rough. This is due to the tendency of the reflected energy to be scattered in many directions by the rough surface, rather than reflected, specularly, back along the illuminating path. The apparent reflectance of the ocean surface can therefore be used as a measure of wind

speed. Other reflectance measurements, using forward- and side-looking radiometer beams, allow wind velocity to be inferred since the reflecting facets of the waves have a preferred orientation that depends on wind direction.

Downward-pointing, millimeterwave, spectroradiometer observations of brightness temperature around the absorption/emission lines of a homogeneously mixed atmospheric constituent can be used to estimate the vertical temperature profile of the atmosphere. Oxygen, which has a cluster of closely spaced resonance lines around 60 GHz, is often used for this purpose (11). The basic principle of such temperature measurements is as follows. The brightness temperature observed at any given frequency represents a weighted mean taken over the observed path. The weighting factor at a particular altitude depends on the density of the oxygen at that altitude and on the attenuation of the intervening atmosphere. Thus at high altitudes (near the radiometer) the contribution to total brightness will be small because the density of the gas is low. At low altitudes (far from the radiometer) the contribution to total brightness will also be low because the intervening atmosphere will greatly attenuate this contribution. The observed brightness temperature will therefore be dominated by a region of the atmosphere at some intermediate height. If measurements are made across a band of frequencies coincident with the skirt of an atmospheric constituent's absorption band where atmospheric attenuation changes rapidly with frequency, then the height from which the dominant contribution arises also changes with frequency. In this way a brightness temperature versus frequency profile can be transformed to a brightness temperature versus altitude profile.

In a related area, microwave radiometry has been applied in the estimation, using zenith pointing radiometers, of integrated atmospheric water content (both liquid and gaseous phases). The large specific, and latent, heats of water mean that mapping of this quantity is important in understanding the global energy balance. If spectroradiometer measurements are made across a significant band of frequencies close to the water vapor absorption line at 22 GHz, then the water vapor's density profile can be inferred. The technique here is essentially the same as that described previously in the context of temperature profiling using the 60 GHz oxygen absorption line(s). The temperature profile is first estimated (using the oxygen line, for example), and the brightness temperature versus frequency profile in the skirt region of the water vapor's absorption line is transformed to a brightness temperature versus altitude profile. The water vapor's density at a given altitude is then inferred from the brightness at that altitude. Similar measurement can be used to find vertical profiles of many other atmospheric constituents with applications to environmental monitoring and pollution control.

Infrared radiometry (especially at 4 μm and 11 μm) can be used for ocean surface temperature mapping with obvious applications to meteorology and climatology. In practice, both these frequencies, and at least one other intermediate frequency, are usually required in order to correct for errors introduced by atmospheric water vapor and scattering by atmospheric aerosols. Multifrequency measurements can yield data with accuracies of 1 K.

Chlorophyll, the dominant pigmentation in most species of plants, has absorption lines close to 440 nm and 650 nm corresponding to the blue and red parts of the visible spectrum, thus giving vegetation its predominantly green color. [In fact,

there are two distinct varieties of chlorophyll, blue-green and yellow-green, which both have absorption lines located about 10 nm either side of these frequencies (11).] In the near infrared region, however, absorption by chlorophyll is very low and reflectance correspondingly high. Infrared and visible reflectance measurements can, together, yield information about the photosynthetic capacity of selected areas of the biosphere (12). The presence of pigmentation other than chlorophyll can also be detected by spectroradiometric measurements and used to discriminate between different plant species. Furthermore, water absorption lines at 1.45 μm and 1.95 μm allow the water content of plant material to be estimated from infrared measurements. High-resolution, multiband, visible and infrared radiance measurements can thus be used to assess the quantity, variety, health, and photosynthetic vigour of much of the earth's surface vegetation.

Visible multiband measurements of radiance can also be used to gather information about the phytoplankton of the oceans' surface layers. A particular problem associated with these measurements is that upwelling radiation originating in the intervening atmosphere must generally be accounted for. One solution relies on the fact that the oceans do not contribute to upwelling radiance in the red part of the spectrum. Measurements in the red region of the spectrum over the oceans can therefore be used to estimate atmospheric upwelling radiance, which in turn can be used to infer atmospheric contributions at other wavelengths.

BIBLIOGRAPHY

1. C. J. Bartleson and F. Grum (eds.), *Visual Measurements*, Orlando: London: Academic Press, 1984.
2. International Commission on Illumination, *The Basis of Physical Photometry*, Publication CIE No. 18.2 (TC-1.2), Paris: Commission Internationale de L'Eclairage, 1983.
3. F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing—Fundamentals and Radiometry*, Reading, MA: Addison-Wesley, 1981.
4. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, New York: Wiley, 1998.
5. N. Skou, *Microwave Radiometer Systems: Design and Analysis*, Norwood, MA: Artech House, 1989.
6. L. P. Bayvel and A. R. Jones, *Electromagnetic Scattering and Its Applications*, London: Applied Science Publishers, 1981.
7. G. Brussaard and P. A. Watson, *Atmospheric Modelling and Millimetre Wave Propagation*, London: Chapman and Hall, 1995.
8. C. Elachi, *Introduction to the Physics and Techniques of Remote Sensing*, New York: Wiley, 1987.
9. T. J. Jackson, Passive microwave remote sensing system for soil moisture: Some supporting research, *IEEE Trans. Geosci. Remote Sens.*, **27**: 225–235, 1989.
10. J. Thirlwell and R. G. Howell, OTS and radiometric slant-path measurements at Martlesham Heath (UK), *Annales des Telecommunications (France)*, **36**: 15–23, 1981.
11. E. Schanda, *Physical Fundamentals of Remote Sensing*, New York: Springer-Verlag, 1986.
12. P. J. Sellers, Canopy reflectance, photosynthesis and transpiration, *Int. J. Remote Sens.*, **6**: 1335–1372, 1985.

IAN A. GLOVER
University of Bath

INTERFEROMETERS AND SPECTROMETERS

INTERFEROMETERS

Interferometers are instruments that measure the phase difference of waves. They allow waves to propagate and interfere inside them and yield interference fringe patterns. From the fringe patterns that the interferometer gives one can determine the phase difference of waves and perhaps other parameters such as length, thickness, displacement, angular velocity, pressure, sound, temperature, magnetic field, and so on. This article concerns only optical interferometers. They are operated over the light (ultraviolet, visible, and infrared) wavelength range from several hundred nanometers to ten micrometers. Other interferometers such as microwave interferometers, radio interferometers, and x-ray interferometers also use electromagnetic waves. Their principles are similar to those of optical interferometers.

There are many kinds of interferometers, such as the Michelson–Mach, Zehnder, Fabry–Perot, and Sagnac interferometers. The widely used Michelson interferometer has two separate optical paths, as shown in Fig. 1(a). The light beam from the optical source S is split into two beams at the semireflecting semitransmitting mirror SM and propagate along paths 1 and 2, respectively. After being reflected by two mirrors M_1 and M_2 , the two beams recombine at the semireflecting mirror SM to yield interference patterns. By moving the mirror in one path, one can change the phase difference of lightwaves in two paths and therefore find fringes shifting. The Michelson

interferometer is often used to measure length, displacement, and many other parameters.

The Mach–Zehnder is another useful interferometer in numerous measurement fields. It can transmit more optical power emerging from the light source to the receiver than the Michelson interferometer. The Mach–Zehnder interferometer also has two separate light paths (more often called *arms*) as shown in Fig. 1(b). A light beam from the optical source S is split at the first semireflecting mirror into two beams. After propagating along the two arms, respectively, the two beams are recombined at the second semireflecting mirror to yield interference. The fringes will shift as the difference of the optical path lengths between arm 1 and arm 2 varies. Because the two arms are separate, various bulk modulators, e.g., electrooptic modulators, Bragg cells, and Faraday magneto-optic modulators, can be inserted in the optical paths. Modulators provide modulations in phase, frequency, and/or polarization states of light waves. They have played important roles in various detection systems.

The Fabry–Perot interferometer belongs to the multiple-beam category of interferometers and comprises two parallel partially transparent plates, as shown in Fig. 1(c). The light beam from the optical source passes through one plate at an angle and enters the cavity between the two plates. The beams are reflected between the two plates repeatedly, and when they emerge gradually from both sides of this cavity, called the Fabry–Perot cavity, they form, respectively, two types of multiple-beam interference fringe patterns. This kind of interferometer produces sharp fringes owing to the multiple reflection. The half-width of the bright fringe is related to the reflectivity of the inner surface of the plate. The higher the reflectivity, the narrower the half-width. The Fabry–Perot interferometer is useful for temperature measurement, optical filtering, and the study of the fine structure of spectrum lines.

Fiber-optic interferometers have formed a rather large and important category among the various interferometers. In fiber-optic interferometers, the lightwave propagates in the optical fiber rather than in the air. As the fiber is flexible, the direction of the light propagation can vary continuously, which is impossible for air-path interferometers. Environmental factors can influence the fiber by changing its dimensions and the refractive indices of its materials, and therefore result in the variations in the phase difference and the polarized state of lightwaves. These influences cause noises and errors in measurements, but on the other hand, they offer the fundamentals of various sensing schemes (1). The relation between the light phase shift and the external factor is quite complicated in practical situations. It depends on the nature of the measurand and the material and structure of the fiber used. The fiber-optic interferometer generally uses a single-mode optical fiber, permitting only one mode to propagate inside it. There is birefringence in the practical fiber that produces the two orthogonal polarization waves that propagate with different phase velocities. Perturbations couple the two polarization modes and thus give noise at the output. The use of the high-birefringence fiber can prevent coupling between the two modes.

Much effort has been spent to apply fiber-optic interferometers to pressure, strain, temperature, acoustic wave,

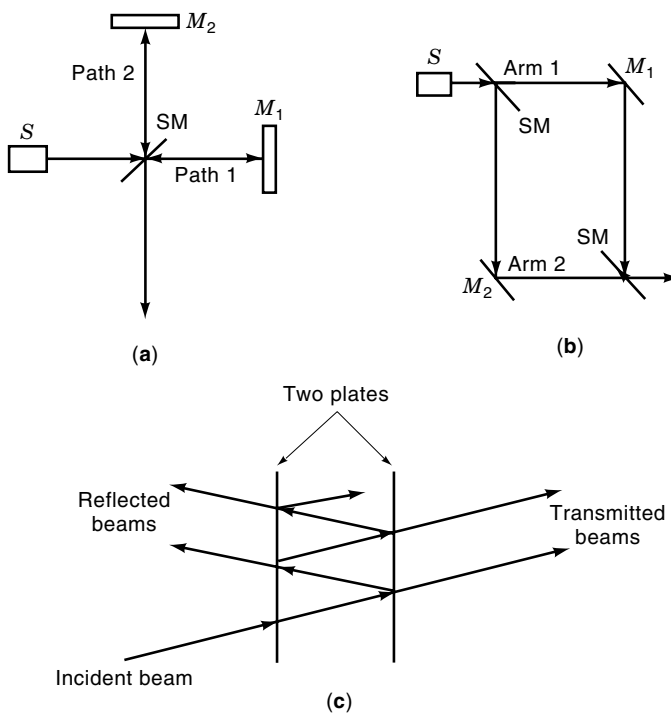


Figure 1. Three famous air-path interferometers. (a) The Michelson interferometer. (b) The Mach–Zehnder interferometer. (c) The Fabry–Perot interferometer. S , light source; SM , semireflecting, semitransmitting mirror; M_1 , M_2 , mirrors.

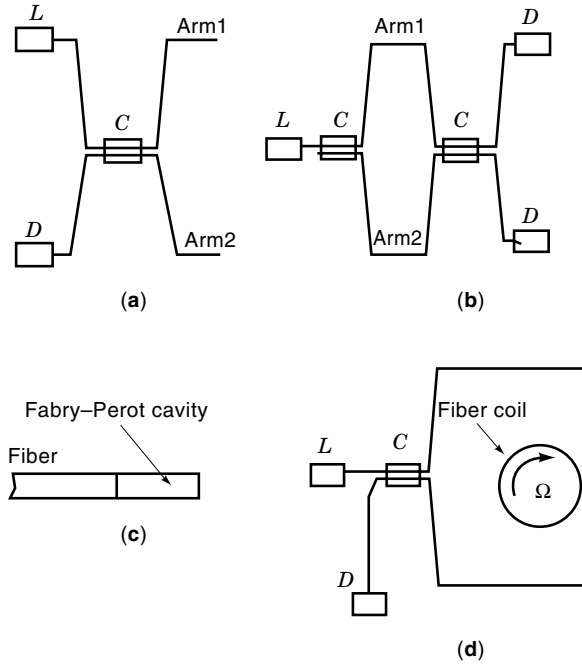


Figure 2. Four famous fiber-optic interferometers. (a) The fiber-optic Michelson interferometer. (b) The fiber-optic Mach-Zehnder interferometer. (c) The fiber-optic Fabry-Perot interferometer. (d) The fiber-optic Sagnac interferometer. *L*, laser; *C*, coupler; *D*, photodetector.

magnetic field, and rotation rate measurements (2). Figure 2 shows schematically several useful fiber-optic interferometers. In these arrangements, bulk devices such as semireflecting mirrors and modulators are replaced by their waveguide counterparts. The measurands induce the phase shift of the signal arm. Fiber-optic interferometers usually use the phase modulators to improve their performance. The phase modulator changes the optical path length of the reference arm and leads to the phase shift of the light wave. As the length of the fiber in the measurand field can be increased readily, the fiber-optic interferometer is apt to reach higher measurement sensitivity than the air path interferometer.

The principle of the fiber-optic interferometer is similar to that of its air path counterpart. Fig. 2(a) shows the fiber-optic Michelson interferometer. The optical beam from a laser launches to the coupler and divides into two beams there. After being modulated in each arm by measurand and modulator, respectively, the two beams recombine at the same coupler and give rise to interference. Compared with the fiber-optic Michelson interferometer, the fiber-optic Mach-Zehnder interferometer has two couplers, as shown in Fig. 2(b). The coupler at which the two light beams are recombined is different from the one at which the light beam is split. Obviously, as the phase difference varies, the output of the photodetector varies.

Figure 2(c) shows the basic configuration of the fiber-optic Fabry-Perot interferometer. There are two reflecting faces inside the optical fiber. In many cases, one reflecting face is just the surface end of the fiber. The fiber portion between the two reflecting faces constitutes a Fabry-Perot cavity. When the measurand changes the optical path length of the cavity, the phase difference of light beams proceeding to the photodetector is changed.

The fiber-optic Sagnac interferometer has a common ring path for two optical beams propagating in opposite directions. Nonreciprocal factors such as rotation rate generate a nonzero phase difference between the two light waves. As shown in Fig. 2(d), light from the optical source is split at the coupler into two beams, which then counterpropagate in the fiber coil. When the fiber coil rotates clockwise about its axis, the clockwise light beam takes a longer time to travel between two fixed points in the fiber coil than the anticlockwise light beam. So there exists a nonzero phase difference between the two counterpropagating lightwaves. This interferometer can be used for inertial navigation as a gyroscope.

The properties of light sources and photodetectors play important roles to interferometer performances (1-3). In most cases, interferometers use lasers as their optical sources because the laser light has a good spatial coherence and a good temporal coherence. Both light sources and photodetectors can introduce noises. There are many noises in interferometers, such as shot noise, thermal noise, $1/f$ noise, mode-mode interfering noise, and so on, which produce errors in output. Perturbations due to the fluctuation in the optical source intensity and the variation in the polarized state of lightwaves also result in errors. A well-designed setup and an appropriate detection scheme can reduce the influences of noises. There are numerous detection schemes, and many of them suit more than one interferometer.

BASIC FORMULAE FOR INTERFERENCE

The interference of two light waves gives a superposed optical field as follows

$$\mathbf{E}(t) = \mathbf{E}_1(t) + \mathbf{E}_2(t)$$

where $\mathbf{E}_1(t)$ and $\mathbf{E}_2(t)$ represent the optical vectors of the two light waves to be superposed. The fringe intensity as well as the output of the photodetector depends on the optical power of the light wave rather than the instantaneous value of the amplitude of the light wave. Thus the fringe intensity can be written as (4)

$$I = I_0 + I_0 \mathcal{V} \cos \phi$$

Here ϕ denotes the phase difference of the two light waves, I_0 is a constant proportional to the light source intensity, and $\mathcal{V}$ gives the interference term that is related to the mixing efficiency. The fringe intensity depends on the phase difference ϕ . When $\phi = m\pi$, $|m| = 0, 1, 2, \dots$, I reaches its maximum I_a , and when $\phi = m\pi$, $|m| = \frac{1}{2}, 3/2, 5/2, \dots$, I equals minimum I_i . The distinctness of the fringe pattern can be specified by the visibility of fringe or the fringe contrast that is defined as $(I_a - I_i)/(I_a + I_i)$.

The visibility of fringe is related to the spatial coherence, the temporal coherence, and the polarized direction of the light waves used. *Spatial coherence* describes how close a practical light source approaches a point optical source. When a practical light source is an extended one, the different points in the source will produce different light waves with different phases at the observing point. Thus the interference effect due to one emitting point of the light source may be counteracted partially by another point of the light source, which results in the

decrease of the visibility of fringe. On the other hand, when the light wave from the optical source is not a strictly monochromatic one, the visibility of fringe will decrease even if the light source is a point source in the case that the phase difference of light waves is large. *Temporal coherence* describes how near a practical light source approaches a monochromatic optical source. According to quantum mechanics, the light wave emerging from an atom is a wave train with a finite length in the time axis. This temporal finite length is called *coherence time*. The interference will occur when light wave trains are overlapped partially, that is, when the time difference of the light wave trains is less than the coherence time. There is a relation between the coherence time $\Delta\tau$ and the spectral width $\Delta\omega$ of the light which can be written as

$$\Delta\tau\Delta\omega \geq 1$$

which means that the wave train is not infinitely long unless the light is a strictly monochromatic one. Another important parameter specifying the temporal coherence is the *coherence length* l_c , which is defined as

$$l_c = c\Delta\tau$$

where c is the velocity of light. Generally speaking, the visibility of the fringe will be reduced as the difference between the optical paths increases, and will approach zero as such difference becomes much greater than the coherence length. However, practical experiments show that even if the path difference is several times longer than the coherence length, the interference effects can still be seen (2).

The polarized directions of the light waves can also affect the visibility. For the linearly polarized situation, when the polarized planes of the two superposed light waves are parallel, $\mathcal{V}$ has a maximum value; whereas when they are orthogonal, $\mathcal{V}$ becomes zero.

CONFIGURATIONS, PRINCIPLES, AND APPLICATIONS OF INTERFEROMETERS

Michelson Interferometer

The Michelson interferometer is shown schematically in Fig. 3. Light waves propagating along paths one and two are recombined at the semireflecting mirror or the coupler to give rise to interference fringe. Under illumination by an extended source, the interference produces fringes of equal inclination localized at infinity if the mirror M_2 and the image of the other mirror M_1 are parallel, and fringes of equal thickness localized near M_2 if they are mutually inclined at a small angle. When the interferometer is illuminated by a white-light source, one can see only zeroth-order fringe and some low-order fringes. High-order fringes are invisible because of the short coherence length of the white light. The fringe pattern can be converted into current or voltage by use of the photodetector. The output of the photodetector that responds to the optical intensity of the fringe will vary as the mirror moves. Obviously the output of the photodetector has a one-cycle variation as the mirror moves a distance equal to half an optical wavelength $\lambda/2$. When the difference of the optical path length between paths 1 and 2 equals $m\lambda$ where $|m| = 0, 1, 2, \dots$, the output of the photodetector reaches its maximum, and when such difference

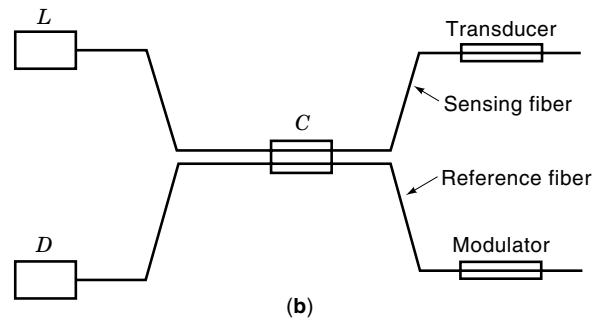
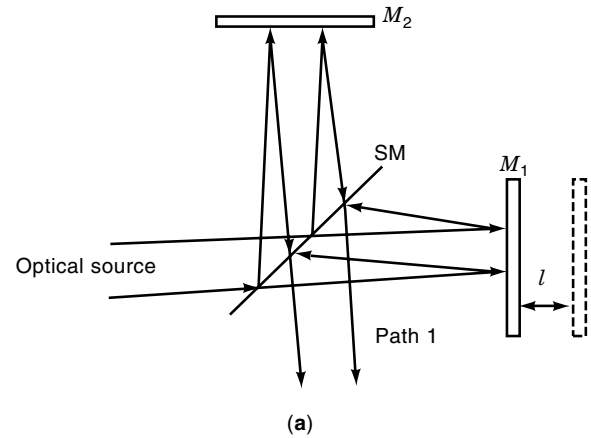


Figure 3. (a) The air-path Michelson interferometer that measures displacement of the plate M_1 . (b) The fiber-optic Michelson interferometer that is used as the interferometric sensor. L , laser; SM , semireflecting, semitransmitting mirror; M_1 , M_2 , mirrors; C , coupler; D , photodetector.

equals $m\lambda$ where $|m| = 1/2, 3/2, \dots$, the output of the photodetector reaches its minimum (4). Measuring these electric signals, one can obtain at least $\lambda/2$ or 300 nm of displacement resolution.

For the fiber-optic Michelson interferometric sensor, the variation measurand alters the length, diameter, and refractive index of the sensing fiber and thus results in the phase shift of the light wave. Take the temperature measurement as an example. In this case, the sensing fiber is attached to the thermal-sensitive material, whereas the reference fiber is attached to the thermal nonactive material. The variation in temperature makes the sensing material as well as the sensing fiber spread and contract so that the phase difference is changed. This kind of thermometer can attain a sensitivity of the order of 1×10^{-5} K (5).

The Michelson interferometer can be used to measure displacement, length, temperature, pressure, gas flow, spectrum distribution, and many other factors.

Mach-Zehnder Interferometer

Figure 4 shows the schematic of a typical Mach-Zehnder interferometer. The output of the photodetector reaches its maximum when the difference of the optical path lengths nl between

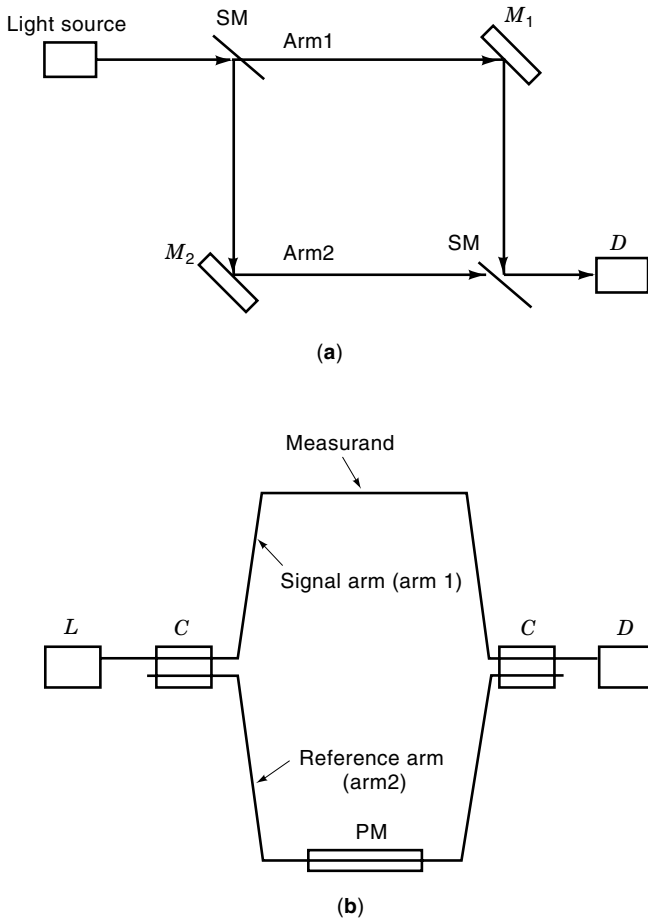


Figure 4. The Mach-Zehnder interferometer, which can transmit more optical power to the photodetector than the Michelson interferometer. (a) The air-path. (b) The fiber-optic. *L*, laser; *SM*, semireflecting, semitransmitting mirror; *M*₁, *M*₂, mirrors; *C*, coupler; *D*, photodetector; *PM*, phase modulator.

two arms 1 and 2 satisfies

$$nl = m\lambda, |m| = 0, 1, 2, \dots$$

and minimum when nl satisfies

$$nl = m\lambda, |m| = 1/2, 3/2, 5/2, \dots$$

where n is the refractive index and l is the path difference of the two arms (1).

For the fiber-optic Mach-Zehnder or other type of interferometers, the detectable phase shift usually has to be as small as 10^{-6} rad or less. Thus the noises caused by lasers, photodetectors, and other components cannot be ignored (2,3). For a stable and reliable readout, an appropriate detection scheme is necessary.

The detection schemes can be classified as homodyne with phase tracking, homodyne without phase tracking, and heterodyne. The homodyne (3) detection system is illustrated in Fig. 5. The output voltages (or currents) of the photodetectors are of

the form for this case

$$\begin{aligned} V_1 &= V_0[1 + \mathcal{V}\cos(\phi_s - \phi_r + \phi_n)] \\ V_2 &= V_0[1 - \mathcal{V}\cos(\phi_s - \phi_r + \phi_n)] \end{aligned} \quad (1)$$

where ϕ_s is the phase shift in the signal arm induced by the measurand, ϕ_r is the phase shift in the reference arm produced by modulation, and ϕ_n is the phase difference between the two arms that is due to noise and other disturbance. In this case, V_0 is the constant proportional to the optical power of the laser. $\mathcal{V}$ is the parameter related to fringe contrast. To obtain the maximum sensitivity, the system should satisfy the quadrature condition, that is, $\phi_s - \phi_r + \phi_n = 2m\pi \pm \pi/2$. The phase-tracking homodyne scheme uses the feedback control technique to maintain the quadrature condition. The electronic circuit transforms V_1 and V_2 into the feedback voltage V_m that applies to the piezoelectric transducer (PZT) device to stretch the reference arm to change ϕ_r . Under the quadrature condition, V_1 and V_2 are proportional to the phase shift induced by the measurand, approximately if this phase shift is small.

Besides the above dc modulation manner, the interferometer can be operated in the dc-ac hybrid modulation way. In this case, V_m is the sum of dc and ac voltages. The dc voltage produces a dc phase bias to maintain quadrature condition, and the ac voltage induces a phase variation with the modulation frequency to facilitate phase detection. The output V_1 can be achieved by $J_1(\max)$ method (6). One advantage of this ac modulation is the attainable high SNR because of the possible use of the lock-in amplifier in the system. Both the dc modulation scheme and the dc-ac hybrid scheme have drawbacks. They have resetting noise and are complicated in structure. However, in practice, the phase-tracking schemes can work quite well.

The no-phase-tracking homodyne system does not need feedback control. In contrast to the phase-tracking homodyne arrangement, the no-phase-tracking homodyne arrangement has a rather simple structure. It avoids the reset problem involved in the phase tracking operation. There are many no-phase-tracking homodyne schemes. Figure 6(a) shows schematically the system using a 3×3 coupler. The beams along the two arms recombine together at the 3×3 coupler instead of at a 2×2 coupler and give rise to outputs different from those given by Eq. (1). Through signal processing with appropriate electronic circuits, a result proportional to the phase difference is achieved. It has been reported that a minimum detectable phase shift of 3×10^{-6} rad at 1 kHz and a dynamic range of 10^7 were attainable for this detection scheme (7).

Figure 6(b) shows another no-phase-tracking homodyne scheme, called homodyne demodulation. The output of the photodetector V_1 in this case has the same form as that in Eq. (1). Because of the ac modulation, V_1 can be written as

$$\begin{aligned} V_1 &= V_0 + V_0 \mathcal{V} \{ [J_0(M) - 2J_2(M) \cos 2\omega_m t + \dots] \cos(\phi_s + \phi_n) \\ &\quad + [2J_1(M) \cos \omega_m t - \dots] \sin(\phi_s + \phi_n) \} \end{aligned} \quad (2)$$

where J_i denotes the Bessel function of order i , and M is modulation depth. Multiplying the first harmonic (ω_m) by the derivative of the second harmonic ($2\omega_m$), then subtracting the product of the second harmonic and the derivative of the first harmonic, and finally integrating this difference, one can obtain a result varying directly with $(\phi_s + \phi_n)$. As the signal ϕ_s and the noise ϕ_n are usually within different bands, it is not difficult to

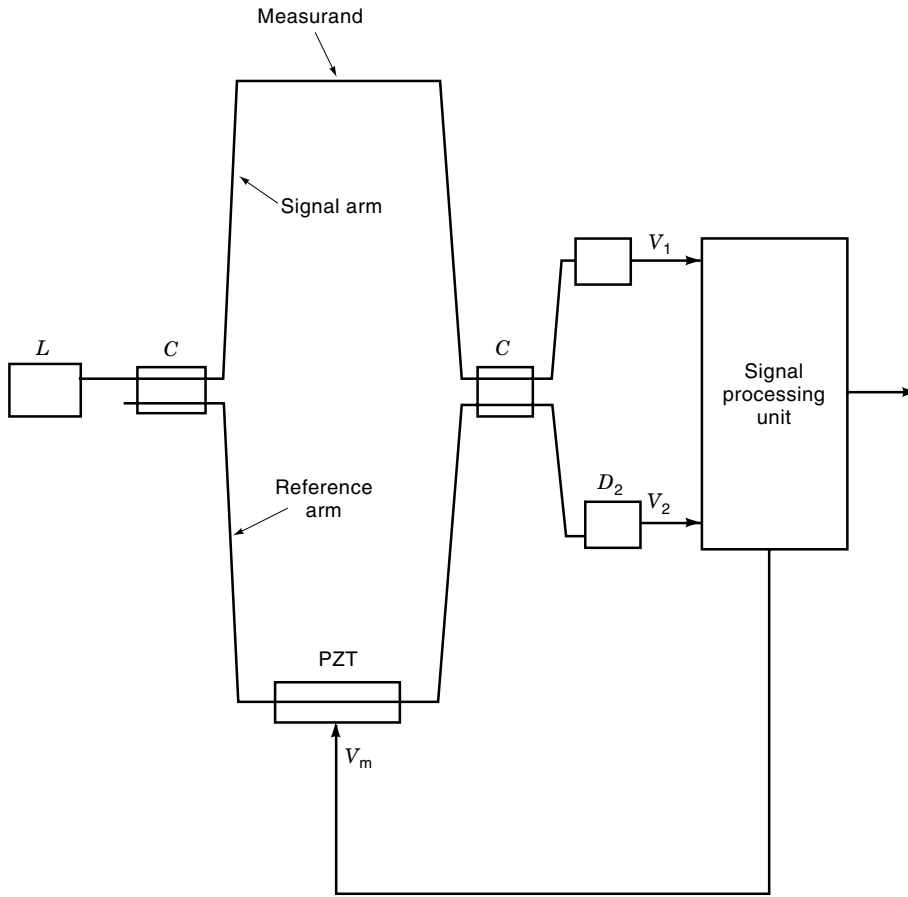


Figure 5. The basic form of the phase-tracking homodyne system. *L*, laser; *C*, couplers; *D*₁, *D*₂, photodetectors.

remove ϕ_n by filtering. Dandridge et al. have shown that the detectable phase shift of 10^{-6} rad at 1 kHz was attainable with this scheme (8).

The $J_1 \dots J_4$ method is also a useful no-phase-tracking homodyne scheme, as shown in Fig. 6(b). The ac phase modulation is not necessary here where the interferometer is used to measure dynamic phase changes. The output of the photodetector *D* has a form similar to that of Eq. (2). According to the recurrence relation of the Bessel functions, the following expression can be obtained

$$X^2 = \frac{4i(i+1)V(i)V(i+1)}{[V(i) + V(i+2)][V(i-1) + V(i+1)]}$$

Here X is the modulation depth induced by the measurand and $V(i)$ is the voltage amplitude of the i th frequency component of the output voltage of the photodetector. This approach has removed the need of stabilizing the source intensity and polarization control, which is welcome in practical applications (9).

The heterodyne detection system uses frequency modulators to produce a frequency shift. The output of the photodetector is an FM signal in this case. It has the form of

$$V_o = 2V_0 \mathcal{V} \cos[(\omega_s - \omega_r)t + \phi_s(t) + \phi_n]$$

where ω_s and ω_r are two angular frequencies of light waves in the signal arm and the reference arm, respectively.

Another heterodyne system is called synthetic heterodyne system. This system uses the PZT element as the frequency modulator. The output signal of the interferometer is calculated or "synthesized" before entering the FM discriminator. The processed signal has the form of

$$V_s = 2V_0 \mathcal{V} J_1(M) \cos(3\omega_m t + \phi_s + \phi_n)$$

The advantage of the heterodyne scheme is that neither feedback circuit nor resetting arrangement is required. The drawback of this method is that the oscillator phase noise and the laser amplitude noise can affect the detection result (10).

The frequency of the laser diode can be modulated by its driving current or the temperature around it. This characteristic provides another heterodyne detection possibility. Interferometers using FM laser diodes have been demonstrated for measuring the impurity inside the fiber and the length of the fiber as distributed sensors (11). They are also useful for multiplexed systems.

Other schemes such as that using two or more light beams with different wavelengths, that using an amplitude modulator rather than a phase modulator, and that using passive compensation with quarter-wave plate are also found in some interferometers.

Fabry-Perot Interferometer

The Fabry-Perot interferometer consists of two parallel plates whose inner surfaces are coated with partially transparent

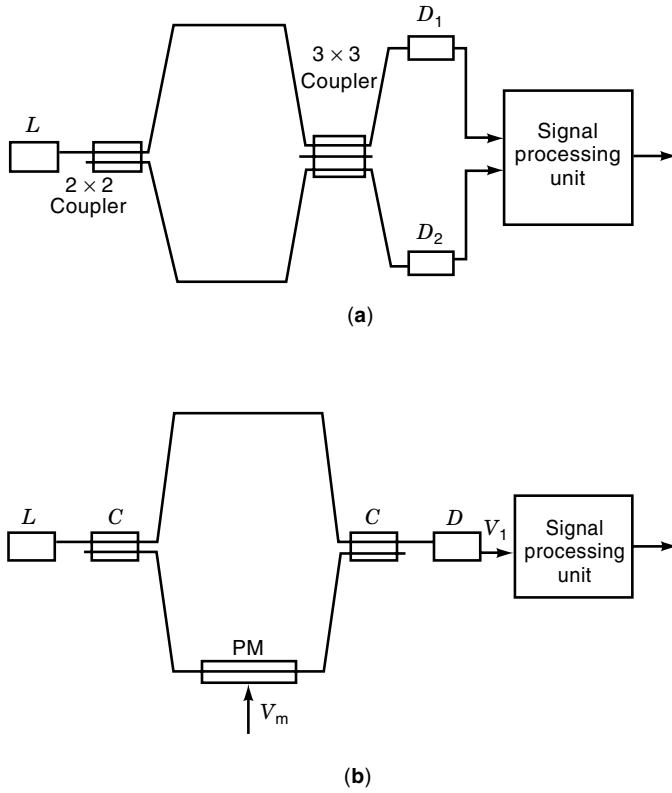


Figure 6. (a) The no-phase-tracking homodyne system using 3×3 coupler. (b) The passive homodyne system using demodulation or $J_1 \dots J_4$ method. L , laser; C , 2×2 couplers; D , D_1 , D_2 , photodetectors; PM, phase modulator.

high-reflectivity films, as shown in Fig. 7(a). Light beams entering the Fabry–Perot cavity will reflect many times between the two plates before going out of the cavity. To generate multiple reflection in the cavity, the plate surface must have high reflectivity, often >0.9 . This can be implemented by coating the surface with multilayer dielectric films.

Because of the multiple reflection in the Fabry–Perot cavity, the Fabry–Perot interferometer gives rise to sharp bright fringes for transmitted beams and sharp dark fringes for reflected beams. The relations between the output light intensities and the reflectivities are specified by Airy's formulae, which are written as follows:

$$I^{(t)}/I^{(i)} = T^2 / [(1 - R)^2 + 4R \sin^2(\phi/2)]$$

$$I^{(r)}/I^{(i)} = 4R \sin^2(\phi/2) / [(1 - R)^2 + 4R \sin^2(\phi/2)]$$

where $I^{(i)}$ denotes the incident optical intensity, $I^{(t)}$ the transmitted optical intensity, $I^{(r)}$ the reflected optical intensity, R and T are the reflectivity and the transmissivity of the plate surface, respectively, and ϕ is the phase difference of multiple light waves. The finesse is a parameter for specifying the sharpness of the fringes. It is defined as the ratio of the separation distance of adjacent fringes and the half-width of the fringe. The greater the finesse, the smaller the half-width and the narrower the fringes.

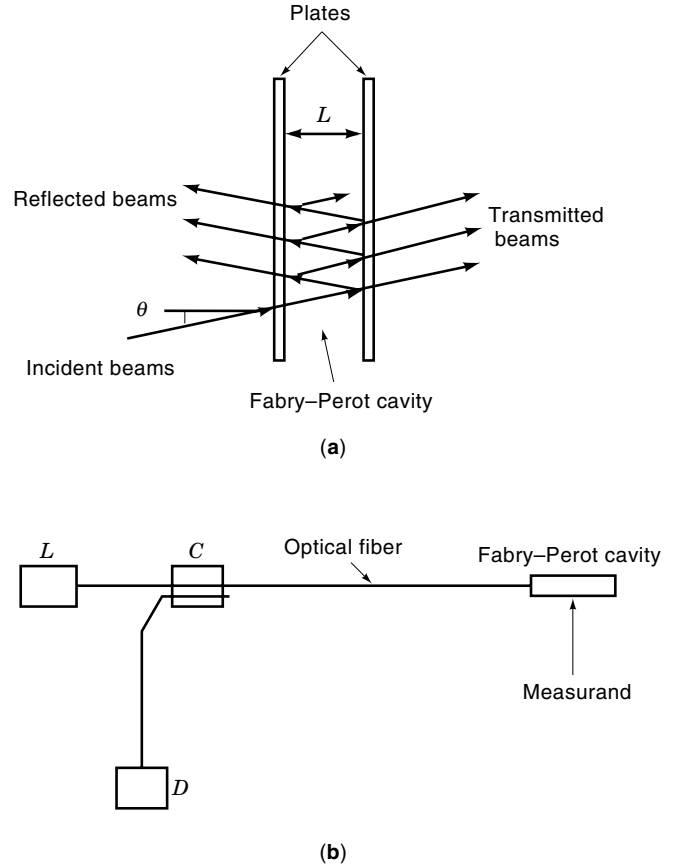


Figure 7. (a) The Fabry–Perot interferometer. (b) The basic configuration of fiber-optic Fabry–Perot interferometric sensor. L , laser; C , coupler; D , photodetector.

For the transmitted rays, the difference of multiple light waves is of the form (4)

$$\phi = (4\pi/\lambda)nL \cos \theta + 2\phi_d$$

where L is the separation distance between the two plates, θ is the incident angle, and ϕ_d is the additional phase change on internal reflection. Conventionally, the Fabry–Perot interferometer with fixed L is referred to as the *Fabry–Perot etalon*.

One application of the Fabry–Perot interferometer is the study of the fine structure of spectral lines. The Fabry–Perot interferometer is suitable for this work because it can give very narrow fringes. In this situation, another parameter, called the free spectral range, is used to specify the upper limit of the wavelength difference between the two light waves. When the wavelength difference approaches this upper limit, the displacement between the two fringe patterns will approach the separation distance of two adjacent fringes of any fringe pattern. Introduction of some auxiliary device, for example, a prism, into the instrument can eliminate this problem.

Another application of the Fabry–Perot interferometer is optical filtering. The Fabry–Perot interferometer only transmits the specific wavelengths that fall into the narrow pass-band corresponding to the sharp bright fringes. Therefore the

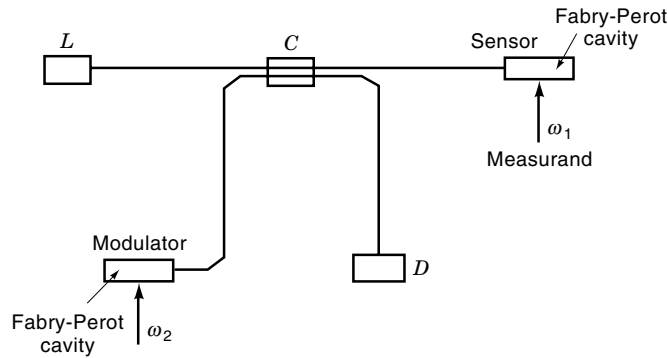


Figure 8. The arrangement of the extrinsic fiber-optic Fabry-Perot interferometer. There are two Fabry-Perot cavities in this system; one is a sensor and one is a modulator. *L*, laser; *C*, coupler; *D*, photodetector.

Fabry-Perot interferometer can serve as bandpass filter. Use of the PZT element can make this optical filter electrically tunable. In this case, applied voltage makes PZT elements stretch or contract so that the length of the Fabry-Perot cavity as well as its passband are varied.

A typical arrangement of the fiber-optic Fabry-Perot interferometric sensor is shown in Fig. 7(b). The measurand changes the optical path length of the cavity and therefore the phase difference of light beams proceeding to the photodetector. The fiber-optic Fabry-Perot interferometric sensors are suitable for measurements of temperature, pressure, and vibration. The materials of the Fabry-Perot cavity can be different from those of the optical fiber, which makes possible the application of the fiber-optic Fabry-Perot interferometric sensor to some specific measurements. Reference (12) gives an example of high temperature measurement. In this case, the sensor is a ceramic Fabry-Perot cavity connected by a multimode uncoated sapphire fiber. The report showed that it attained a resolution of 0.2°C over a range of 310° to 970°C.

Since there is no separated reference arm for the fiber-optic Fabry-Perot interferometer, a particular phase bias arrangement should be designed. Figure 8 shows an example of this. There are two Fabry-Perot cavities in this system, one is used as the remote sensor and the other as the modulator. The two Fabry-Perot cavities are modulated at two different frequencies, respectively. The output of the photodetector is the sum of the infinitely many harmonic terms. In this output the phase shift induced by the measurand can be determined by use of the Bessel recurrence relation. This detection system can eliminate the influences due to source intensity variation, fringe contrast fluctuation, and random phase drift and provide a linear readout of the phase shift (13).

FIBER-OPTIC SAGNAC INTERFEROMETER

The Sagnac interferometer is an instrument that detects the nonreciprocal phase difference of two counterpropagating lightwaves. The measurement of rotation rate as a gyro is perhaps the most important application of this interferometer. The basic configuration of the fiber-optic gyro is shown schematically in Fig. 9. The rotation of the fiber coil about its axis makes

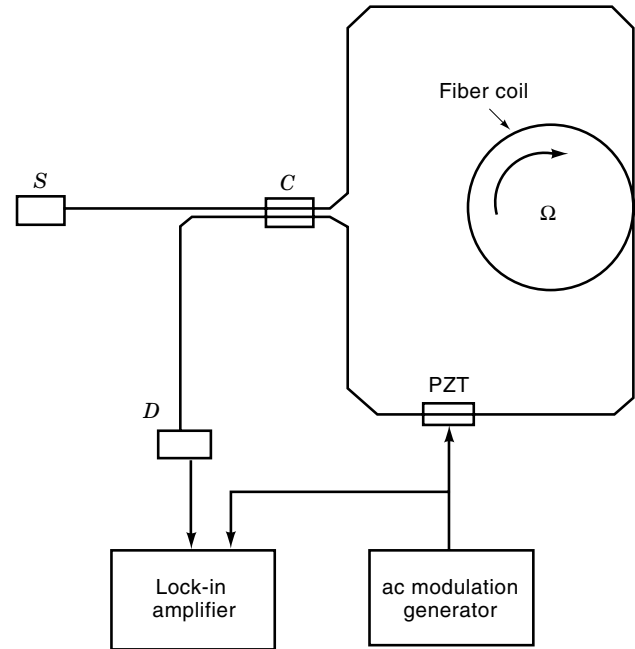


Figure 9. The basic structure of the fiber-optic Sagnac interferometer with phase bias. It is an open-loop system. The phase modulator is placed close to one end of the fiber coil to produce effective phase modulation. *S*, optical source; *C*, coupler; *D*, photodetector.

the two optical paths unequal for the two counterpropagating beams so that a nonzero phase difference is produced between the two light waves. It can be verified that this Sagnac phase difference is

$$\phi = 8\pi NA\Omega/\lambda c$$

where Ω is the rotation rate, and N and A are the number of turns and the area of the fiber coil, respectively (14). This phase difference is nonreciprocal. Its magnitude and sign depend on the rotation rate and direction. Nevertheless, since the interference intensity varies with $\cos \phi$, the sensitivity fades completely as the rotation rate approaches zero. Moreover, the rotation direction cannot be discriminated in terms of the output of the photodetector in this situation. To solve this problem, several methods have been devised to introduce a nonreciprocal 90° phase bias into the system.

The detection schemes for the fiber-optic gyro are classified as open-loop scheme and closed loop scheme. Figure 9 shows an open-loop system. It is an all-fiber phase-bias gyroscope. An ac phase modulator PZT is inserted into the fiber loop to introduce 90° phase bias. It seems that this modulator cannot produce the nonreciprocal phase bias because the changes in paths would be identical for the two light waves. However, in practice, as the two counterpropagating light waves pass through the PZT device at different moments, the path lengths for the two lightwaves are different. The relation between the output and the phase difference is similar to that shown in Eq. (2). In this case, fiber polarizers and polarization controllers may be required to maintain the identical polarization paths. The main feature of this arrangement is that all its components are constructed on a continuous optical fiber. This helps to remove

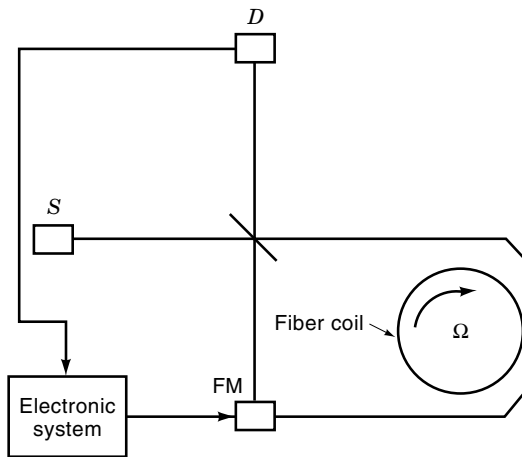


Figure 10. The typical arrangement of the phase-nulling fiber-optical gyro, where the frequency modulator rather than the phase modulator is used. S, optical source; D, photodetector; FM, frequency modulator.

the noise caused by reflections at the joints of components. The experiment shows that this interferometer has a phase error of 2×10^{-6} rad, corresponding to $0.2^\circ/\text{h}$ of long-term stability (15).

Figure 10 shows a phase-nulling fiber-optic gyroscope. A frequency modulator rather than the phase modulator is used in this closed-loop system. The two light waves counterpropagate along the fiber coil before and after passing through the frequency modulator, and produce a beat-frequency signal that reflects Sagnac phase shift. Through the feedback system, this beat-frequency signal makes the frequency change in such a way that the phase shift due to this frequency change just nulls out the Sagnac phase shift. There are two difficulties involved in this scheme: one is the need for a low frequency modulator and the other is about the dispersion of fiber. Some methods are available for eliminating these problems (3).

The heterodyne scheme can also be used to detect the Sagnac phase shift. In this system, a frequency modulator such as the Bragg cell is employed to modulate the frequency. The FM signal with an intermediate frequency includes the information of the Sagnac phase shift. The electronic circuit discriminates this FM signal. To prevent the influence of backscattering light waves, a coding technique by which the modulation frequency varies in a triangular fashion is used. The heterodyne system can eliminate the influence of the $1/f$ noise and avoid the AM effect (16).

A special problem for the Sagnac interferometer is the influence of coherent backscattering light waves that result in unwanted interference. One method to eliminate this influence is the use of incoherent sources such as the superluminescent diode, which emits short-coherence-length light.

Besides rotation rate, other factors such as current can produce nonreciprocal phase difference in the fiber-optic Sagnac interferometer. The current passing across the fiber coil generates a magnetic field. The Faraday effect induces a nonreciprocal phase shift between the two circularly polarized light waves counterpropagating through the fiber coil. It is verified that this nonreciprocal phase shift is proportional to the current. Many practical disturbances to the measurement can influence the measurement accuracy. P-A. Nicati et al. have analyzed disturbances to the Sagnac interferometer (17).

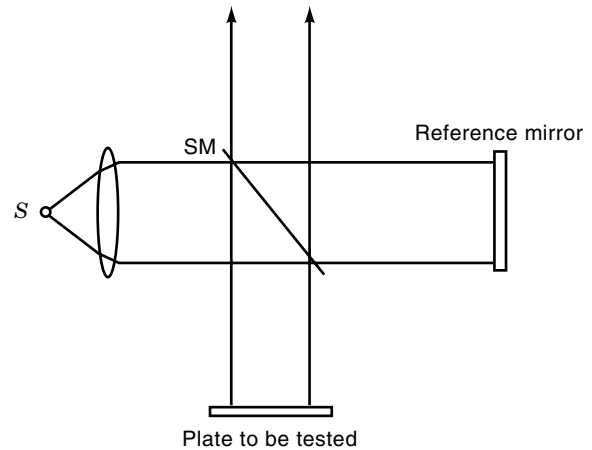


Figure 11. The basic form of the Twyman-Green interferometer, which is used to test the deformation of the plate. S, light source; SM, semireflecting, semitransmitting mirror.

The Twyman-Green interferometer is the variant of the Michelson interferometer. Figure 11 shows the basic form of the Twyman-Green interferometer. It uses broad beams, and it can be used to test the parallelism of the wave fronts of two light waves. The observer will see a uniform illumination if the two wave fronts are parallel and distinct fringes if there is some distortion between them. The fringes will shift as the mirror moves. The shifting direction is related to the distortion direction. The feature of this interferometer makes it possible to test the surface flatness of large objects.

Heterodyne detection scheme and image sensors such as the CCD linear array and area array are usually employed to receive the spatially distributed fringes produced by the Twyman-Green interferometer. The heterodyne detection system transforms the sign of the phase difference, which depends on the distortion direction of the tested surface, to the sign of the time difference of the zero crossing of the sensing signal and that of the reference signal. Thus the spatial distortion of the surface can be determined by measuring the temporal phase difference between the sensing signal and the reference signal (18).

The fiber-optic dual-mode interferometer is a one-fiber interferometer, that is, there are two modes propagating in only one fiber rather than one mode propagating in two separate fibers. This kind of interferometer is useful in temperature and pressure measurements. The variation of the temperature leads to the variation of the phase difference between the two modes and therefore makes a fringes shift. W. Eickhoff has demonstrated a scheme for measuring temperature which used the highly and linearly birefringent fiber with internal lateral stress and had a sensitivity of $1.66 \text{ rad} \cdot \text{K}^{-1} \cdot \text{m}^{-1}$. The advantage of such an interferometer is that variation of source intensity does not influence the readout (19).

In the resonant ring interferometer the output beam reenters the interferometer to produce resonance. Because there is multiple-beam interference in the ring, this interferometer, like the Fabry-Perot, can give rise to sharp fringes. It can measure various parameters with high sensitivities and resolutions. Figure 12 shows two basic forms of the resonant ring interferometer (20).

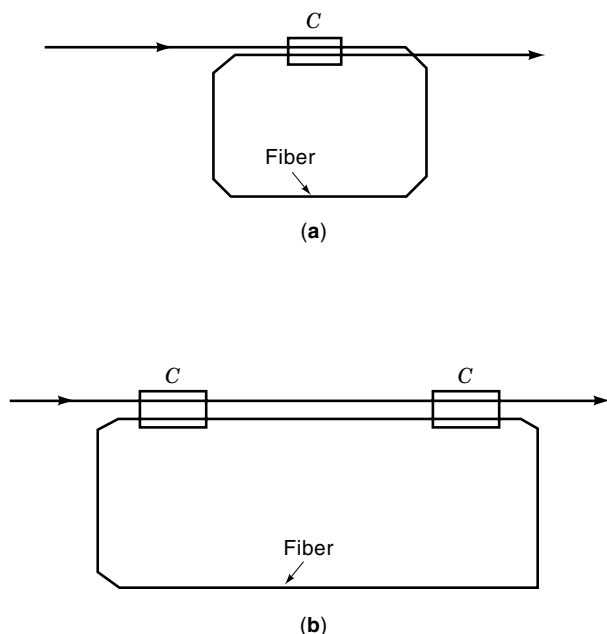


Figure 12. Two forms of resonant ring interferometers, in which a portion of output light is fed back into the interferometer. *C*, couplers.

Other interferometers such as Young's interferometer and the Jamin interferometer are also useful in practical applications (1).

RECENT DEVELOPMENTS

Many effective interferometric schemes have been advanced during the past decade. White-light interferometers utilize wideband light sources. With white light it is possible to discriminate the zeroth-order fringe readily and thereby make absolute measurements. T. Li et al. have given an example of a white-light Michelson interferometer for absolute distance measurement (21). An interferometer combination has also been used for absolute measurement (22). Recently, multiplexing of white-light fiber-optic interferometric transducers has been demonstrated for temperature and pressure measurement (23). A digital technique has been incorporated with the interferometer. A compact interferometer with PC interface has been reported for displacement measurement (24). A heterodyne double-pass differential interferometer with a PC-based control system is used to calibrate the subnanometric displacement actuator, whose uncertainty reaches 0.7 nm (25).

The measurement abilities of the interferometer have been greatly expanded. Besides displacement, length, vibration, temperature, pressure, strain, acoustic waves, current, magnetic field, and rotation rate, applications of interferometers to measurements on gravitation and gravitational waves (26, 27), energized underground electric cables (28), hydrocarbon pollutants in the bulk of a liquid (29), and dipole-orientation distributions of a polymer (30) have been reported. Recently, many specially designed Sagnac interferometers have been demonstrated for measurements (for example) of surface profiles (31), chromatic dispersion of fibers (32), and rotation rate (33). An all-optical Sagnac interferometer can largely eliminate the

influence of the source intensity noise, with a noise suppression ratio of 35 for an unmodulated system (34).

There is much interest in the interferometric spectrometer. Chern and Chao introduced an enhanced Michelson interferometer in connection with a modified forward-backward linear prediction algorithm. It eliminates the drawbacks of the conventional FFT method, which involves a large amount of calculation (35). For spectrum measurement, it has been reported that use of the correlated unbalanced interferometric method can attain a resolution 100 times better than that of the Fabry-Perot interferometer or on the order of 0.5 kHz in some special situations (36). Some research shows that the Fabry-Perot interferometer acts as a low-pass device in response to time-varying optical input (37). A longer cavity can provide higher spectral resolution, but may result in worse dynamic response. Similar conclusions can be drawn for the ring interferometer.

Another development related to spectral devices is the electrically tunable optical filter, which is useful in communication systems for wavelength-division multiplexing. In 1990 Patel reported a Fabry-Perot interferometric filter whose cavity is filled with nematic liquid-crystal material. The voltage applied to the device regulates the refractive index of the material in the cavity, rather than the length of the cavity, to change the passband. This kind of electrically tunable optical filter has a simple structure. It consumes less power than conventional devices. However, it is polarization-sensitive. An improvement has been made by using the same material but with twisted structure. In this case, when the applied voltage was sufficiently high (2 V), the device was polarization-insensitive. The experiment showed that the linewidth of the transmission peak of this device is about 0.5 nm. Its response time is of the order of millisecond. Further research results have shown that the substitution of a ferroelectric liquid crystal for the nematic liquid crystal can increase the response speed of this kind of optical filter (38).

Similarly to the interference in the time domain, interference between two light waves can occur in the frequency domain. One can see this phenomenon even if two light-wave trains are separated by a time interval longer than the wave-train duration. Figure 13 schematically shows a frequency-domain interferometer useful for the study of femtosecond phase spectroscopy. There are three pulses propagating in this interferometer: probe, excitation, and reference. They pass

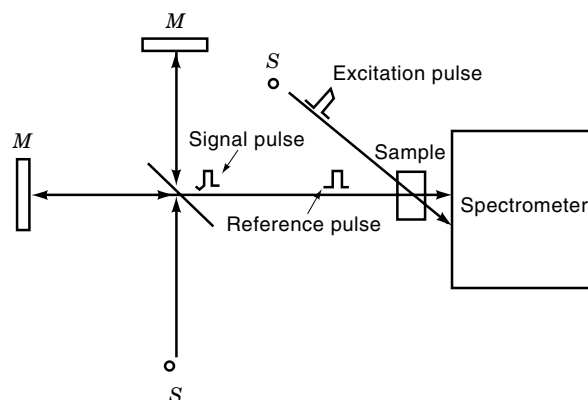


Figure 13. The frequency-domain interferometer for study of femtosecond phase spectroscopy. *S*, optical source; *M*, mirror.

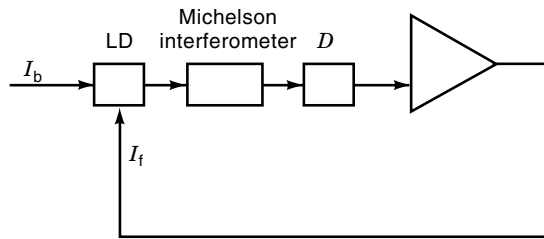


Figure 14. The hybrid bistable and multistable interferometer. LD, laser diode; D, photodetector; I_b , bias current; I_f , feedback current.

through the same sample at different moments. The reference pulse passes through the sample first, followed by the excitation pulse and finally the probe pulse. The reference pulse does not experience the change in the complex refractive index caused by the excitation pulse, but the probe pulse does. One can determine the femtosecond phase spectrum in terms of the interference between the probe and the reference pulses with and without excitation. The advantage of this method is that there is no need for accurate temporal overlapping of the femtosecond pulses (39).

Misawa and Kobayashi constructed a femtosecond interferometer. They designed a Sagnac interferometer for measurement of continuous phase spectra with white light. The zeroth-order fringe of white light can be found readily in the Sagnac interferometer because of its reciprocal structure. The sample is inserted into the optical loop. The reference pulse arrives at the sample earlier than both the excitation and the probe pulse. The probe and reference pulses propagate in opposite directions, and the probe pulse reaches the sample after the excitation pulse passes through it. According to the interference results between the probe pulse and the reference pulse, one can estimate the nature of the sample (40).

Besides the measurement of femtosecond phase spectra, the interferometer finds application in the generation and detection of ultrafast optical signals. A mode-locked ring laser was reported to have a pulse output with 0.4 ps width (41). A simple common-path-designed Sagnac interferometer was demonstrated for detection of ultrafast optical signals; it is operated at normal incidence with an objective for pumping as well as probing (42). An interferometer can also be used in the study of soliton collisions (43).

Nonlinear or digital use of the interferometer is an attractive research area. Figure 14 shows the block diagram of a hybrid bistable and multistable interferometer (44). With the help of electronic feedback, the interferometer can be

operated in a bistable or multistable state. In this case, the external light input minus the output of the interferometer controls the injection current of the laser diode. The bistable and multistable states are observed in the output from the laser diode. Besides this hybrid arrangement, all-optical devices such as the twin-core fiber nonlinear Mach-Zehnder interferometer (45) and the bistable fiber-optic unbalanced Michelson interferometer (46) are available for bistable operation. Jinno and Matsumoto reported a nonlinear Sagnac interferometer. Highly stable logic operations, including Inversion, AND, OR, and XOR, were achieved at multigigabit-per-second rates (47).

Some new bistable schemes have been discussed recently. Figure 15 shows a fiber-reflection Mach-Zehnder interferometer. It is able to switch the light from reflection to transmission or vice versa by changing the phase difference of light waves when couplers have splitting ratios in a range ($0.146 < \kappa < 0.854$). Beyond this range the interferometer transmits most of the light and reflects only a small portion of it—a feature that will make this interferometer useful for optical switching (48).

BIBLIOGRAPHY

1. B. Culshaw, *Optical Fiber Sensing and Signal Processing*, London: Peter Peregrinus Ltd., 1984.
2. T. G. Giallorenzi et al., Optical fiber sensor technology, *IEEE J. Quantum Electron.*, **QE-18**: 626–665, 1982.
3. S. L. Chuang, *Physics of Optoelectronic Devices*, New York: Wiley, 1995.
4. M. Bass, *Handbook of Optics*, New York: McGraw-Hill, 1995.
5. T. Kimura et al., Heat propagation in differential thermal analysis using fiber-optics technology, *IEEE Trans. Instrum. Meas.*, **IM-38**: 599–603, 1989.
6. D. A. Jackson, A. Dandridge, and S. K. Sheem, Measurement of small phase shifts using a single-mode optical-fiber interferometer, *Opt. Lett.*, **5**: 139–141, 1980.
7. K. P. Koo, A. B. Tveten, and A. Dandridge, Passive stabilization scheme for fiber interferometers using (3×3) fiber directional couplers, *Appl. Phys. Lett.*, **41**: 616–618, 1982.
8. A. Dandridge, A. B. Tveten, and T. G. Giallorenzi, Homodyne demodulation scheme for fiber optic sensors using phase generated carrier, *IEEE Trans. Microw. Theory Tech.*, **MTT-30**: 1635–1641, 1982.
9. V. S. Sudarshanam and K. Srinivasan, Linear readout of dynamic phase change in a fiber-optic homodyne interferometer, *Opt. Lett.*, **14**: 140–142, 1989.
10. J. H. Cole, B. A. Danver, and J. A. Bucaro, Synthetic-heterodyne interferometric demodulation, *IEEE J. Quantum Electron.*, **QE-18**: 694–696, 1982.
11. R. Juskaitis et al., Distributed interferometric fiber sensor system, *Opt. Lett.*, **17**: 1023–1025, 1992.
12. A. Wang et al., Sapphire-fiber-based intrinsic Fabry-Perot interferometer, *Opt. Lett.*, **17**: 1021–1023, 1992.
13. V. S. Sudarshanam and R. O. Claus, Split-cavity cross-coupled extrinsic fiber-optic interferometric sensor, *Opt. Lett.*, **18**: 543–545, 1993.
14. W. K. Burns, *Optical Fiber Rotation Sensing*, San Diego: Academic, 1994.
15. R. A. Bergh, H. C. Lefevre, and H. J. Shaw, All-single-mode fiber-optic gyroscope, *Opt. Lett.*, **6**: 198–200, 1981.
16. B. Culshaw and I. P. Giles, Frequency modulated heterodyne optical fiber Sagnac interferometer, *IEEE J. Quantum Electron.*, **QE-18**: 690–693, 1982.

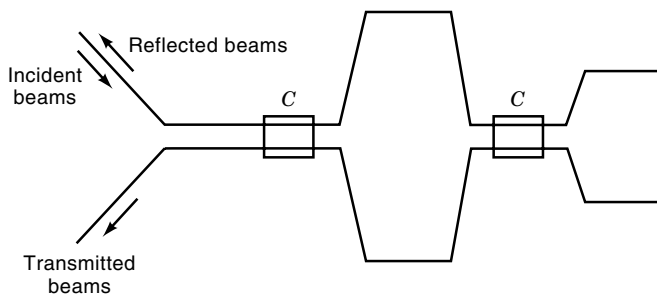


Figure 15. The configuration of the fiber-reflection Mach-Zehnder interferometer. C, couplers.

17. P-A. Nicati and P-A. Robert, Numerical analysis of second-order polarization effects in a Sagnac current sensor, *IEEE Trans. Instrum. Meas.*, **IM-39**: 219–224, 1990.
18. N. A. Massie, R. D. Nelson, and S. Holly, High performance real-time heterodyne interferometry, *Appl. Opt.*, **18**: 1797–1803, 1979.
19. W. Eickhoff, Temperature sensing by mode-mode interference in birefringent optical fibers, *Opt. Lett.*, **6**: 204–206, 1981.
20. K. Arai et al., Sensitivity-enhancing scheme of a polarimetric heterodyne sensor using a birefringent fiber loop, *Opt. Lett.*, **15**: 1103–1105, 1990.
21. T. Li et al., White-light scanning fiber Michelson interferometer for absolute position-distance measurement, *Opt. Lett.*, **20**: 785–787, 1995.
22. E. Gelmini, U. Minoni, and F. Docchio, Tunable, double-wavelength heterodyne detection interferometer for absolute-distance measurements, *Opt. Lett.*, **19**: 213–215, 1994.
23. W. J. Bock and W. Urbanczyk, Coherence multiplexing of fiber-optic pressure and temperature sensors based on highly birefringent fibers, *IEEE Trans. Instrum. Meas.*, **IM-49**: 392–397, 2000.
24. S. Donati, L. Falzoni, and S. Merlo, A PC-interfaced compact laser-diode feedback interferometer for displacement measurements, *IEEE Trans. Instrum. Meas.*, **IM-45**: 942–947, 1996.
25. A. Sacconi, G. B. Picotto, and W. Pasin, The IMGC calibration setup for microdisplacement actuators, *IEEE Trans. Instrum. Meas.*, **IM-48**: 483–486, 1999.
26. P. T. Beyersdorf, M. M. Fejer, and R. L. Byer, Polarization Sagnac interferometer with postmodulation for gravitational-wave detection, *Opt. Lett.*, **24**: 1112–1114, 1999.
27. J-M. Chartier et al., A portable iodine stabilized He–Ne laser and its use in an absolute gravimeter, *IEEE Trans. Instrum. Meas.*, **IM-42**: 420–422, 1993.
28. A. V. Mamishev et al., Nonintrusive sensing techniques for the discrimination of energized electric cables, *IEEE Trans. Instrum. Meas.*, **IM-45**: 457–461, 1996.
29. D. P. Hand et al., Optical fiber interferometry for photoacoustic spectroscopy in liquids, *Opt. Lett.*, **20**: 213–215, 1995.
30. W. Brinker et al., Phase-shift interference microscope for the investigation of dipole-orientation distributions, *Opt. Lett.*, **20**: 816–818, 1995.
31. T. Shirai, T. H. Barnes, and T. G. Haskel, Surface-profile measurement by means of a polarization Sagnac interferometer with parallel optical feedback, *Opt. Lett.*, **24**: 297–299, 1999.
32. K. S. Abedin, M. Hyodo, and N. Onodera, Measurement of the chromatic dispersion of an optical fiber by use of a Sagnac interferometer employing asymmetric modulation, *Opt. Lett.*, **25**: 299–301, 2000.
33. K. Taguchi et al., Experimental investigation of a semiconductor ring laser as an optical gyroscope, *IEEE Trans. Instrum. Meas.*, **IM-48**: 1314–1317, 1999.
34. P. Polynkin, Josiel de Arruda, and J. Blake, All-optical noise-subtraction scheme for a fiber-optic gyroscope, *Opt. Lett.*, **25**: 147–149, 2000.
35. S.-J. Chem and K.-J. Chao, An enhanced Fourier transform spectrometer with a search algorithm, *IEEE Trans. Instrum. Meas.*, **IM-45**: 127–135, 1996.
36. I.-W. Oh and H. R. D. Sunak, Measurement of the spectral linewidth of semiconductor laser for use with coherent optical communication system, *IEEE Trans. Instrum. Meas.*, **IM-36**: 1054–1059, 1987.
37. P. Tremblay and R. Ouellet, Frequency response of a Fabry–Perot interferometer used as a frequency discriminator, *IEEE Trans. Instrum. Meas.*, **IM-40**: 204–207, 1991.
38. J. S. Patel, Electrically tunable ferroelectric liquid-crystal Fabry–Perot filter, *Opt. Lett.*, **17**: 456–458, 1992.
39. E. Tokunaga and A. Terasaki, Frequency-domain interferometer for femtosecond time-resolved phase spectroscopy, *Opt. Lett.*, **17**: 1131–1133, 1992.
40. K. Misawa and T. Kobayashi, Femtosecond Sagnac interferometer for phase spectroscopy, *Opt. Lett.*, **20**: 1550–1552, 1995.
41. K. Kuroda and H. Takakura, Mode-locked ring laser with output pulse width of 0.4 ps, *IEEE Trans. Instrum. Meas.*, **IM-48**: 1018–1022, 1999.
42. D. H. Hurley and O. B. Wright, Detection of ultrafast phenomena by use of a modified Sagnac interferometer, *Opt. Lett.*, **24**: 1305–1307, 1999.
43. Y. Sakai, R. J. Hawkins, and S. R. Friberg, Soliton-collision interferometer for the quantum nondemolition measurement of photon number: Numerical results, *Opt. Lett.*, **15**: 239–241, 1990.
44. J. Ohtsubo and Yun Liu, Optical bistability and multistability in an active interferometer, *Opt. Lett.*, **15**: 731–733, 1990.
45. B. K. Nayar et al., All-optical switching in a 200 m twin-core fiber nonlinear Mach–Zehnder interferometer, *Opt. Lett.*, **16**: 408–410, 1991.
46. N. Furstenuau, Bistable fiber-optic Michelson interferometer that uses wavelength control, *Opt. Lett.*, **16**: 1896–1898, 1991.
47. M. Jinno and T. Matsumoto, Ultrafast all-optical logic operations in a nonlinear Sagnac interferometer with two control beams, *Opt. Lett.*, **16**: 220–222, 1991.
48. G. Dickinson, D. A. Chapman, and D. A. Gorham, Properties of the fiber reflection Mach–Zehnder interferometer with identical couplers, *Opt. Lett.*, **17**: 1192–1194, 1992.

KEN XU

Huazhong University of Science
and Technology

ACOUSTIC WAVE INTERFEROMETERS

Propagation of sound in a medium is a wave phenomenon and thus exhibits interference. In fact, interference is observed when two coherent waves are allowed to be superimposed in space. It is possible to obtain two coherent waves by splitting a single wave either at its wavefront, by passing through two or more apertures, or in its amplitude, by reflection and transmission at the interface of media of different phase velocity and/or density. The interference pattern at any point depends on the phase difference (relative time delay) caused by the propagation of these initially coherent waves through different acoustic paths.

When two sound waves interact with a particle of the propagation medium, the resultant displacement, velocity, and pressure are the vector sum of the effects due to the separate waves. If the final distribution of energy for this particle is different from what it would have been for each separate wave, interference has occurred. This phenomenon may be used to determine the relative physical properties of the media through which these waves have traveled prior to their superimposition. If two or more coherent waves propagate along a different acoustic path with respect to the reference material or to each other, their superimposition produces fringes (i.e., alternating regions of constructive and destructive interference, with a maximum or minimum of intensity, respectively). The shape of these regions depends on the geometry of the media causing the path difference. Thus a properly designed acoustic interferometer (or an interferometer, in general) behaves as a differential device since it transforms the phase difference of the initially coherent wave into intensity modulation.

A variety of experimental acoustic (and ultrasonic) techniques for material characterization have been developed in recent decades: for example, time-of-flight methods, photoelastic and acoustoelastic techniques, pulse-echo overlap methods, and phase detection methods (1). Among all these techniques, the interferometric ones, based on the aforementioned phenomenon of interference, present several advantages as far as accuracy and reproducibility of the measurements are concerned. In fact, they are not based on the measurement of absolute values, like other techniques, but rather on the determination of differences between reference signals and perturbed ones (e.g., phase shifts, frequency shifts) (2). Such differences are responsible for destructive or constructive interference between signals, which can be measured with high precision by using simple experimental setups. Other advantages of acoustic interferometry are as follows:

1. Its applicability to very small samples
2. Its ability to measure simultaneously both velocity and attenuation
3. The possibility of using noncontact detectors
4. The higher precision that may be obtained from the direct measurement of the difference between unperturbed and perturbed quantities

A proper logical control device is connected to the detector in order to trigger, both in amplitude and frequency, the signal generator of the acoustic wave until a specific interferometric pattern is achieved. This capability is exploited when the interferometric device is used to measure the triggering required for reproducing the same pattern under different external conditions. In this case the precision and sensitivity of the instrument can be highly enhanced.

Although all interferometers are based on the same simple general principle, several different realizations may be designed. First, acoustic waves may interfere directly or can be used to produce an electric signal or to modify the path of an optical wave, and then be used for interference. Second, different techniques may be used for generating the path difference. The simplest way is to let the wave experience multiple reflections at the interfaces of the propagation medium. Alternatively, the input and output signals may be correlated or two different input signals injected in different locations may be used.

Acoustic interferometers are used both for measuring the actual phase velocity and attenuation factor of the propagation medium and for evaluating their variation in the presence of a perturbation (e.g., pressure variation, presence of impurities) with respect to a reference case (ambient pressure, homogeneous specimen, etc.). In the first case, the physical properties are directly correlated to the position and intensity of the maxima and minima of the interference pattern. In the second case, higher accuracy may be obtained if some device characteristics (e.g., position, frequency) are properly modified in order to reproduce the same interference pattern as in the unperturbed case (in particular, the conditions for destructive interference).

To obtain constructive and/or destructive interference fringes, the acoustic path difference between the two interfering beams must be at least one-half of a wavelength. This puts constraints on the specimen thickness (or allowed variation of

it) and the frequency of the acoustic wave employed. For low frequencies, one may need a thicker specimen in order to obtain several fringes, whereas for higher frequencies the required variation in an acoustic path may be smaller, provided that the attenuation is low enough to have a signal-to-noise ratio greater than one.

Acoustic interferometric techniques may be realized by means of a proper experimental setup or a specific interferometric device. In the next section, we discuss the first case, in which interference of acoustic waves is realized and the interference pattern is directly sensed by a transducer. The following section is devoted to the description of indirect interferometric techniques, which analyze the interference in an electronic device (e.g., an oscilloscope) between the electric signal feeding the transducer and the one induced by the acoustic wave at the receiver. Then the operational principles of quadrature detector systems, which may be used to improve the performance of interferometers, are briefly described. Since in recent years great effort has been devoted to developing techniques based on optical sensors and interferometers, in the last section a brief comparison is made between acoustic and other interferometric techniques.

DIRECT ACOUSTIC INTERFEROMETRIC TECHNIQUES

In direct interferometric techniques the interference occurs among acoustic waves and the interference pattern is sensed by a transducer. One of the earliest demonstrations of acoustic interference was carried out by Quincke in 1866 (3) by using an apparatus suggested by J. Herschel. It consisted of a tube, which branched into two tubes differing in length by an odd number of half wavelengths of sound in air and joining again at the other end into a single tube. Stewart (3) modified this interferometer by using equal-length branches but varying the frequency in order to obtain interference. A more sophisticated tool for measuring the velocity in gases consisted of a column formed by an ultrasonic transducer and an adjustable reflecting piston (4).

In more recent times, a simple setup, called a standing wave acoustic interferometer, has been used for the characterization of molten materials, such as metals or alloys, and dense fluids (5,6). A continuous wave (cw) is injected from the left side of a sandwich, composed of two external layers (A) of known properties and an intermediate layer (B) of the molten or fluid material, whose acoustic properties need to be determined (Fig. 1). The output signal is detected by a receiver (R) at the right side of the setup. The impedance mismatch at the interfaces between the materials A and B generates multiple reflections that interfere with each other. A resonant constructive standing wave is formed when the thickness x of the layer B is equal to integer multiples of half wavelengths, while antiresonant conditions

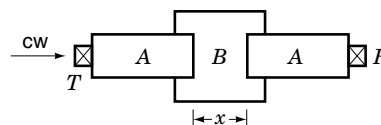


Figure 1. Scheme of a standing wave acoustic interferometer for the determination of the acoustic properties of the material B.

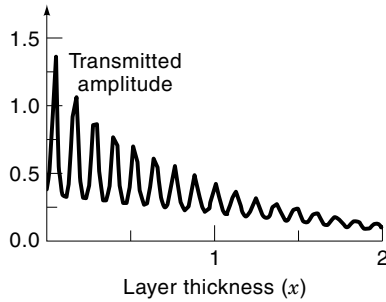


Figure 2. Interference pattern obtained with the standing wave acoustic interferometer.

occur when

$$x = (n + 1/2) \frac{\lambda}{2} \quad (1)$$

By varying the thickness x (and the geometry) of B, the transmitted ultrasonic signal traces out an interference pattern with alternating maxima and minima (Fig. 2). By fitting the curve amplitude versus thickness, according to simple theoretical expressions, the phase velocity and attenuation factor may be easily calculated.

A particularly suitable setup consists of two movable metal rods partially submerged in a tank containing the molten material. Thus the thickness of B can be easily varied with continuity. In this case, the analytical expression for the wave amplitude at the receiver is given by (5)

$$\frac{1}{|A|} = \frac{1}{4} \sqrt{\frac{2(4 + H) \cosh\left(\frac{\beta x}{Q}\right) + 8F \sinh\left(\frac{\beta x}{Q}\right)}{+ 2(4 - H) \cos(2\beta x) - 8G \sin(2\beta x)}} \quad (2)$$

where F , G , and H are functions of the acoustic impedances of the rod and the molten material and β and Q are the real part of the wave number in the molten material and the quality factor, respectively. Both may be evaluated through a least-squares fitting of the experimental data (Fig. 2) with Eq. (2). Thus, the same measurement allows one to compute the phase velocity v_B and the specific attenuation factor α for the medium B:

$$v_B = \frac{\omega}{\beta}; \quad \alpha = \frac{\beta}{2Q} \quad (3)$$

where ω is the known carrier-wave frequency.

Several effects may affect the accuracy of the standing wave acoustic interferometric technique (e.g., beam spreading in layer B or lack of parallelism and flatness of the A faces). However, the technique is rather accurate as far as the measurement of the velocity is concerned.

A similar approach has been used (7) for determining the speed of sound in gases by evaluating the resonant frequencies in a cylindrical interferometer. A transducer is used for generating an acoustic wave in a gas inside an acoustical cavity. Longitudinal modes are formed, due to interference between multiple reflecting waves, with resonance frequencies

$$f_n = \frac{u}{2L}(n + \delta) \quad (4)$$

where L is the length of the cavity, u the sound speed, n the order of the resonance, and δ a phase shift correction due to the finite diameter of the cavity. The resonant frequencies are measured through an oscilloscope, and a fitting with Eq. (4) yields optimal values for the parameters u and δ .

INDIRECT ACOUSTIC INTERFEROMETRIC TECHNIQUES

In indirect acoustic interferometric techniques the interference occurs between the electric signal feeding the transducer and the one induced by the acoustic wave at the receiver. The simplified circuit diagram (neglecting almost all electronic components) of a typical interferometric device (8) is reported in Fig. 3. A generator feeds a signal into a doubler (not reported in Fig. 3), which splits the signal into two components, called reference and testing signals. The latter is amplified by a power amplifier and converted into an ultrasonic pulse by a piezoelectric (or other) transducer. The signal detected by the receiver passes through a low-noise amplifier and is recombined in the region H with the unperturbed reference signal, thus producing interference. In Fig. 3(b), a zooming of the region H is reported for a two-channel quadrature detection system (described later) (9). Before interference the reference signal is split in A into two arms in quadrature. Then each one is combined in C with the testing signal split in B into two arms in phase. The two channels are usually named cosine and sine channels.

This technique is particularly efficient for evaluating the variations of a physical parameter of the specimen due to external perturbations. For such a purpose, the frequency of the reference signal is modulated in order to maintain quadrature (i.e., a destructive interference) when the intensity of the

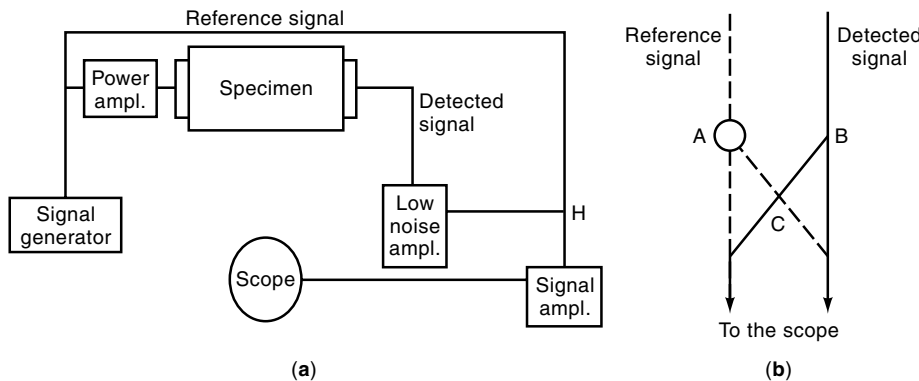


Figure 3. (a) Simplified circuit diagram for an indirect acoustic interferometric setup. (b) Scheme of the quadrature detection system.

perturbation is changed. The fractional change in the interferometer frequency with respect to the nonperturbed case gives, through simple relationships, the variation of the physical parameter.

To clarify the method, we discuss its use (10) for the determination of the acoustic nonlinearity parameter B/A , defined as

$$\frac{B}{A} = 2\rho_0 c_0 \frac{\Delta c}{\Delta p} \quad (5)$$

where ρ_0 is the medium density, c_0 the sound velocity, and Δc the velocity variation resulting from a pressure change Δp for an isentropic transformation. A sample material maintained at a constant temperature is pressurized with different values of p . For each pressure, the effective path length between the transducers changes due to a variation of the sound velocity in the propagation medium. An adjustment of the transmitted frequency is necessary in order to maintain quadrature. Then the parameter B/A can be easily evaluated from the fractional change in the interferometer frequency $\Delta f/f_0$ necessary to maintain quadrature:

$$\frac{B}{A} = 2\rho_0 c_0^2 \frac{\Delta f}{f_0 \Delta p} \quad (6)$$

where f_0 is the reference frequency. In this situation the velocity variation is measured with a very high accuracy (up to 1%).

A similar approach is based on the measurement of the cross-correlation function between the signals from the receiver (R) and the transmitter (T) (8,11–19). For example (11,12), continuous waves of suitable frequency (1 MHz to 20 MHz) are split into two parts. One portion of these waves excites an immersion transducer, while the other (reference signal) is fed into a phase-sensitive detector. The specimen is submerged in water and oriented in such a way that the incident beam scans its surface at a fixed angle. Receiver and transmitter are mounted on the same bridge and initially located and oriented so that the received signal (testing signal) is maximum. The amplitude of the signal resulting from the crosscorrelation between the reference and testing signals is measured for different positions of the transducer. It depends, of course, on the length of the effective acoustical path of the testing wave from T to R , leading to plots similar to the one reported in Fig. 2.

When the geometry of the specimen is properly defined, this crosscorrelation technique leads to results that can be easily interpreted in terms of the relevant physical variables. A particularly interesting case occurs when the material specimen is cut in the shape of a wedge (11) (Fig. 4). In fact, if the material

is homogeneous, the only difference between waves entering the wedge at different locations (e.g., T and T' in Fig. 4) is their length y of propagation path within the material, since all other factors, such as boundary conditions, remain identical. Therefore, the results depend only on the interference effects and are not affected by the angular dependence of reflection, refraction, and mode conversion, which in general greatly complicate the physical picture.

Assuming that the wedge angle is small, the ultrasonic signal S_T , sent by the transmitter, and S_R , received by the transducer R , are given by

$$\begin{aligned} S_T &= F_T A_0 \cos(\omega t + \Phi_0) \\ S_R &= T_{12} T_{21} F_R A_0 \exp[-\alpha_1(l-y) - \alpha_2 y] \\ &\quad \cos\left(\omega\left\{t - \left[\frac{l-y}{v_1}\right] - \frac{y}{v_2}\right\} + \Phi_0\right) \end{aligned} \quad (7)$$

where F_T and F_R are the response factors of the transmitter and of the receiver and associated electronics; T_{12} and T_{21} are the transmission factors between water and the specimen material and vice versa; α_1 , α_2 , v_1 , and v_2 are the attenuation coefficients and phase velocities in water and in the material, respectively; A_0 is the initial amplitude of the cw; and l is the fixed distance between the transducers T and R .

After some algebraic manipulation, one finds that the cross-correlation factor between the two signals S_T and S_R is given by (11)

$$\Psi = \Psi_0 + C \exp(-\alpha y) \cos(by + \Phi) \quad (8)$$

where Ψ_0 , C , α , b and Φ are parameters that can be found by least-squares fitting the experimental curves (see, e.g., Fig. 5a). Since

$$\alpha = \alpha_2 - \alpha_1 \quad (9)$$

$$\beta = \omega \left(\frac{1}{v_2} - \frac{1}{v_1} \right) \quad (10)$$

one immediately obtains the attenuation coefficient and phase velocity for the material, since the values of the corresponding quantities for water are well known. It is worthwhile to remark that Fig. 5(a) yields α_2 and v_2 by simple inspection, since α_2 is given by the coefficient of the envelope curve $\exp(-\alpha y)$ and v_2 by the distance between successive maxima and minima.

This technique may also be used for the detection of flaws in the material specimen. In fact, in the presence of defects the crosscorrelation plot is locally modified [see Fig. 5(b)] and from the difference between the two plots (the so-called signature of the defect) the main features of the flaw may be evaluated.

Since for practical purposes material plates are much more important than wedges, the method has been extended to the treatment of plates (12). The main ingredient (i.e., the continuous variation of the acoustic path in the material specimen y) is obtained by slowly varying the angle of incidence. Also in this case it is possible to obtain the attenuation coefficient and phase velocity of the material through a best fitting of the theoretical formula. The fitting formulas are, however, more

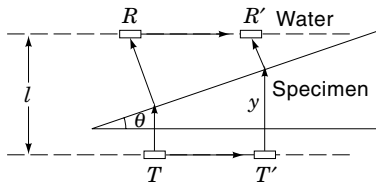


Figure 4. Experimental setup for indirect interferometric measurements of the acoustic slowness and attenuation in a wedge-shaped material specimen.

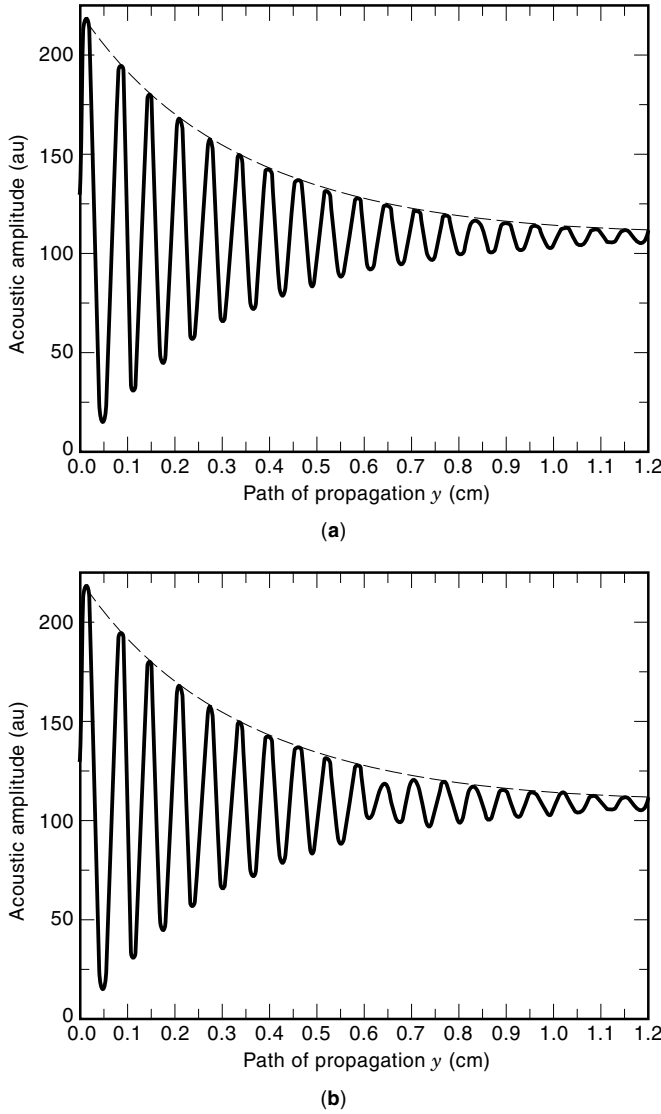


Figure 5. Typical experimental crosscorrelation curves obtained using the wedge-shaped specimen shown in Fig. 4. (a) Specimen without defects. (b) Specimen with a hole drilled into it. The frequency is in both cases $f = 11$ MHz.

complicated and the measured attenuation is no longer absolute, since there are boundary effects depending on the incidence angle.

QUADRATURE DETECTOR SYSTEMS

A quadrature detector system is an interferometer based on comparison between the correlations of the testing signal with two reference signals in quadrature in two different channels (cosine and sine channels). To illustrate the operation of an ideal quadrature detector system (see Fig. 3), let us consider a wave train of angular frequency ω and wavelength λ crossing a path length L in a medium with propagation velocity c . If $w = \cos(\omega t)$ is the transmitted signal, the received signal is

in the form

$$s = A \cos\left(\omega t - \frac{2\pi L}{\lambda}\right) \quad (11)$$

where A is its amplitude. The testing signal is then demodulated in the two channels of the quadrature detector with two reference signals $u = 2 \cos(\omega t)$ and $v = -2 \sin(\omega t)$. Neglecting high-frequency contributions (filtered out by a low-pass filter), the two detected signals are $u_1 = A \cos(2\pi L/\lambda)$ and $u_2 = A \sin(2\pi L/\lambda)$. The pair (u_1, u_2) are components of a vector $\mathbf{u}$ with constant modulus, which rotates when L varies. The number of fringes η crossed by the vector $\mathbf{u}$ can be easily counted by recording the zeroes of u_1 and u_2 . The change in path length is then given by

$$\Delta L = (\eta + F) \frac{\lambda}{4} \quad (12)$$

where F is the fractional number of fringes, obtained by interpolation between the counts.

All interferometers with quadrature detector systems are affected by a set of errors (9), which in many cases severely limit the precision and accuracy of the measurements. These errors are due to a lack of quadrature between the reference signals (the phase shift is not exactly $\lambda/4$), to an unequal gain in the detector channels, and to an error in the zero offsets of the device. Moreover, in real systems, due to electronic and mechanical overtones (nonlinearity), higher-harmonics effects are always present and will severely affect the evaluation of transient times. Therefore, while for an ideal quadrature detector system the precision is limited only by the system noise, in real cases the endpoint of the vector $\mathbf{u}$, is no longer on a circle and F is affected by an error, causing errors in ΔL up to a few percent of λ , which is very large compared to the accuracy of a good interferometer. If r is the channel gain ratio, p and q are the offsets in the cosine and sine channels, and α is the reference signal quadrature error, the output signal from the detector becomes

$$u_1^D = u_1 + p \quad (13)$$

$$u_2^D = \frac{1}{r}(u_2 \cos \alpha - u_1 \sin \alpha) + q \quad (14)$$

describing an ellipse. The real signals can, however, be easily reconstructed from the ones used in Eqs. (13) and (14), since we are considering systematic errors and therefore the corrections p , q , r , and α must be calculated only once. By several measurements of u_1^D and u_2^D for a few sets of known values of ΔL , the parameters p , q , r , and α can be obtained by least-squares fitting the distorted ellipse equation to the experimental data. Such an approach allows one to reach a resolution on the order of 10^{-8} m, 100 times higher than the one obtained with uncorrected data.

By using a similar approach, it is possible to introduce corrections taking into account the transformation of the ellipse into a deformed ellipse due to the presence of higher-order harmonics. Such corrections are obtained by using the least-squares method iteratively (20).

OTHER INTERFEROMETRIC TECHNIQUES

There are many other types of devices, e.g., acoustic spectrometers (21) and acoustic microscopes (22) which use the underlying principles of interferometry for their operation but have not been labeled as interferometers. Acoustic spectrometers are used to measure frequencies of mechanical resonances (standing waves in the thickness of the specimen) caused by interference between the normal incident acoustic waves and the ones reflected from the back surface of a parallel plate. These resonances are obtained by varying the frequency of the incident plane wave or by varying the thickness (e.g., corroded steel plates) for a broadband incident pulse. The resonant frequencies and Q -factor of these resonances are used to compute phase velocity and attenuation. Both of these quantities can be used to characterize the material.

The acoustic microscope is also an example of a device based on interference phenomena. It can be designed for various applications to materials and tissue characterization. For example, a specially designed acoustic lens (with aperture angle larger than Rayleigh critical angle) facilitates the oscillations in a quantity $V(z)$ which is the record of the modulus of the measured voltage as a function of the z distance between the focal point of the acoustic lens and the specimen surface. These oscillations are due to interference between the normal specular reflection from the fluid-loaded surface of a solid specimen and a ray passing through the Rayleigh critical angle. The velocities and attenuation of elastic wave propagation mode can be obtained from the periodicity and decay of $V(z)$. Theoretical fit of the velocities with those obtained from $V(z)$ determines elastic constants which are useful for characterization of thin films deposited on an elastic substrate (see Achenbach et al., Ref. 23).

Different types of interferometers based on optics have also been developed (e.g., the Mach interferometer) for the purpose of detecting the gas flow in supersonic wind tunnels. Displacements of the interference fringes in the optical wavefront, traveling normal to the gas flow and the reference beam, can be used to determine the density change in the gas flow. A coupling of the light beam to the specimen (gas flow) does not perturb its initial phase. Conversely, ultrasonic nondestructive evaluation may require the use of contact transducers, which perturb the system and therefore the accuracy of specific measurements. For instance, wave attenuation measurements are severely limited by the mode conversion of the wave at the interface between the medium and the transducer. In fact, such an effect becomes, in many cases, the dominant contribution to the measured attenuation factor. To overcome this difficulty, considerable effort has been devoted to the development of optical techniques for the generation and detection of ultrasonic waves. Such techniques are based on the so-called photoacoustic effect; for example, the generation of an ultrasonic pulse through a laser beam incident on an elastic specimen (24) and the evaluation of the acoustic wave amplitudes through modulations of the optical phase of a light beam (25–31). In the latter, the phase shift may be demodulated to a change in optical intensity through an optical fiber interferometer. The acoustic wave amplitude and/or the acoustic properties of the medium can be easily obtained from these measurements. Optical methods are particularly suitable for determination of the dispersion law of the medium, since optical interferometers measure directly the phase shift for any given frequency of the induced wave.

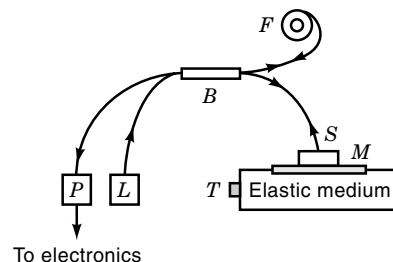


Figure 6. Simple scheme for a fiber optics sensor.

Without going into details, which are beyond the scope of this article, we illustrate the photoacoustic effect in Fig. 6, based on a Michelson-type interferometer (25–27). Let us consider an elastic medium with surfaces polished or coated with a reflective tape M [e.g., Mylar (32) tape] in order to increase its reflectivity. A transducer T driving a continuous plane wave produces standing waves at discrete frequencies. A light beam generated by a polarized laser L is launched into a fiber optic system through a beam splitter B . One arm (reference arm) is launched into a fixed-length light guide F . The other (signal) arm is sent to the sensor S , where it is collimated through a cylindrical lens and reflected off the vibrating surface. The reflected back reference and signal arms are then brought to interfere with each other and an intensity-varying output signal is created. A photodiode P connected to electronic devices can then measure the changes in light intensity.

If the specimen is enclosed in a rigid tube (e.g., water in a Plexiglas tube; Ref. 33), since the external surface is not sensitive to vibrations, a membrane must be inserted in a hole of the tube in order to sense the pressure variations in the enclosed medium. The standard setup shown in Fig. 6 may be used to implement different techniques by means of fiber optics or laser interferometers. They may achieve a very high sensitivity for displacements, of the order of $1 \text{ \AA}$ for the Michelson interferometer, and a flat bandwidth response, from 1 kHz to 20 MHz in the case of the Mach-Zender interferometer.

The Mach-Zender interferometer may be either bonded to the surface (29) or embedded in the specimen (30,31). In both cases, the acoustic wave displacement induces a pressure on the light guides (31), leading to axial compression/tension and, therefore, to a phase shift of the propagating light beam. It is sensitive to both in-plane and out-of-plane displacements, but in the case of composites, it has been found to be insensitive to small angular misalignments of fibers with respect to the incoming wave front. The embedded configuration is shown to be several times more sensitive to the wave motion (up to 12 Pa per 1 m length of lightguide in the measuring channel) (e.g., in the case of Lamb waves) than the surface-bonded one. However, the latter configuration may be more practical.

A dual-probe interferometer (34) is based on the interference of two signals received from two fiberoptic receivers located in different positions on the surface of the specimen. A similar approach has been used by Clark, Delsanto, and Mignogna (35,36) using piezoelectric transducers for the determination of times of flight of ultrasonic Rayleigh waves in acoustoelasticity applications (texture and stress measurements in polycrystalline aggregates). However, this technique could be problematic since the signal is affected by the contact transducer it must pass. A different approach, based on the Doppler effect, is at the basis

of the Fabry–Perot interferometer (37). If an elastic specimen is subject to vibrations, the reflected light is Doppler shifted and the wavelength shift is related to the surface velocity of the target.

Two major problems lead to signal fading in optical fiber interferometers. The first is represented by a fluctuation in the state of optical polarization within the optical fiber and can be removed by using polarization control or polarization-maintaining fibers. The second concerns changes in the optical phase difference between the interfering channels due, for example, to thermal effects. As a consequence, the device sensitivity can drop to zero when the measurement is performed. To avoid variations in sensitivity, several stabilization schemes have been developed. In the case of homodyne systems, feedback is used in order to lock the interferometer at quadrature; heterodyne techniques allow the environmental noise to be effectively cancelled, but lead to a rise in the demodulation circuitry complexity. Passive stabilization schemes are very efficient but require complex signal processing.

In general, photoacoustic transducers and sensors present several disadvantages compared with piezoelectric transducers. The efficiency of sound generation by thermal sources is rather low and the sensitivity of ultrasonic detection is poorer. However, since these are noncontact devices, they can scan curved surfaces conveniently and measure displacement. Since the optical wavelength is several orders of magnitude smaller than the wavelength of acoustic pulses, the resulting accuracy in the determination of the displacement amplitudes becomes correspondingly better.

SUMMARY

Acoustic interference is due to the superposition of two or more initially coherent sound waves, which are superimposed after they have traveled different acoustic paths. Acoustic interferometry has been used for determining material properties and their dependence on parameters such as temperature or pressure; or the presence of impurities or defects by determining the phase velocity and/or attenuation as a function of these variables. When the geometry of the specimen is properly defined, the phase velocity and attenuation in the specimen can be determined accurately as a function of the physical parameters by using a suitable acoustic interferometer. We have shown various examples of configurations of different interferometers and discussed their usefulness in studying the properties of fluids as well as of solids.

BIBLIOGRAPHY

1. E. P. Papadakis, Ultrasonic velocity and attenuation: Measurement methods with scientific and industrial applications. In W. Mason and R. Thurston (eds.), *Physical Acoustics*, vol. 12, New York: Academic Press, 1976, pp. 277–374.
2. C. M. Fortunko, et al., Absolute measurements of elastic wave phase and group velocities in lossy materials, *Rev. Sci. Instrum.*, **63**: 3477–3486, 1992.
3. Alexander Wood, *Acoustics*, New York: Dover, 1966.
4. T. F. Hueter and R. H. Bolt, *Sonics*, New York: Wiley, 1955.
5. K. W. Katahara, et al., An interferometric technique for measuring velocity and attenuation in molten rocks, *J. Geophys. Res.*, **86**: 11779–11786, 1981.
6. P. M. Nasch, M. H. Manghnani, and R. A. Secco, A modified ultrasonic interferometer for sound velocity measurements in molten metals and alloys, *Rev. Sci. Instrum.*, **65**: 682–688, 1994.
7. M. B. Ewing, M. L. McGlashan, and J. P. M. Trusler, The speed of sound in gases, *J. Chem. Thermodynamics*, **17**: 549–559, 1985.
8. C. H. A. Huan, et al., High-sensitivity ultrasonic interferometer for the detection of magnetic phase transitions, *J. Appl. Phys.*, **61**: 3193–3195, 1987.
9. P. L. M. Heidemann, Determination and correlation of quadrature fringe measurement errors in interferometers, *Appl. Optics*, **20**: 3382–3385, 1981.
10. E. C. Everbach and R. E. Apfel, An interferometric technique for B/A measurement, *J. Acoust. Soc. Amer.*, **98**: 3428–3438, 1995.
11. N. K. Batra and P. P. Delsanto, Ultrasonic interferometric characterization of highly attenuative materials, *J. Acoust. Soc. Amer.*, **85**: 1167–1172, 1989.
12. N. K. Batra, et al., Ultrasonic interferometric characterization of material plates, *Proc. IEEE Ultrason. Symp.*, Cannes, France, pp. 1215–1218, 1994.
13. A. P. Sarvazyan and T. V. Chalikian, Theoretical analysis of an ultrasonic interferometer for precise measurements at high pressures, *Ultrasonics*, **29**: 119–124, 1991.
14. S. P. Klimov, V. V. Tyutekin, and A. E. Vovk, Automated acoustic interferometer, *Meas. Techniques*, **32**: 1198–1201, 1989.
15. V. A. Sukatskas, et al., Density measurement of a liquid with an ultrasonic interferometer of constant length, *Meas. Techniques*, **31**: 1126–1129, 1988.
16. I. V. Lebedeva and S. P. Dragan, Determination of acoustic characteristics in tubes by means of two microphones, *Meas. Techniques*, **31**: 806–807, 1988.
17. Yu. S. Il'inykh, V. I. Levtsov, and S. S. Sekoyan, Analysis of dynamic characteristics of an ultrasonic interferometer with astatic external control system, *Meas. Techniques*, **31**: 597–601, 1988.
18. L. Cusco and J. P. M. Trusler, Identification of environmentally acceptable low-sound speed liquids, *Int. J. Thermophys.*, **16**: 675–685, 1995.
19. Y. N. Barabenenkov and V. I. Pasechnik, A study of correlations properties of thermal acoustic radiation, *Acoust. Phys.*, **41**: 494–496, 1995.
20. K. H. C. Jhung, et al., Effects of higher-order harmonics on phase determination in an ultrasonic interferometer, *J. Acoust. Soc. Amer.*, **91**: 2025–2029, 1992.
21. A. Migliori and J. L. Sarro, *Resonant Ultrasound Spectroscopy*, New York: Wiley, 1997.
22. A. Briggs, *Acoustic Microscopy*, New York: Oxford University Press, 1992.
23. J. D. Achenbach, J. O. Kim, and Y.-C. Lee, Measuring thin-film elastic constants by line-focus acoustic microscopy, in Andrew Briggs (ed.), *Advances in Acoustic Microscopy*, New York: Academic Press, 1995, Volume 1, pp. 153–208.
24. T. W. Murray, J. B. Deaton, Jr., and J. W. Wagner, Experimental evaluation of enhanced generation of ultrasonic waves using an array of laser sources, *Ultrasonics*, **34**: 69–77, 1996.
25. S. A. Cheyne, C. T. Stebbings, and R. A. Roy, Phase velocity measurements in bubbly liquids using a fiber optic laser interferometer, *J. Acoust. Soc. Amer.*, **97**: 1621–1624, 1995.
26. S. G. Pierce, B. Culshaw, and W. R. Philp, Synchronized triggering for ultrasonic transient detection using optical fiber interferometers and application to the measurement of broadband Lamb wave dispersion characteristics, *Meas. Sci. Technol.*, **7**: 1665–1667, 1996.
27. M. G. Somekh, et al., An accurate noncontacting laser based system for surface wave velocity measurement, *Meas. Sci. Technol.*, **6**: 1329–1337, 1995.

28. S. G. Pierce, et al., Broadband Lamb wave measurements in aluminum and carbon/glass fiber reinforced composite materials using non contacting laser generation and detection, *Ultrasonics*, **35**: 105–114, 1997.
29. S. G. Pierce, et al., Surface-bonded and embedded optical fibers as ultrasonic sensors, *Appl. Opt.*, **35**: 5191–5197, 1996.
30. W. J. Staszewski, et al., Wavelet signal processing for enhanced Lamb wave defect detection in composite plates using optical fiber detection, *Opt. Eng.*, **36**: 1877–1888, 1997.
31. S. X. Short, et al., An experimental study of acoustic vibration effects in optical fibers current sensors, *IEEE Trans. Power Deliv.*, **11**: 1702–1705, 1996.
32. Mylar: Manufactured by DuPont de-Nemours, Barley Mill Plaza, Wilmington, DE, 1980.
33. Plexiglas: Manufactured by AtoHaas Americas Inc., 100-T Independence Mall W., Philadelphia, PA 19105.
34. H. P. Ho, et al., Direct and indirect dual-probe interferometers for accurate surface wave measurements, *Meas. Sci. Technol.*, **5**: 1480–1490, 1994.
35. P. P. Delsanto and A. V. Clark, Jr., Rayleigh wave propagation in deformed orthotropic materials, *J. Acoust. Soc. Amer.*, **81**: 952–960, 1987.
36. P. P. Delsanto, R. B. Mignogna, and A. V. Clark, Jr., Ultrasonic texture and stress measurements in anisotropic polycrystalline aggregates, *J. Acoust. Soc. Amer.*, **87**: 215–224, 1990.
37. D. A. Oursler and J. W. Wagner, Full-field vibrometry using the Fabry-Perot etalon interferometer, *Appl. Opt.*, **31**: 7301–7307, 1992.

N. K. BATRA
Naval Research Laboratory
P. P. DELSANTO
A. ROMANO
M. SCALERANDI
Politecnico di Torino

MASS SPECTROMETERS

APPLICATIONS OF MASS SPECTROMETRY

Mass spectrometers precisely determine the masses of ionized atoms or molecules with extreme sensitivity (1,2). Such investigations allow the following measurements.

The Identification of Atoms or Molecules

From a reasonably resolved mass spectrum, the original atoms or molecules (see Fig. 1) are identified by determining their atomic or molecular weight precisely. From such a mass spectrum the binding energies of different isotopes of one or of several elements are also determined (3,4) if the mass-resolving power $m/\Delta m$ of the mass analyzer is large enough (see also the section "The Magnetic Ion Trap Mass Spectrometer").

The Investigation of Macromolecules

The mass analysis of macromolecules is helpful for investigating very large biomolecules (see Fig. 2) and also atomic or molecular clusters. The goal of such investigations is to precisely determine the molecular weight.

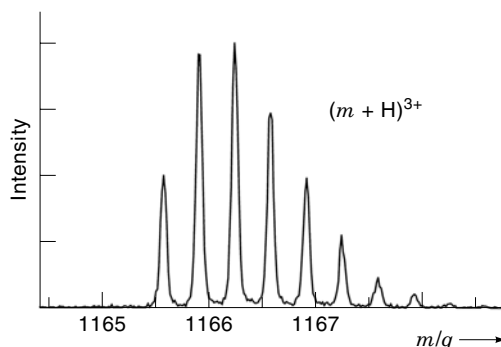


Figure 1. Mass spectrum of β -chain insulin of mass $m \approx 3494.9$ u, recorded as $(m + H)^{3+}$, that is, with one proton attached. This mass spectrum was recorded in a time-of-flight mass spectrometer (see Fig. 9) of mass resolving power $m/\Delta m = 17000$ (FWHM), that is, measured as the full line width at half maximum. Since carbohydrate molecules all contain about one ^{13}C -atom for every one hundred ^{12}C -atoms there is a mass multiplet of ions for every molecule. When the molecule is large, the probability is high that it contains one or several ^{13}C -atoms. Thus, the most abundant molecule is usually not the one that contains only ^{12}C -atoms.

The Determination of the Structure of Molecules

To understand the structure of complex organic molecules one can isolate ions of one specific mass in a first-stage mass analyzer, fragment this ion by gas collisions in some intermediate gas cell, and analyze the mass spectrum of the molecule fragments in a second-stage mass analyzer. This (MS/MS) technique (5) is used, for instance, in the amino acid sequencing of proteins (6).

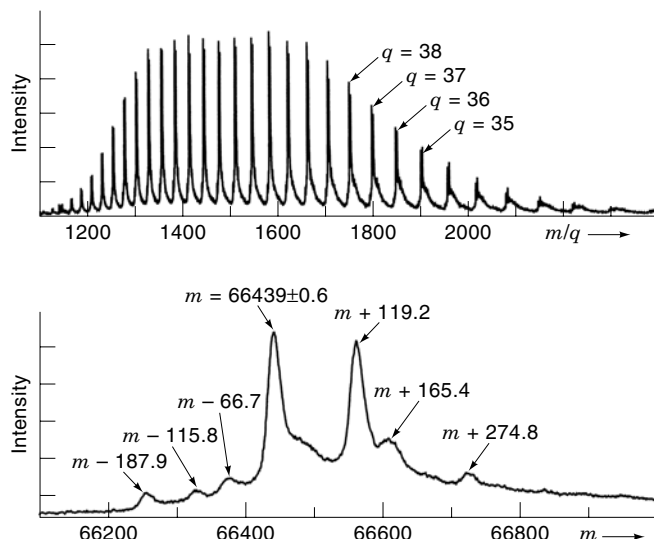


Figure 2. The mass spectrum of ≈ 30 to ≈ 55 times charged ions of albumin. This molecule has a molecular weight of about $(66,439 \pm 0.6)$ u. The time-of-flight mass spectrometer used for this investigation had only a mass resolving power $m/\Delta m \approx 12,000$ so that individual ion masses could not be separated from each other. Also shown is the mathematical combination of all of these mass distributions, from which one determines the molecular mass of albumin with certain adducts with a precision of only 10 ppm. The main adducts here had a mass of $m_1 = 119.2$ u.

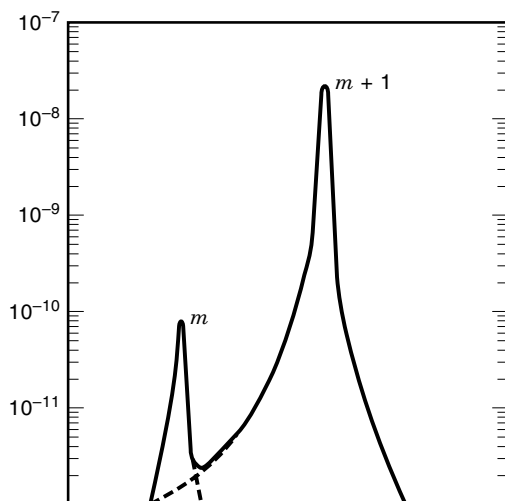


Figure 3. Mass cross-contamination caused by the tails of an intense neighboring mass line. The shown spectrum was recorded by a rather good sector field mass analyzer that had a (FWHM) mass-resolving power of $m/\Delta m \approx 5000$ and was operated at a pressure of only a few times 10^{-8} mbar.

The Measurement of Isotopic Distributions

When some isotopes of a certain element decay radioactively, the isotopic intensity distribution reveals the age of the sample. Some of the isotopes exist only in very small quantities while others are abundantly available. For this reason it is very important to use a mass analyzer that provides for a rather small mass cross-contamination from one mass to the next. The system must be designed properly for this purpose, and it must have a very good vacuum to reduce residual gas scattering of the most abundant ion species (see Fig. 3). Especially low mass cross-contamination is achieved by using high-energy ions accelerated in tandem accelerators (7). In such systems, for instance, the isotopic ratio $^{14}\text{C}/^{12}\text{C}$ can be investigated. This is important because ^{14}C , which has a half-life of 5730 years, is constantly produced in the earth's atmosphere by the sun, so that the $^{14}\text{C}/^{12}\text{C}$ ratio is a good measure for the age of old natural products, such as wood or bones.

The Detection of Small Amounts of Specific Atoms or Molecules

To investigate the pollution, for instance, of water or air samples one can use the high sensitivity and specificity of mass spectrometers and combine them with the selectivity of gas chromatographs (GC/MS), liquid chromatographs (LC/MS), or of complex but powerful capillary zone electrophoresis (CZE) systems. In such chromatographs the molecules are separated according to their chemical adsorption properties (8), that is, simultaneously injected chemically different molecules leave the chromatograph successively. A specific substance is identified in such a system by the mass spectrum of the chromatograph's effluent recorded at a specific time (see Fig. 4).

FUNDAMENTALS OF MASS SPECTROMETRY

The very diverse applications described are all united by the use of the same instrumentation. Thus it is neces-

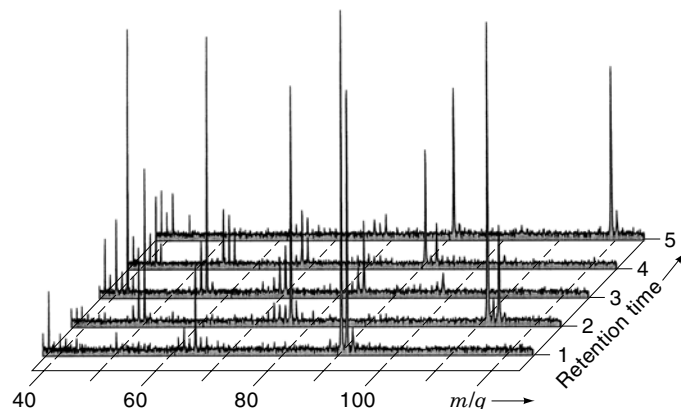


Figure 4. The record of a GC/MS investigation of traces of about 10^{-12} g of (1) toluene, (2) chlorobenzene, (3) nonane, (4) chloroheptane, and (5) *t* butylbenzene. For such a measurement several thousand or 10,000 mass spectra are recorded of the chromatograph effluent, and each mass spectrum is characteristic of the effluent at a specific time. For clarity, however, here only one mass spectrum is shown for each GC peak. A typical record, as shown here, requires several hundred seconds and in some cases up to a few thousand seconds because one must wait until the substance with the longest retention time leaves the gas chromatograph.

sary to describe the mass spectrometric techniques in some detail.

An atomic or a molecular ion to be mass analyzed is characterized by

1. its mass m measured in *mass units* u , that is, the mass of one-twelfth of a ^{12}C -atom;
2. its charge q measured in *charge units*, that is, the negative value of the charge of one electron or $\approx 1.6022 \times 10^{-19}$ C;
3. its kinetic energy $K = qV$ measured in electron volts where V is the potential difference by which the ion has been accelerated; and
4. its velocity v measured in kilometers per second or in millimeters per microsecond (9) is given by

$$v \approx 9.82269 \sqrt{\frac{2K}{m}} \quad (1)$$

The numerical multiplier is found from $c/\sqrt{m_u} \approx 299,792,458/\sqrt{931,494,300}$ where c is the velocity of light in millimeters per microsecond and m_u is the energy equivalent of one mass unit in electron volts.

MASS ANALYZERS

To distinguish ions of different masses, the electromagnetic fields used can either provide a lateral dispersion, that is, a mass-dependent beam deflection or a longitudinal dispersion, that is, a mass-dependent flight-time difference. In the second case one must use a chopped and bunched ion beam, whereas in the first case bunched ion beams and dc beams can be used.

Laterally Dispersive Mass Analyzers

Laterally dispersive mass analyzers consist mainly of a magnetic deflecting field of flux density B measured in Tesla (T). In

a magnetic flux density B , nonrelativistically fast ions ($v \ll c$) of mass m , energy K , and charge q move along radii ρ measured in millimeters according to the expression

$$B\rho \approx \frac{\sqrt{2Km}}{9.82269q} \quad (2)$$

Usually the ions produced are all accelerated by the same potential difference V_0 so that their kinetic energies are all $K_0 = qV_0$. Thus their velocities $v_0 \approx 10\sqrt{2K_0/m}$ are also the same. However, because of the ionization process, the ions always have an energy spread $\pm\Delta K$ which is usually smaller than 1 eV.

When ions of different masses $m = m_0 \pm \Delta m$ and of a range of energies $K = K_0 \pm \Delta K$ enter into a magnetic sector field, the ions are finally separated from the beam axis by $\pm[(x_B|m)(\Delta m/m_0) + (x_B|K)(\Delta K/K_0)]$ with $(x_B|m) = (x_B|K)$ for a given magnetic field. Here $\pm(x_B|m)(\Delta m/m_0)$ determines the desired mass separation between ions of masses $m_0 + \Delta m$ and $m_0 - \Delta m$, and $(x_B|K)(\Delta K/K_0)$ causes a detrimental beam widening because of the continuous distribution of ion energies from $K_0 - \Delta K$ to $K_0 + \Delta K$. This energy spread also causes the ions to diverge from the ion source at angles of divergence given by

$$\gamma_0 = \pm \frac{v_i}{v_0} = \pm \sqrt{\frac{\Delta K}{K_0}} \quad (3)$$

because the velocity distribution of the unaccelerated ions is isotropic.

For a given magnetic field this widening of a mass line caused by the ions' energy distribution is unavoidable. However, for a combination of two or more sector fields, one can arrange the fields such that the energy dispersions compensate for each other, and the overall mass dispersion is $(\bar{x}|m)\Delta m \neq 0$. For a two-sector field system there are two possible arrangements (see Fig. 5):

1. The first solution uses an electrostatic sector field in addition to the magnetic sector which is dimensioned so that the forwardly calculated energy dispersion $(\bar{x}_E|K)(\Delta K/K_0)$ of the electrostatic sector field is equal to the backward calculated energy dispersion $(\bar{x}_B|K)(\Delta K/K_0)$ of the magnetic sector field (9,10), as indicated in Fig. 5(a). Because the electrostatic sector field has no mass dispersion, the mass dispersion of the magnetic sector field is also the mass dispersion of the full system.
2. The second solution uses a small magnetic sector field to compensate for the energy dispersion of the main magnetic sector (11), as indicated in Fig. 5(b). When the two sector magnets operate at different potentials and the energies K_1 and K_2 denote the ion energies in the two fields, one must postulate that the forwardly calculated energy dispersion $(\bar{x}_{B1}|K)(\Delta K/K_1)$ of the first magnetic sector field equals the backward calculated energy dispersion $(\bar{x}_{B2}|K)(\Delta K/K_2)$ of the second sector field (11). However, such a combination of two sector fields is useful only for $K_1 \neq K_2$ because only in this case the mass dispersions of the two sector magnets do not cancel each other for the case of a vanishing energy dispersion.

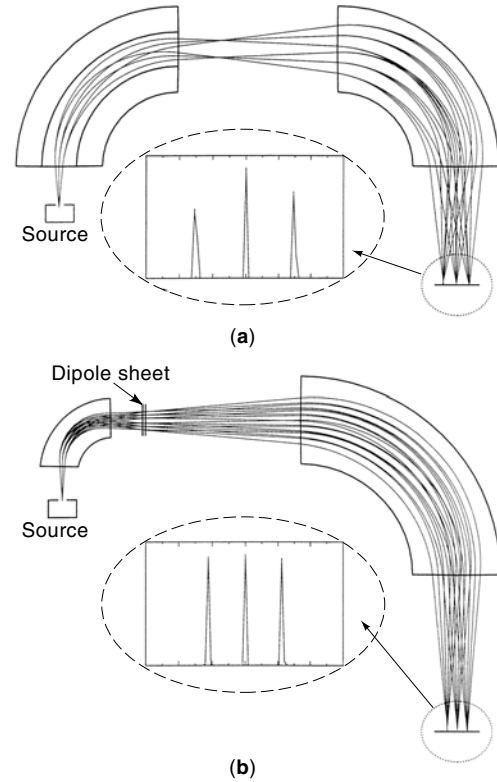


Figure 5. Ion trajectories are shown in two types of angle and energy focusing mass spectrometers which in the example shown both use the same geometry for the final magnetic sector-field mass analyzer. In Fig. 5(a) this magnetic sector field is preceded by (9,10) an electrostatic sector field and in Fig. 5(b) by a magnetic sector field placed at a different potential (10). It is assumed here that the ions are accelerated in a dipole sheet between the two stages. Note that in both systems 18 ion trajectories are shown characterized by two energies and three masses which leave the ion source at three different angles of inclination. For both systems there are also only three beams at the end (characterized by the three ion masses) independent of the angles at which the ions left the ion source and independent of the energy of these ions.

Comparing the two solutions, one sees that the overall mass dispersion is a little larger for the first case than for the second. The mass cross-contamination, however, is lower for the second solution (11) because the two momentum analyzers are used in series whereas only one momentum analyzer exists for the first solution (10).

Longitudinally Dispersive Mass Analyzers

Laterally dispersive mass analyzers have proven to be effective and powerful tools exhibiting the lowest mass cross-contamination of all known mass analyzers. However, longitudinally dispersive systems are becoming more and more popular because they are mechanically simpler. There are three types of such systems:

1. high-frequency mass analyzers (12,13,14) in which low mass ions pick up higher speeds than high mass ions within one frequency cycle. Thus, low mass ions swing in this field with larger amplitudes that can become larger than the electrode separation.

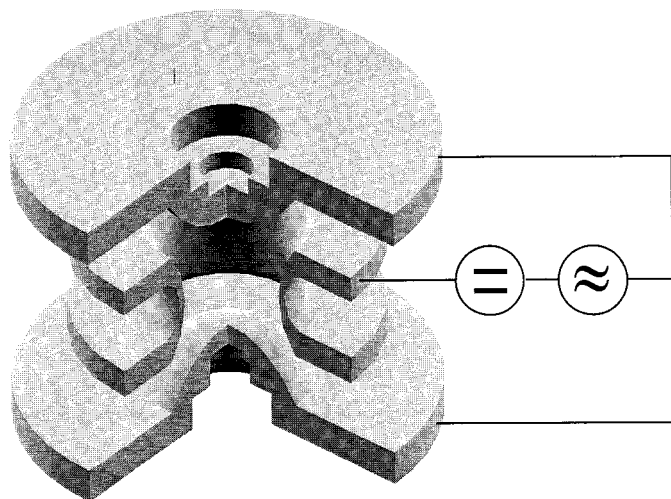


Figure 6. In the ion trap shown the ions can swing in the rotationally symmetrical quadrupolar ac field. If the dc power supply is left off, only ions above a certain mass move along stable orbits. The addition of a dc power supply also provides a limit for the high mass ions. In all cases the ion motion is mainly up and down in the figure shown here.

2. mass spectrometers in which ions circulate in a homogeneous magnetic field (15–17) or in race tracks (3,9).
3. time-of-flight mass analyzers (18–23) in which the mass-dependent flight times are observed directly for ions that have started simultaneously.

High-Frequency Mass Analyzers. High-frequency mass analyzers cause ions of energies of a few electron volts to swing in electric ac fields of some 100 V/mm and frequencies of $\nu \approx 1$ MHz. Because low mass ions swing with larger amplitudes than high mass ions, ions of too low mass are intercepted by the electrodes and only high mass ions survive. Adding a dc potential to the electrodes (see Figs. 6 and 7) also eliminates the heaviest ions (12,13) so that only ions within a small range of masses move along stable trajectories. There are two basic configurations of such mass analyzers.

The Quadrupole Ion Trap. The *quadrupole ion trap* uses rotationally symmetrical, hyperbolically shaped electrodes (see Fig. 6). Into such a *trap* ≈ 10 eV ions are introduced which then under the action of the high-frequency field mainly swing up and down between the two electrodes of the two-surface hyperboloids of Fig. 6. If the ac voltage is increased over time, the amplitudes of these swings also increase, and ions of higher and higher masses impinge on the electrodes or if appropriate holes are provided (14), leave the trap and can be recorded on some sensitive ion detector. By varying the voltages on these electrodes appropriately over time or during different time intervals, ions over a wide mass range or over a deliberately narrowed one can be made to move along stable orbits. One very useful MS/MS sequence requires that during the time interval Δt_1 the voltages are chosen so that stable ion motion is guaranteed only for ions of one particular molecule mass m_0 , whereas during the time interval Δt_2 the voltages are chosen so that ions of a range of masses all perform stable motions. Though, at the beginning of this time interval Δt_2 , only ions of mass m_0 are in the trap, these ions fragment because of collisions with residual gas atoms. With properly chosen ac and dc voltages,

these fragments are all stored in the trap. Then during the time interval Δt_3 , the voltages are scanned so that the mass range for the ions that perform stable motion is constantly decreased. Thus a mass spectrum is recorded of the fragment ions which become unstable successively according to their mass values.

The Quadrupole Mass Filter. A second well-established high-frequency mass analyzer consists of four rod-like electrodes (see Fig. 7) with ≈ 10 eV ions injected along the quadrupole axis. For appropriate ac and AD voltages of several 100 V, only ions within a very small mass range pass this *mass filter* (12,13), that is, move in the z -direction about 0.1 mm in each high-frequency period and at the same time swing transversally between the electrodes with amplitudes of about 5 mm. At the end, a mass spectrum is recorded in an ion detector if the ac and dc amplitudes are scanned appropriately over time.

The Magnetic Ion Trap Mass Spectrometer. A mass spectrometer of very high mass-resolving power $R = m/\Delta m$ is a system in which ions move in a magnetic flux density B along radii ρ according to Eq. (2) with a velocity v according to Eq. (1). Thus the flight time per turn in microseconds is given by

$$\bar{t} \approx 2\pi \frac{\rho}{v} \approx 0.0651212 \frac{m}{B} \quad (4)$$

where B is in Tesla and m is in mass units. This flight time is independent of the ion energy K . There are several ways to determine this flight time per turn.

1. If the radius ρ is large enough, one can determine the time $\bar{t}$ directly for one turn by small pulsed beam deflectors (15). In principle, this method is also applicable for many turns (9).

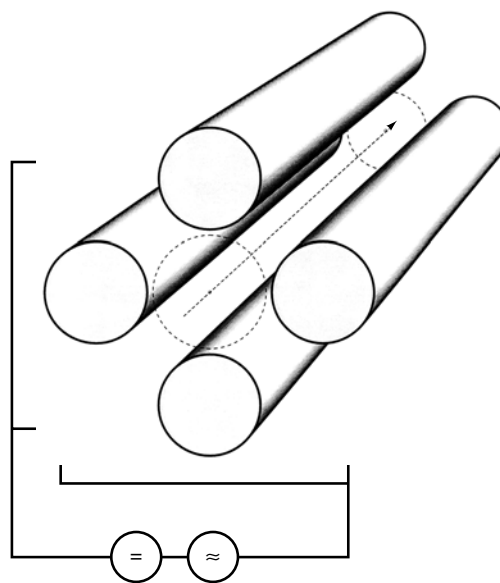


Figure 7. The quadrupole mass filter. Under the action of ac and dc fields, low energy ions swing laterally between the electrodes shown, and they move with constant velocity in the direction of the z -axis. Analogously to the ion motion in an ion trap, here also a pure ac field would allow only ions above a certain mass to move along stable orbits, whereas an added dc field would make the trajectories of the heavier ones unstable.

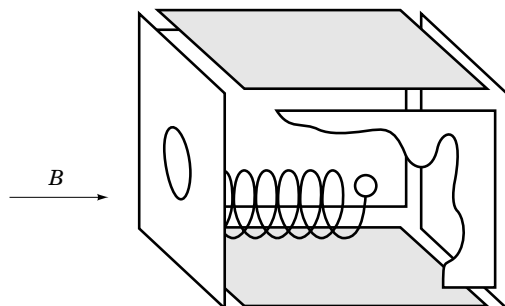


Figure 8. A Fourier transform mass analyzer that uses a high field magnetic solenoid (14). The mass-resolving power increases with an increased magnetic flux density B .

2. One can amplify the potentials induced on electrodes close to the ion path (see Fig. 8). After a Fourier analysis these induced voltages reveal mass specific frequencies $\nu = 1/\bar{t}t$ and thus the desired mass spectrum (3,16).
3. If the radius ρ is small enough, one can superimpose a high frequency electrostatic field to the magnetic one and register the finally left ions in the system by accelerating them axially out of the magnetic field. In this case the ions' azimuthal velocities are transformed into axial velocities that can be measured by a time-of-flight technique (17) which identifies the ions in resonance.

All these systems deliver rather high mass resolving powers of $m/\Delta m \geq 100,000$ or more which can be used for molecule mass analysis (16) or for the determination of nuclear mass defects (24) and other basic information like the CPT invariance (25). Note here that the mass resolving power of such systems is proportional to the number of rotations of the ion cloud and that the mass resolving power per turn can be rather small (16,17). Thus for a given experimental time it is advantageous to increase the magnetic flux density as much as possible. Note also that for $B = 7$ T, singly charged 100 eV ions of 100 u would move along circles of radius $\rho \approx 2.057$ mm with $\bar{t}t \approx 0.930$ μ s.

Instead of having ions move along circles in a homogeneous magnetic field, one can also arrange a number of magnetic and/or electrostatic sector fields into a *race track* with intermediate field-free regions (3,9,18) and then determine the flight time $\bar{t}t$ per turn via Fourier transform techniques. In principle this flight time $\bar{t}t$ depends on the ion mass and also on the ion energy. However, there are still two ways to achieve high mass-resolving power:

1. reduce the energy spread in the ion beam either by an electron cooler (3) for high energy ions or by gas collisions (26) for low energy ions.
2. introduce magnetic and/or electrostatic quadrupole lenses into the ring and excite them so that the overall system is energetically isochronous (9,18), that is, faster ions are sent around the ring along properly elongated trajectories.

Both race track systems have been used successfully for mass measurements of stable (14) and of short-lived nuclei (4).

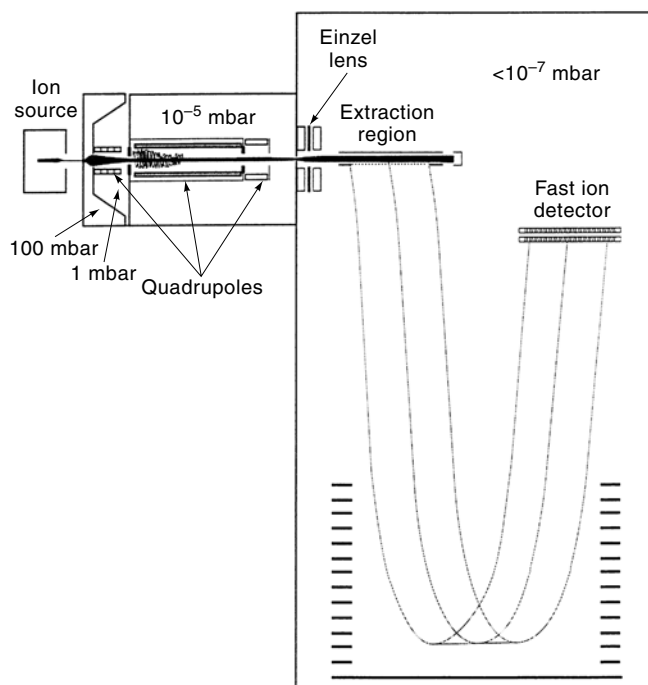


Figure 9. A time-of-flight mass analyzer for heavy molecules. Note the folded flight path in the z -direction and the large ion mirror used (23). Also note the orthogonal ion injection in the x -direction and the pulsed ion acceleration in the z -direction (21). The electrospray ion source (31) forms charged droplets and sends them into the vacuum system where in flight they evaporate all the solvent, so that highly charged ions remain. Then these ions are passed through three quadrupoles the last of which is powered in a dc mode and focuses the 10 eV ions into the extraction region. The other two are two gas-filled ac only quadrupoles the first of which *heats* the ions to boil off any molecule adducts or to fragment the main molecule (6), and the second of which because of its different gas pressure *cools* the ions by gas collisions. This procedure works continuously so that finally the ions all have more or less the same velocity in the x -direction before being pulse accelerated orthogonally in the z -direction.

Time-of-Flight Mass Analyzers

A different approach is a time-of-flight (TOF) mass spectrometer (MS) into which a bunch of ions of different masses is injected and the ion arrival is recorded at some downstream ion detector (see Fig. 9). Differing from all scanning mass analyzers, there is no limit in a TOF-MS on the mass values of the ions under investigation. Therefore TOF-MS systems are powerful tools for the investigation of large biomolecules or cluster ions.

The ion source for a TOF system requires that all ions of equal mass start from some point at the same time. This is guaranteed by

1. a pulsed ion acceleration in which a pulsed electric field acts on some cloud of ions. If all of these ions are stationary initially and then are accelerated in the z -direction, they all reach a properly placed ion detector simultaneously if their masses are the same. In other words, *they will be bunched* because the ions that start from $z_0 + \Delta z$ receive a little less energy than those that start from

$z_0 - \Delta z$, but the first ones must also travel a little further (18,19). There are three ways to implement this method:

- (a) introduce ions at a specific z_0 , and wait until they move apart to different z -values because of their thermal energies (19). In this case the final ion positions z are correlated to their initial energies which improves the bunching properties considerably.
- (b) store (20) the ions in the potential well caused by a beam of electrons of energy K_0 . Depending on the electron beam current I_e , the potential in the middle of the beam is $V_e \approx 15200(I/\sqrt{U_0})$. For $U_0 = 70$ eV and $I = 0.0005$ A, one thus finds that $V_e \approx 0.75$ V. In this case the final bunch length is usually determined by the "turnaround time," that is the time in which an ion that moved initially in the $-z$ -direction has reversed its velocity.
- (c) introduce a low energy ion beam perpendicularly to the z -direction, that is, the direction of the pulsed ion acceleration (21). In this arrangement some "ion storage" is also achieved because it takes a relatively long time for the ions to perpendicularly traverse the region from which they are then accelerated. Most importantly, however, there is almost no ion motion in the z -direction initially. Thus the "turnaround time" is reduced compared to 1(b).

For such a pulsed ion acceleration, the final energy distribution in the bunched ion beam is large, for instance, $\Delta K \approx 100$ eV.

2. a pulsed ion generation which is, for instance, the case for the laser ionization of a sample. A very good method is (20) to embed some organic material into a matrix, for instance one of glycerol, so that each molecule to be investigated is surrounded only by matrix molecules. Irradiating this sample by a short intense laser pulse releases such molecules as ions. This method is called *matrix-assisted laser desorption and ionization* (MALDI). The performance of this technique is improved by accelerating these ions in an electrostatic field that is switched on shortly after the laser pulse (19).

The time focus (19,21) achieved by pulsed ion acceleration is a good technique for achieving reasonable mass-resolving powers for a so-called *linear TOF-MS*. One can improve the simultaneous arrival of ions of equal mass by using this time focus as the source for a folded-flight-path TOF-MS (18,20). In such a system the ions are reflected by some *electrostatic ion mirror* in an energy isochronous manner, that is, so that the more energetic and thus faster ions reach the final ion detector only via a properly dimensioned detour (18,23) achieved by their deeper penetration into the reflector field. One can also say: *to first order the overall flight time becomes independent of the ions' energy spread ΔK* . Investigating q -times charged ions of energy $(K_0 \pm \Delta K)$ that enter a homogeneous repeller field E this condition of isochronicity postulates that

$$L_1 + L_2 = \frac{4qE}{K_0} \quad (5)$$

where L_1 and L_2 denote the lengths of the field-free regions between the ion source and the repeller field and between the

repeller field and the final ion detector, respectively. In this case the overall flight time is $t + \Delta t$ with $\Delta t/t = (\Delta K/K_0)^2/8 + \dots$. Using an ion reflector composed of two regions (23) of properly dimensioned field strengths or a grid-free ion reflector (18,20) that produces a properly dimensioned field region, the performance of such a TOF-MS can even be improved, so that finally only much smaller effects proportional to $(\Delta K/K_0)^3$ remain.

ION SOURCES

Many different ion sources have been developed for different mass spectrometers and different applications. Only in a few cases this ion formation can occur inside a mass analyzer, as for instance inside of an "ion trap." In most cases an external ion source is used that consists of two parts: the ionization device and the ion accelerating and beam-forming device.

Electron Impact Ion Sources

Energetic electrons ionize vaporized atoms or molecules by collisional impact. The electron energy here must be large enough that at least one electron in the shell of the atom or molecule in question is removed. Optimally, (2,27) beams of electrons of 70 eV or 80 eV are used with currents of usually a few 100 μ A. In many cases these beams are held together by magnetic fields of perhaps 0.01 T produced easily by permanent magnets. This magnetic field usually confines the electron beam to diameters ≤ 2 mm and at the same time elongates the electron path thus increasing the overall ionization probability for a given electron current. Usually the ions are pulled from the ionization region through some narrow orifice by an accelerating electrode at a potential of some 100 V or even 1000 V. However, the ions often are also pushed toward this orifice by some relatively large pusher electrode at a potential of a few volts.

Plasma Ion Sources

In a plasma ions already exist and only need to be extracted. However, electrostatic fields cannot penetrate deeply into a conductive plasma and thus ions can be extracted only from the plasma surface which consequently is depleted of ions to some depth after a very short time. Thus a pulsed ion acceleration that extracts ions from a replenished *plasma surface* is especially effective (28). As one should expect, the extracting field strength shapes the plasma surface and thus greatly influences the ion optical properties of the extracted ion beam.

Thermal Ion Sources

If the ionization energy I_p for some atom under consideration is smaller than the work function W of a substrate, the atoms leave this substrate partially as ions. The ratio between ionized and neutral atoms evaporating from a filament heated to a temperature T is given by

$$\frac{N^+}{N^0} = \exp \frac{W - I_p}{kT} \quad (6)$$

where $k = 8.62 \times 10^{-5}$ is the Boltzmann constant measured in electron volts per K. Re and also W or Ta substrates have large work functions which cause them to efficiently remove one electron from evaporating alkali atoms, that is Li, Na, K, Rb, Cs, Fr and from some earth alkali atoms during evaporation. With

Rb or Sr on 2500 K hot Re, and $I_p(\text{Rb}) = 4.18 \text{ eV}$ or $I_p(\text{Sr}) = 5.7 \text{ eV}$ one finds that $N^+/(N^0 + N^+)$ equals 100% or 85% if the work function of the Re surface is assumed to be 6.1 eV. However, other atoms, for instance, U are also ionized with some efficiency (2,27). Similarly, one can also attach one electron to atoms with large electron affinities (for instance, Br) if they are evaporated from a substrate that has a very low work function, for instance, LaB_6 .

This simple ionization technique is improved considerably if the substrate forms the inner surface of a hot cavity (29). In this case most elements are ionized with efficiencies of about 1%. The reason for this enhancement is that the ions are extracted through a small hole in this cavity whereas the atoms stay in the cavity and thus get another chance to be ionized.

Laser Ion Sources

As one might expect, one can directly ionize specific isotopes by high-power lasers of some resonant frequencies. However, often one uses only lasers of broader frequency bands and thus ionizes all isotopes of a specific element simultaneously (30). In some cases one also uses the property of a finely focused laser beam to produce a small plasma at some surface as, for instance, for the previously mentioned MALDI sources (22).

Electrospray Ion Sources

A very interesting ion source for high mass biomolecules uses the electrospray technique (31). In one of these ion sources a very dilute solution of the molecules of interest is sprayed in fine droplets into some gas at about 1 bar. Then the gas and these droplets are sucked through a small orifice into a vacuum vessel where the droplets will evaporate most of the solvent and thus rapidly decrease in size. If there were only one molecule in the droplet there is only one molecule left at the end when all the solvent has evaporated. Because the droplets were multiply charged from the beginning, as is the case in the spray from any waterfall, the final molecule is multiply charged (see Figs. 1 and 4).

CONCLUSION

Mass spectrometry began as a special technique to determine the isotopic distribution of all elements (9), but it has grown into a very general, sensitive, and specific analytical technique for ionized atoms or molecules. For this reason mass spectrometers have been used in very diverse applications from nuclear physics to pharmacology. The fastest growing application certainly is the analysis of biomolecules and the trace detection of organic molecules.

BIBLIOGRAPHY

1. F. White and G. Wood, *Mass Spectrometry: Applications in Science and Engineering*, New York: Wiley, 1981.
2. M. Gross, *Mass Spectrometry in the Biological Sciences: A Tutorial*, Dordrecht: Kluwer, 1992.
3. B. Schlitt et al., *Hyperfine Interactions*, **99**: 117, 1996.
4. H. Wollnik et al., *Nucl. Phys.*, **A616**: 346, 1997.
5. K. L. Busch, G. L. Glush, and S. A. McLuckey, *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, New York: VCH, 1988.
6. A. Dodonov et al., *Rapid Commun. Mass Spectrom.*, **11**: 1649, 1997.
7. R. C. Finkel and M. Suter, "AMS in the Earth Sciences: Techniques and Applications," in *Adv. Anal. Geochem.* **1**: 211, 1990.
8. F. Brunner, *The Science of Chromatography*, Amsterdam: Elsevier, 1985.
9. H. Wollnik, *Optics of Charge Particles*, Orlando: Academic Press, 1987.
10. F. W. Aston, *Philos. Mag.*, **38**: 709, 1919.
11. H. Wollnik, *Nucl. Instr. Methods*, in press, 1998.
12. W. Paul and H. Steinwedel, *Z. Naturf.*, **A8**: 448, 1953.
13. P. H. Dawson, *Quadrupole Mass Spectrometry and its Applications*, Amsterdam: Elsevier, 1976.
14. G. C. Stafford et al., *Int. J. Mass Spectrom. Ion Proc.*, **60**: 85, 1984.
15. A. G. Marshall and F. R. Verdun, *Fourier Transform in NMR, Optical and Mass Spectrometry*, New York: Elsevier, 1990.
16. G. Wendt, *Metrologia*, **22**: 174, 1986.
17. M. St. Simon et al., *Physica Scripta*, **T59**: 406, 1995.
18. H. Wollnik, in *Mass spectrometry in biomolecular science*, M. Caprioli, A. Malorni and G. Sindona, (eds.), Dordrecht: Kluwer, 1996, p. 111.
19. W. C. Wiley and I. H. McLaren, *Rev. Sci. Instr.*, **26**: 1150, 1955.
20. R. Grix et al., *Int. J. Mass Spectrom. Ion Proc.*, **93**: 323, 1989.
21. G. J. O'Halloran et al., Tech. Doc. ASD TDR Report 62-644, Bendix Co., 1964.
22. M. Karas, U. Bahr, and U. Giessmann, *Mass Spectrom. Rev.*, **10**: 335, 1991.
23. B. A. Mamyrin et al., *Sov. Phys. JETP*, **37**: 45, 1973.
24. D. Beck et al., *Nucl. Phys. A*, **626**: 343c, 1997.
25. G. Gabrielse et al., *Phys. Rev. Lett.* **74**: 3544, 1995.
26. G. Savard et al., *Phys. Lett. A*, **158**: 247, 1991.
27. B. Wolf, *Handbook of Ion Sources*, Boca Raton: CRC Press, 1995.
28. Y. Shirakabe et al., *Nucl. Instr. Methods*, **A 337**: 11, 1993.
29. A. Latushinsky and V. Raiko, *Nucl. Instr. Methods*, **125**: 61, 1975.
30. P. v. Duppen, *Nucl. Instr. Methods*, **B 126**: 66, 1997.
31. J. B. Fenn et al., *Science*, **246**: 64, 1989.

HERMANN WOLLNIK
Universität Giessen

MICROSCOPY

ELECTRON MICROSCOPES

In 1873, it was proven by Ernst Abbe that the resolving power of a light microscope will always be limited by the wavelength of the light, which is of the order of $1\text{ }\mu\text{m}$, so that there could be no hope to visualize much smaller objects such as atomic scale structures. (In the 1980s, near-field optical scanning techniques were developed that can bring the resolution down by two orders of magnitude.) Fifty years later, a new impulse was given to the problem by the hypothesis of Louis De Broglie about the wave nature of particles so that other particles could also serve as "light." In 1931 Ernst Ruska developed the first transmission microscope TEM that uses electrons instead of photons. In 1986 Ernst Ruska was awarded the Nobel Prize for his pioneering work. Electrons are the best candidates since they can easily be generated by a heated filament or extracted from a point by an electric field and they are easily deflected by electric and magnetic fields. When accelerated to, say, 100 keV, their wavelength is much smaller ($3\text{ pm} = 3 \times 10^{-12}\text{ m}$) than that of visible light. They can also be detected on a photoplate, a fluorescent screen, or an electronic camera. On the other hand, they can only propagate in vacuum and they can only penetrate through very thin objects ($<10^3\text{ nm}$), so that vacuum and specimen preparation techniques are crucial.

In past decades, electron microscopy has matured to an indispensable tool for materials and biomedical research. Numerous are the Nobel Prizes for research work in which electron microscopy revealed crucial evidence. Despite the very small wavelength of the electrons, it has only recently been possible to visualize individual atoms. The reason for this is that magnetic lenses inevitably suffer from aberrations that, contrary to glass lenses in a light microscope, cannot easily be corrected. Recently the problem of compensating the spherical aberration has become solved on the research level.

When individual atoms, the building blocks of nature, can be revealed, electron microscopy enters into a new era. Indeed, then it becomes possible to determine the atomic structure of matter quantitatively and accurately, even for objects for which only limited prior knowledge is available. Although the goal is not yet achieved, the progress is very encouraging. In the future, as materials science evolves into materials design, and mesostructures into nanostructures, the role of electron microscopy will become even more important.

In recent years we have also seen the development of related techniques such as scanning transmission electron microscopy (STEM), electron holography, ptychography, and add-on techniques that use complementary information such as X-ray analysis (EDX), and energy filtering. There is now a growing tendency to incorporate all these techniques into one versatile instrument under full computer control, where they are all considered as different ways to obtain complementary information, mainly of chemical nature.

The development of scanning electron microscopy (SEM) has taken place parallel to that of TEM. The first SEM was constructed by Manfred von Ardenne in the 1930s. The SEM is conceptually simpler than the TEM. With the aid of magnetic lenses, the electron beam is focused into a small point that is scanned over the surface of the object. The backscattered

electrons, or other secondary signals, can then be detected and synchronously displayed on a cathode-ray-tube (CRT) screen. The SEM is relatively easy to use and does not require extensive specimen preparation and is also less expensive than the TEM. On the other hand, it can only be used to study the surface area and, because of the beam spread in the object, its resolution is rather limited and quantitative interpretation of the images is more difficult.

TRANSMISSION ELECTRON MICROSCOPY

The Instrument

A transmission electron microscope is very similar to a light microscope (Fig. 1) in which the light bulb is replaced by an electron source and the glass lenses by magnetic (or electric) lenses. The focusing of the lenses can be varied by changing the lens currents so that the setting of the microscope (focusing, magnification, etc.) can easily be altered. The condenser lens shapes the electron flow into a parallel beam. The specimen is a very thin object mounted in a holder that, through an airlock system, can be brought into the electron beam. Holders exist with a variety of degrees of freedom (translation, rotation,

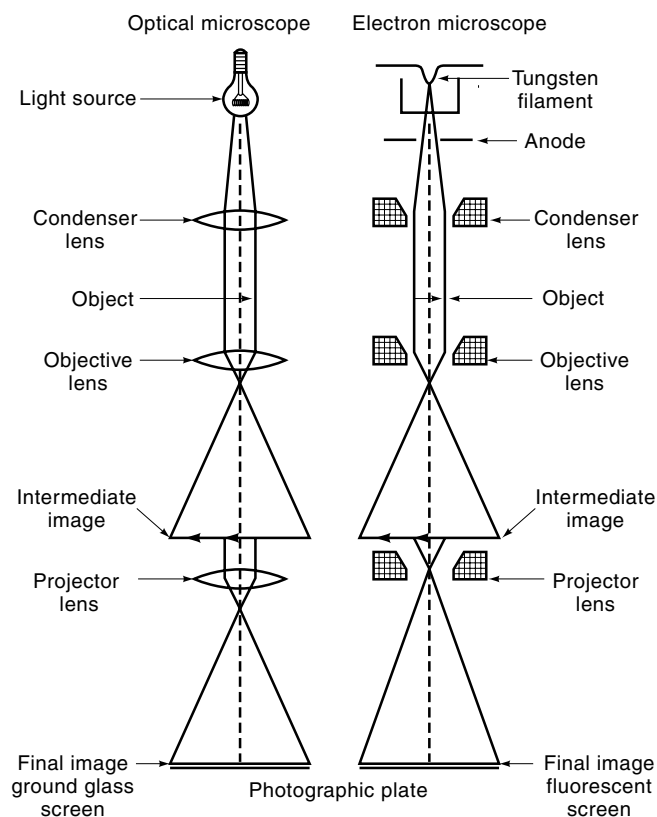


Figure 1. Comparison between an electron microscope (right) and a light microscope (left). All the components of the light microscope have their analog in the electron microscope. The acceleration electrons have a wave character and act as light.

tilt, temperature, stress). The objective lens produces an enlarged and rotated image, which in its turn, through a system of intermediate and projector lenses, is further magnified and projected onto a fluorescent screen, a photoplate, or an electronic camera.

Another mode of operation is not to image the object but to image the focal plane of the objective lens. In this way one can observe the electron-diffraction pattern directly onto the fluorescent screen.

The resolving power of the electron microscope is mainly determined by the quality of the objective lens. With modern instruments, a resolution of about 0.15 nm can be reached.

In modern electron microscopes the microscope is operated under computer control and the alignment can be done automatically. The visual electron source, which is a thermal filament, is increasingly replaced by a field-emission source, from which the electrons are extracted by an electric field and which yields a much higher brightness.

Image Formation

Intuitive Description. The imaging process in the electron microscope can be sketched in a very simple way as follows (Fig. 2). The object is illuminated by a plane wave. Due to the interaction of the electrons with the atoms of the object, the electron wave is modulated by the structure of the object. By propagating this object exit wave through the electron microscope to the image plane this electron wave is blurred, that is, each point of the wave is spread into the so-called point-spread function or impulse response function. The wave as well as the point-spread function is a complex valued function with an amplitude and a phase component. Finally, however, in the image plane only the intensity is recorded so that the phase information is lost. Full recording of this phase information can be done using holographic techniques (discussed later).

Wave Optical Description. As in the light microscope, the formation of the image must be described by wave optics (Fig. 3).

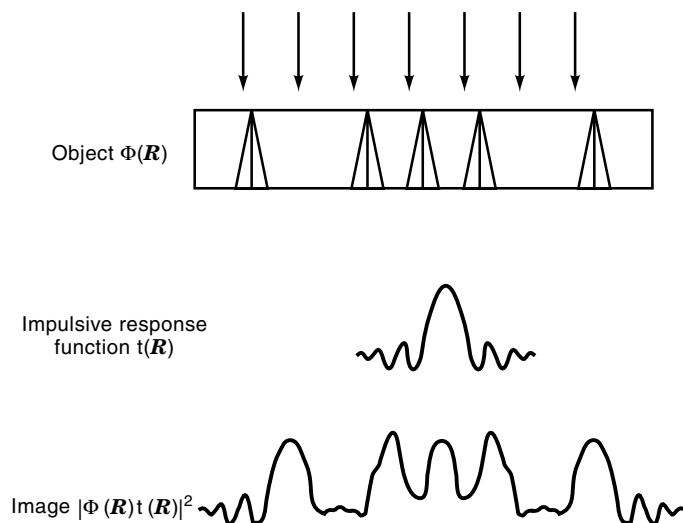


Figure 2. Schematic representation of the imaging process. The electron wave is modulated by the object and then blurred by the electron microscope.

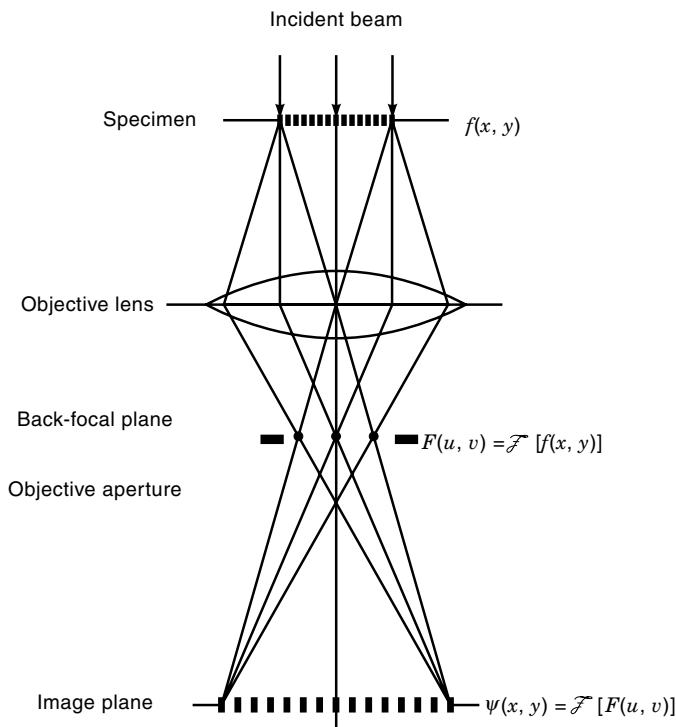


Figure 3. Wave optical representation of the image formation by the objective lens in a transmission electron microscope. The corresponding mathematical operations are indicated [see text with $\mathbf{R} = (x, y)$ and $\mathbf{G} = (u, v)$].

The electron wave is diffracted by the object, and each diffracted beam is focused by the objective lens into one point in the focal plane. The electron wave in the focal plane can be considered as the Fourier transform of the exit wave of the object. The high spatial frequencies (small details) correspond to beams far from the optical axis and the small spatial frequencies (large details) are located close to the optical axis. In case the object is periodical, such as is the case of a perfect crystal, the Fourier transform is discrete, that is, the diffraction pattern in the back focal plane consists of discrete spots. These are the different diffracted beams. In case the object is aperiodic, the image in the focal plane shows diffuse intensity. In the second step of the imaging process, the points in the focal plane act again as sources of waves, which interfere in the image plane. The image wave found in the image plane is the reverse Fourier transform of the wave found in the focal plane; therefore, under ideal imaging conditions with no imaging aberrations or apertures, the image wave should be identical to the object exit wave. In short, the whole imaging process is wave optically considered as a sequence of two Fourier transforms, the first one leading from the object plane to the back focal plane of the objective lens, and the second one from the focal plane to the image plane. In practice, however, the objective lens has intrinsic aberrations. The most important ones are the spherical aberration and the defocus. As a consequence, the Fourier components of the electron wave in Fourier space suffer from hampering phase shifts arising from these aberrations that are more important for the high spatial frequencies and that blur the image.

Usually the Fourier components, falsified by more than a tolerable threshold, are masked out to minimize the blurring

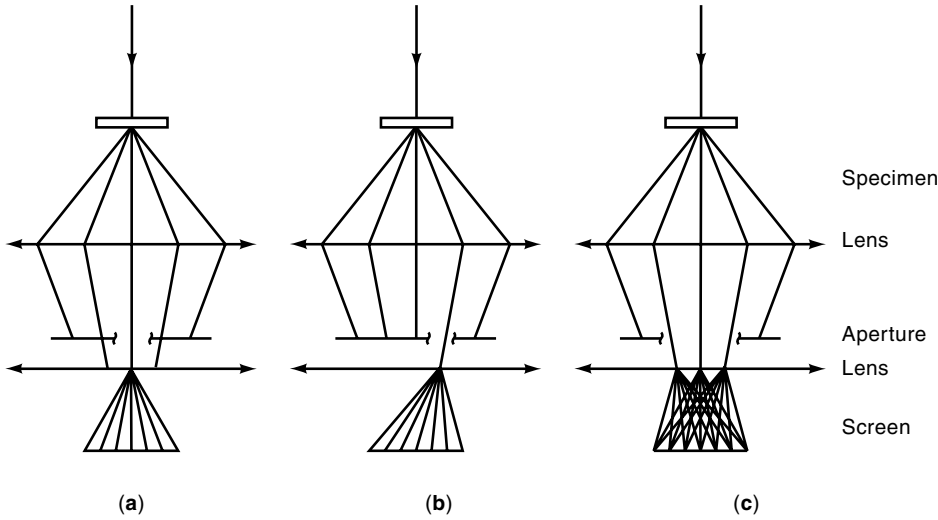


Figure 4. Different imaging modes. (a) Bright-field imaging: only the transmitted beam contributes to the image. (b) Dark field: only one diffracted beam contributes to the image. (c) High-resolution imaging: the image is formed by interference of many diffracted beams.

effect. On the other hand, one has to find a compromise because, for high-resolution imaging, as many Fourier components as possible should be admitted to the image. In a sense, one can then consider the image wave as a truncated Fourier reconstruction of the object exit wave with only a limited number of Fourier components, which may be falsified by the aberrative phase shift. The image formed by interference of the residual waves cannot be interpreted easily.

The resolution, in a sense the smallest detail present in the image, is then determined by the highest spatial frequency that is allowed to contribute to the image formation. One distinguishes between the point resolution, which is the highest spatial frequency that contributes without extra phase shift and which gives rise to directly interpretable details, and the information limit, which is the highest spatial frequency in the image formation but with a phase shift that complicates a direct interpretation of the details.

For special investigations of crystalline objects, it is useful to include only one diffracted beam for imaging (Fig. 4). If the central beam is used, one obtains a so-called bright field image (BF); if only a diffracted beam is included, a so-called dark field image results. In these cases, since there is no interference between diffracted beams, the image contrast can more easily be interpreted in terms of the number of electrons scattered in the corresponding particular direction. However, high resolution cannot be obtained with these techniques.

In any case, for the interpretation of the images, it is equally important to understand the interaction between the electrons and the object.

Mathematical Formulation

Transfer Function. By interaction with the object, the object exit wave $\psi(\mathbf{R})$ is modulated both in amplitude and phase. Hence $\psi(\mathbf{R})$ is complex. $\mathbf{R}$ is the position vector in the object exit plane. According to Fraunhofer's diffraction theory the diffracted wave in the direction given by the reciprocal vector $\mathbf{g}$ (or spatial frequency) is given by the Fourier transform of the object function, that is,

$$\phi(\mathbf{g}) = F_{\mathbf{g}}\psi(\mathbf{R}) \quad (1)$$

The wave in the back focal plane of the objective lens is then the Fourier transform of the object wave. If one directly images the focal plane, one can see the diffraction pattern, given by $|\phi(\mathbf{g})|^2$. If the object is periodic, such as a crystal, the diffraction pattern will consist of sharp spots. A continuous object will give rise to a continuous diffraction pattern. The second stage of the imaging process is described by an inverse Fourier transform that reconstructs the object function $\psi(\mathbf{R})$ (usually enlarged) in the image plane (Fig. 3). The intensity in the image plane is then given by $|\psi(\mathbf{R})|^2$.

In practice, by inserting an aperture in the focal plane of the objective lens, it is possible to obtain an image in which only selected beams contribute. On passing through the objective lens, each electron beam $\mathbf{g}$ undergoes a phase shift and an amplitude reduction (damping). Hence the transfer function takes the form

$$T(\mathbf{g}) = A(\mathbf{g}) \exp[-i\chi(\mathbf{g})]D(\mathbf{g}) \quad (2)$$

$A(\mathbf{g})$ describes the effect of the beam-selecting aperture and the damping caused by incoherent effects (such as fluctuations in voltage and lens currents). $\chi(\mathbf{g})$ is the phase shift, given by

$$\chi(\mathbf{g}) = \frac{\pi}{2}(C_s\lambda^3g^4 + 2\epsilon\lambda g^2) \quad (3)$$

with C_s the spherical aberration constant of the objective lens, ϵ the focus distance, and λ the electron wavelength. The wave function at the image plane is now given by the inverse Fourier transform

$$\psi_{\text{im}}(\mathbf{R}) = F_{\mathbf{R}}^{-1}T(\mathbf{g})\phi(\mathbf{g}) \quad (4)$$

and the image intensity by

$$I_{\text{im}}(\mathbf{R}) = \|\psi_{\text{im}}(\mathbf{R})\|^2 \quad (5)$$

Equation (4) is called the coherent approximation; it is valid for thin objects. In general, the expressions are more complicated. One then has to consider the image as an incoherent superposition of images that are slightly different, due to fluctuations

in the conditions of the electron microscope. For a more general description we refer the reader to Ref. 1.

The total image formation process is illustrated in Fig. 2.

Impulse Response Function. As intuitively described earlier the image transfer can also be described in real space as a blurring effect, and as also follows from Eq. (4) using the convolution theorem,

$$\Phi(\mathbf{R}) = \psi(\mathbf{R}) * t(\mathbf{R}) \quad (6)$$

where $\psi(\mathbf{R})$ is the object wave in real space and $t(\mathbf{R})$ is the Fourier transform of the transfer function. For an hypothetical ideal pointlike object, the object wave $\psi_{\text{im}}(\mathbf{R})$ would be a δ function or “impulse” [$\psi(\mathbf{R}) = \delta(\mathbf{R})$] so that $\Phi_{\text{im}}(\mathbf{R}) = t(\mathbf{R})$, that is, the microscope would reveal $t(\mathbf{R})$, which therefore is called impulse response function. If the transfer function would be constant (i.e., perfectly flat) in the whole spatial frequency range, the impulse response function would be a δ function so that $\psi_{\text{im}}(\mathbf{R}) = \psi(\mathbf{R})$, that is, the wave function in the image plane represents exactly the wave function of the object. In a sense the microscope is perfect. However, in practice the transfer function cannot be made constant as is shown in Fig. 5. The impulse response function is still peaked as shown in Fig. 6.

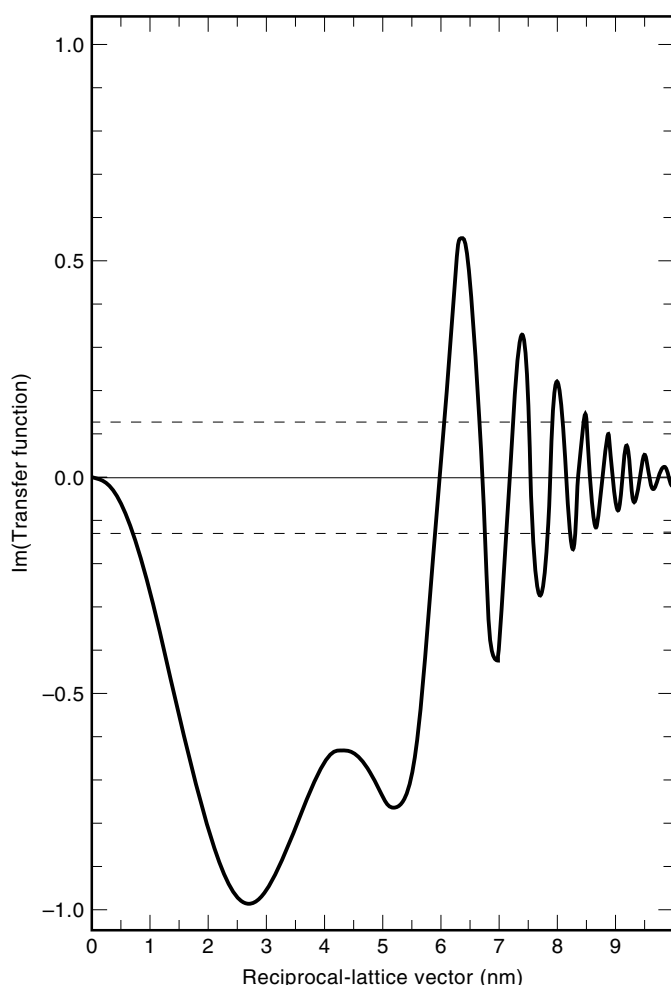


Figure 5. Transfer function (imaginary part) of a 300 keV electron microscope at optimum defocus.

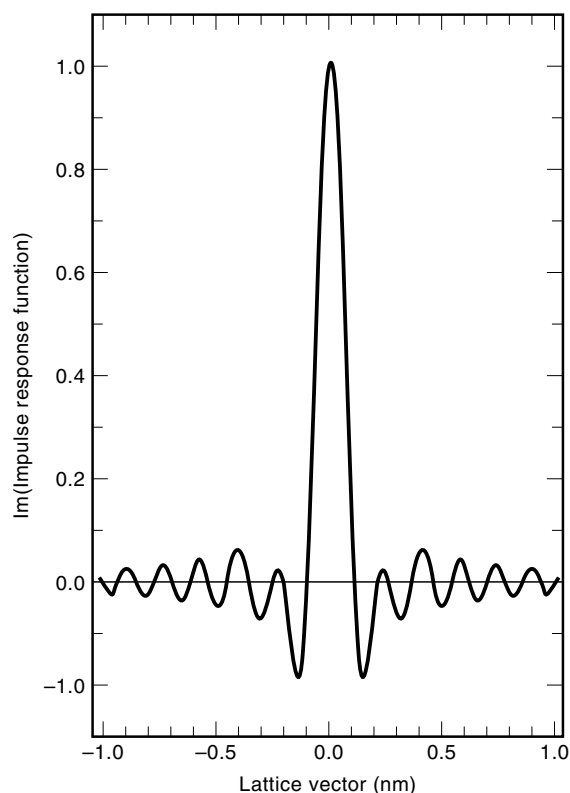


Figure 6. Impulse response function (imaginary part) corresponding to the transfer function of Fig. 5.

Hence, as follows from Eq. (6), the object wave $\psi(\mathbf{R})$ is smeared out (blurred) over the width of the peak. This width can then be considered as a measure for the resolution in the sense as originally defined by Rayleigh. The width of this peak is the inverse of the width of the constant plateau of the transfer function in Fig. 5. From another point of view one can argue that if all the spatial frequencies have the same phase shift, the information is transferred forward and keeps a point to point relation to the object.

However, the information beyond this plateau is still contributing to the image but with a wrong phase. It is scattered outside the peak of the impulse response function and it is thus redistributed over a larger area in the image plane.

Phase-Contrast Microscopy. In an ideal microscope, the image wave function would exactly represent the object wave function and the image of a pure phase object function would show no contrast. This can be compared with imaging a glass plate with variable thickness in an ideal light microscope.

If the phases of the diffracted beam (Fourier components) are shifted over $\pi/2$ with respect to the central beam the image contrast would directly reveal the phase of the object. In light microscopy, phase contrast can be obtained by inserting an annular quarter-wavelength plate in the focal plane of the objective lens.

In electron microscopy, phase-contrast imaging can be achieved by making the transfer function as constant as possible. From Eq. (3) it is clear that phase shifts occur due to spherical aberration and defocus. However, the effect of spherical aberration, which, in a sense, makes the objective lens too strong for the most inclined beams, can be compensated somewhat by slightly underfocussing the lens.

The focus setting is called the optimum focus or Scherzer focus and its value can be calculated as

$$\epsilon = -1.2 C_s^{1/2} \lambda^{1/2} \quad (7)$$

The transfer function for this situation is depicted in Fig. 5. The phase shift $\chi(\mathbf{g})$ is nearly equal to $-\pi/2$ with respect to the central beam for a large range of spatial coordinates, up to the value

$$|\mathbf{g}| = 1.5 C_s^{-1/4} \lambda^{-3/4} \quad (8)$$

The image then reveals directly the phase of the object. Now a thin material object acts as a phase object in which the phase is proportional to the electrostatic potential of the atoms projected along the viewing direction. Hence, if the object would be very thin, optimum focus imaging would directly reveal atom columns as dark dots and empty spaces as light areas. However, this argument only holds for spatial frequencies that are within the range given by Eq. (8). Furthermore the thickness up to which an object can be considered as a weak phase object is very small (e.g., 1 nm) and is rarely met in practice.

Resolution. One has to distinguish between point resolution (or structural resolution) as the finest detail that can be interpreted in terms of the structure and the information limit, which is the finest detail that can be resolved by the instrument, irrespective of a possible interpretation.

The point resolution can be obtained from the inverse of the maximal spatial frequency equation (8) as

$$\rho_s = 0.65 C_s^{1/4} \lambda^{3/4} \quad (9)$$

The point resolution is also equal to the “width” of the impulse response function (Fig. 6). The information beyond ρ_s is transferred with a nonconstant phase and, as a consequence, is redistributed over a larger image area.

The information limit ρ_I can be defined as the finest detail that can be resolved by the instrument. This limit is mainly determined by incoherent effects (fluctuations on the energy of the electrons and on the lens currents, divergences of the illuminating beams). The information limit is usually smaller than the point resolution. Typical values are $\lambda = 2$ pm (300 keV), $C_s = 1$ mm, $\rho_s = 0.2$ nm, $\rho_I = 0.13$ nm. The point resolution can be improved by reducing the incoherent effects (e.g., energy and current fluctuations), for instance, by using a field emission source and a stable voltage supply.

The Specimen

Specimens for high-resolution electron microscopy are prepared using the same techniques as for usual transmission electron microscopy, that is, ion beam milling, chemical and electrolytical thinning, cleavage, crushing, and so on. The only requirements are that the specimen should be sufficiently thin, that is, less than about 10 nm. Furthermore, the specimen should be sufficiently clean and free of contamination. Depending on the type of specimen one can use different thinning methods such as ion beam milling, electrochemical etching, crushing, and dimpling.

For details of specimen preparation we refer the reader to Ref. 2.

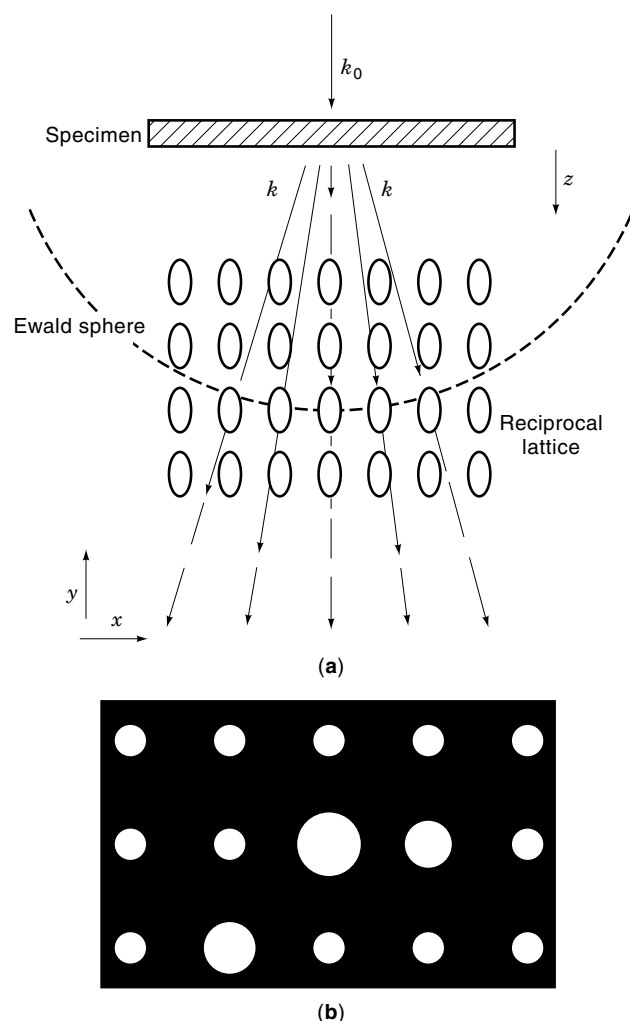


Figure 7. Formation of the diffraction pattern. The simultaneously excited electron beams can be used for image formation.

Crystalline specimens are oriented with a zone axis parallel to the incident beam so that all the diffracted beams of that zone are excited simultaneously and maximal information is present in the image. This is shown in Fig. 7. In this situation, the electrons propagate parallel to a zone axis, that is, parallel to the atom rows, which makes a possible interpretation of the images in terms of the projected structure meaningful. As an example, Fig. 8 shows a drawing of a model of a dislocation viewed along a zone axis.

It is also possible, using an aperture placed in the focal plane of the objective lens, to select a particular set of beams so that the images contain specific information. If the central beam is not included, these images are called dark-field images.

After finding a suitably thin part with the proper orientation, one has to adjust the focus. When the specimen is very thin, the zero focus corresponds to minimal contrast.

In practice, one takes a series of images at gradually different focus settings, recorded approximately around the optimum defocus. This is called a through-focus series. When dealing with a specimen that is unstable in the electron beam, the specimen can be completely destroyed within a period of seconds. Here a minimal exposure technique has to be used.

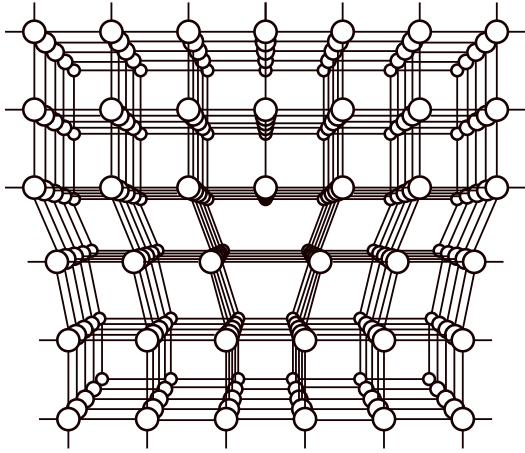


Figure 8. Structure model of a crystal containing a dislocation viewed along the atomic columns.

Electron Diffraction

Diffraction Mode. When the focal length of the intermediate lens is increased, by weakening its excitation, so as to make the back focal plane of the objective lens coincide with the object plane of the projector lens, a magnified image of the diffraction pattern is projected on the screen. Neither specimen orientation nor selected area are hereby changed; the diffraction pattern is thus representative of the selected area. It should be noted that the area selected in the diffraction mode is much larger than the field of view for imaging.

Geometry of the Diffraction Pattern. The diffraction conditions can be formulated in two different ways: either emphasizing direct space or reciprocal space. In direct space the diffraction conditions are known as *Bragg's law* (3). Attention is focused on sets of parallel lattice planes with interplanar spacing d_H , where H represents the three Miller indices of the considered family of lattice planes. "Reflection" will occur for certain "glancing" angles θ_H satisfying Bragg's law; $2d_H \sin \theta_H = n\lambda$, where λ is the wavelength of the radiation used and n is an integer. This relation expresses the condition that the difference in path length between waves scattered by two successive lattice planes is an integer number of wavelengths (Fig. 9). For 100 kV

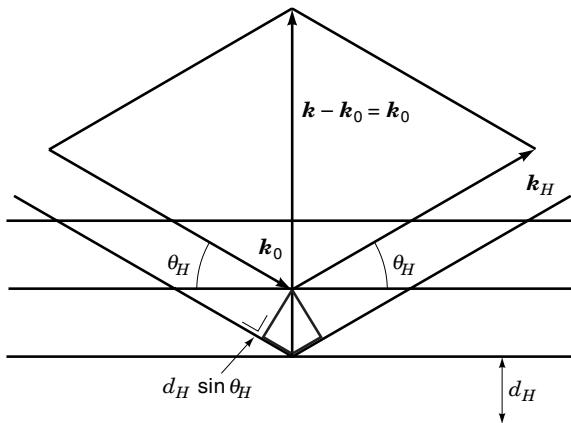


Figure 9. "Reflection" of electron waves by successive lattice planes leading to the Bragg condition.

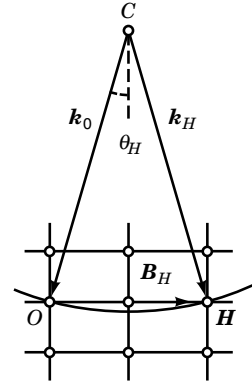


Figure 10. Ewald construction of diffracted wave. A diffracted wave is formed if Ewald's sphere ($\approx$ plane) intersects a reciprocal-lattice node H . k_0 is the wave vector of the incident wave, k_H is the wave vector of diffracted wave, and B_H is the reciprocal-lattice vector belonging to node H .

electrons $\lambda \sim 0.004$ nm and hence the Bragg angles are very small, of the order of a few degrees at most.

In terms of reciprocal space the *Ewald diffraction condition* (4) can be formulated as $k_H = k_0 + B_H$, where k_0 is the wave vector of the incident wave and k_H that of the scattered wave and B_H is a reciprocal-lattice vector $B_H = h_1 b_1 + h_2 b_2 + h_3 b_3$. The b_j are the base vectors of the reciprocal lattice, defined in terms of the base vectors a_i of the direct lattice by the relations $a_i \cdot b_j = \delta_{ij}$ (Kronecker δ) ($i, j = 1, 2, 3$). The Ewald condition expresses the conservation of energy and linear momentum on elastic scattering; its physical content is the same as that of Bragg's law. Geometrically it expresses the condition that a sphere with radius $|k| = 1/\lambda$ and center C in $-k_0$ (Ewald's sphere) must pass through a point H of the reciprocal lattice for diffraction to occur. The scattered wave direction is then obtained by joining the center of Ewald's sphere with the excited reciprocal-lattice point H (Figs. 7 and 10).

In electron microscopy the specimen has to be a thin foil since electrons are strongly "absorbed" in solids. This causes an anisotropic relaxation of the Bragg condition along the foil normal. Diffraction now occurs also for conditions deviating somewhat from the exact Bragg angles. This is represented in reciprocal space as a change of the sharp spots into infinitely thin rods, called "relrods" perpendicular to the foil surfaces. These relrods have an intensity profile represented in Fig. 11 according to the kinematical theory of electron diffraction (see the Appendix).

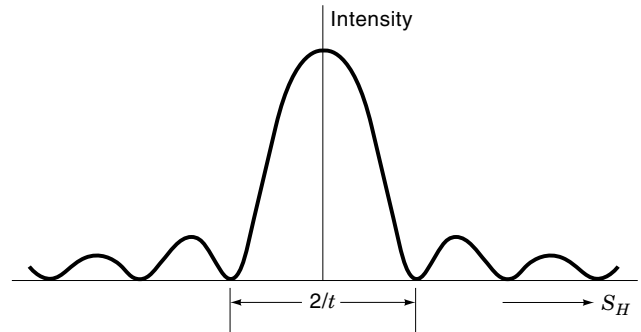


Figure 11. Intensity profile along a "relrod" for a foil with thickness $> t$.

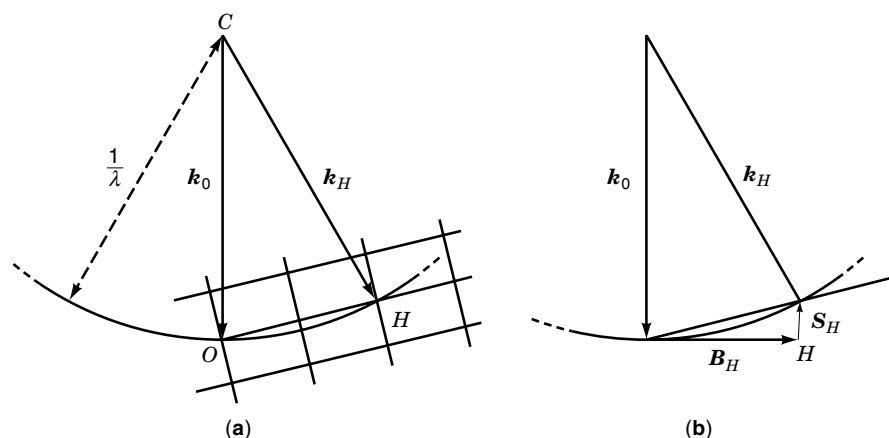


Figure 12. Illustration of the diffraction geometry in reciprocal space, defining the excitation error s_H . (a) $s_H = 0$. (b) $s_H < 0$.

Since the radius of Ewald's sphere ($1/\lambda$) is large as compared to the mesh size of the reciprocal lattice ($1/d_H$), the sphere can be approximated by a plane. The diffraction pattern can thus be considered as the projection of an undeformed planar section of the reciprocal lattice.

The distance s_H by which Ewald's sphere misses the reciprocal-lattice node H is called the *excitation error*. It is a vector parallel to the foil normal that connects the reciprocal-lattice node and the intersection point with Ewald's sphere. It is by convention called positive when pointing in the sense of the propagating electrons, that is, when the node is inside Ewald's sphere. In Fig. 12(b), s_H is negative.

Convergent Beam Electron Diffraction. By converging the incident beam into a cone, the diffraction spots become disks with a fine structure of intensity variations. This way of diffraction is highly sensitive to small changes in the crystal structure. It can be used very efficiently to determine the space group symmetry of a crystal, to measure very accurately the unit cell dimensions, thermal expansion coefficients, local strain, and crystal potentials. For more detail, we refer to Ref. 5.

Imaging

Diffraction-Contrast Imaging

Principles. Diffraction-contrast images are preferably obtained under two-beam conditions since this allows an easier interpretation. The specimen is oriented such that apart from the direct beam only one scattered beam is strong. The selector aperture allows us to select either the direct beam or the intense scattered beam. The selected diffraction spot is then highly magnified by the lenses. The image obtained in this way is a highly magnified map of the intensity distribution in the corresponding diffraction spot. If the direct beam is magnified this is called a *bright-field image* [Fig. 13(a)]; if the diffracted beam is selected, the image is a *dark-field image* [Fig. 13(b)]. Since the electrons propagate within a narrow column by diffracting back and forth, such columns can be considered as picture elements forming the image (Fig. 14). This is the so-called column approximation (6,7).

The amplitude of either the scattered or the transmitted beam emerging in a point on the exit surface of the foil is the sum of the contributions of all volume elements along a column parallel to the incident beam. This amplitude depends on the variation of the excitation error s_H along the column, which itself depends on the local orientation of the diffracting planes.

In a perfect foil s is constant and it is the same along all the columns; the image is an area of uniform intensity. If a strain pattern is present in the foil s will vary differently along different columns and as a result an image is formed. Diffraction-contrast images are thus very sensitive to strain patterns but they do not reveal the crystal structure.

Interface Contrast

Examples. Interfaces situated in planes inclined with respect to the foil surfaces divide the foil into two wedge-shaped parts. These two parts are related by a translation if the interface is a stacking fault or an antiphase boundary with $\mathbf{R} = \mathbf{R}_0$; they differ in orientation, that is, in s when the interface is a domain boundary (Fig. 15).

In both cases sets of parallel fringes are formed, which are parallel to the closest intersection line of the interface and the surface. Their intensity profiles are different, however. Planar interfaces parallel to the foil surfaces only cause a brightness difference in the faulted area.

In Fig. 16 the two types of fringes are compared. The most important feature is the nature, bright (B) or dark (D), of the edge fringes (i.e., the first and the last fringe). Bright-field

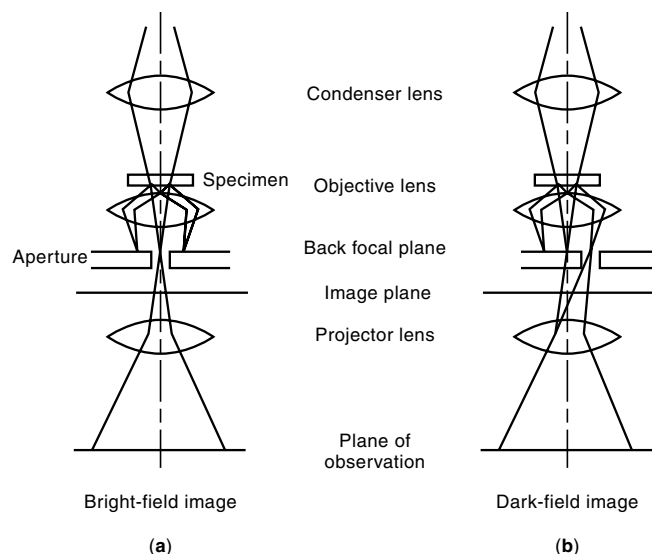


Figure 13. Ray paths for the (a) bright-field mode and (b) dark-field mode.

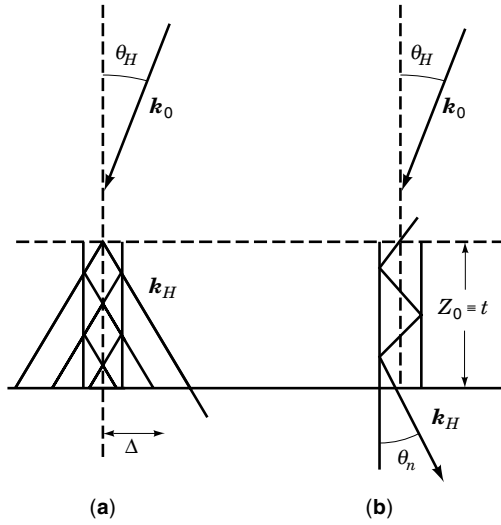


Figure 14. Repeated reflection of electron waves in a column of crystal. According to the (a) kinematical theory and (b) dynamical theory.

fringe patterns produced at stacking faults have edge fringes of the same nature, whereas in the dark-field fringe pattern the edge fringes have opposite nature. The reverse is true for domain boundary fringes; the edge fringes are of opposite nature in the bright-field image, whereas in the dark-field image the nature of the two edge fringes is the same.

From the nature of the first and last fringes one can conclude, for instance, whether a stacking fault in a face-centered cubic crystal is either *intrinsic* (i.e., of the type ...*abcababc*...) (8) or *extrinsic* (i.e., of the type ...*abcbacabc*...). Figure 16, right refers to a domain boundary whereas Fig. 19, left is due to a stacking fault.

Dislocation Contrast (9,10). The contrast produced at dislocation lines can be understood by noting that the reflecting lattice planes in the regions on two opposite sides of the dislocation line are tilted in the opposite sense. Hence the Bragg condition for reflection is differently affected on the two sides of the line. On one side, the diffracted intensity may be enhanced because the Bragg condition is better satisfied (s is smaller), whereas it is decreased on the other side because s is larger, leading to a black-white line contrast shown schematically in Fig. 17 for the case of an edge dislocation. In this schematic representation the line thickness is proportional to the local beam intensity. In bright-field images dislocation lines are thus

imaged as dark lines, slightly displaced from the actual position of the dislocation line towards the “image side.”

This model implies that imaging in reflections associated with families of lattice planes that are not deformed by the presence of the dislocation will not produce a visible line image; the image is then said to be *extinct* (Fig. 18). The extinction condition can to a good approximation be formulated as $\mathbf{H} \cdot \mathbf{b} = 0$, where $\mathbf{b}$ is the Burgers vector of the dislocation. If extinction occurs for two different sets of lattice planes with normals $\mathbf{H}_1$ and $\mathbf{H}_2$, the direction of the Burgers vector $\mathbf{b}$ is parallel to $\mathbf{H}_1 \times \mathbf{H}_2$.

Images of dislocations can be simulated accurately by numerically solving the equations that describe the dynamical scattering of the electrons in the object (see the Appendix).

Fast computer programs (10) have been developed to calculate such images for various strain fields and various diffraction conditions. An example of the agreement between the observed and the computed image that can be achieved is shown in Fig. 19 after Ref. 10.

Weak-Beam Images (11). The width of the bright peak that images a dislocation in the dark-field imaging mode decreases with increasing s . This effect is exploited systematically in the weak-beam method, which allows one to image the dislocations as very fine bright lines on a dark background, using a reflection that is only weakly excited, that is, for which s is large. Unfortunately the contrast is weak and long exposure times are required (Fig. 20).

High-Resolution Imaging

Images of Thin Objects. A very thin object acts as a phase object, in which the phase is proportional to the projected potential along the electron path. The reason is that, in an electrostatic potential, the electron changes speed, which results in a phase shift. Then the exit wave of the object can be written as

$$\psi(\mathbf{R}) \approx 1 + i\sigma V_p(\mathbf{R}) \quad (10)$$

$V_p(\mathbf{R})$ is the projected potential of the object.

In the phase-contrast mode [Appendix A1, Eq. (18)], the phase shift of $\pi/2$ changes i into -1 so that the image intensity is

$$I(\mathbf{R}) \approx 1 - 2\sigma V_p(\mathbf{R}) \quad (11)$$

The image contrast of a thin object is proportional to its electrostatic potential $V_p(\mathbf{R})$ projected along the direction of incidence.

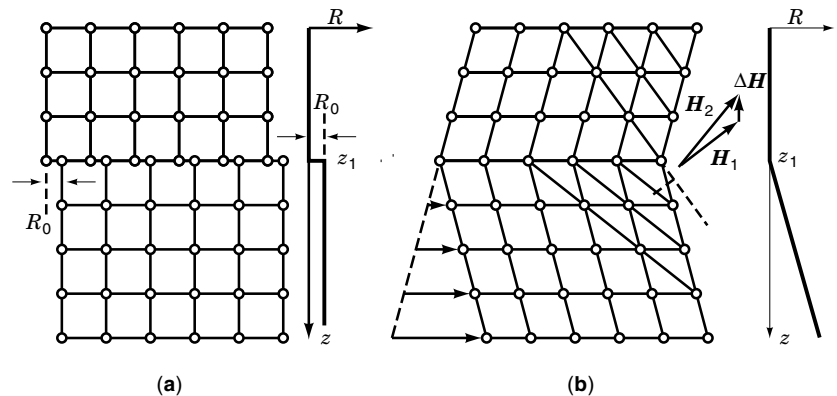


Figure 15. Displacement fields of planar interfaces: (a) stacking fault and (b) domain boundary.

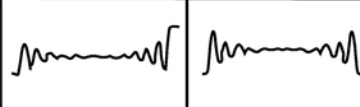
	$\alpha = 2\pi \mathbf{H} \bullet \mathbf{R}_0$					$\delta = S_1 t_{H_1} - S_2 t_{H_2}$			
	BF		DF			BF		DF	
	F	L	F	L		F	L	F	L
	$\sin \alpha > 0$	B	B	B		D	$\delta > 0$	B	D
$\alpha \neq \pi$	-	-	-	-	-	-	-	-	-
$\sin \alpha < 0$	D	D	D	B	$\delta < 0$	D	B	D	D
									

Figure 16. Comparison of fringe pattern characteristics due to a stacking fault (left) and to a domain boundary (right). The abbreviations F, L, B, and D denote first fringe, last fringe, bright, and dark, respectively.

Building-Block Structures. It often happens that a family of crystal structures exists, of which all members consist of a stacking of the simple building blocks but with a different stacking sequence. This is, for instance, the case in mixed-layer compounds, polytypes, and periodical twins, but also periodic interfaces such as antiphase boundaries and crystallographic shear planes can be considered as mixed-layer systems. If the blocks are larger than the resolution of the microscope, each

block will show its characteristic contrast. In this way, stacking of the blocks can be directly “read” from the image. The relation between image and structure is called the image code.

An example is shown in Fig. 21 for the case of the binary alloy Au_4Mn in which Au or Mn atoms are located on two sublattices of a single fcc lattice. In this image the Mn atoms are visualized as bright dots. The Au atoms are not visible. This kind of image can be interpreted unambiguously.

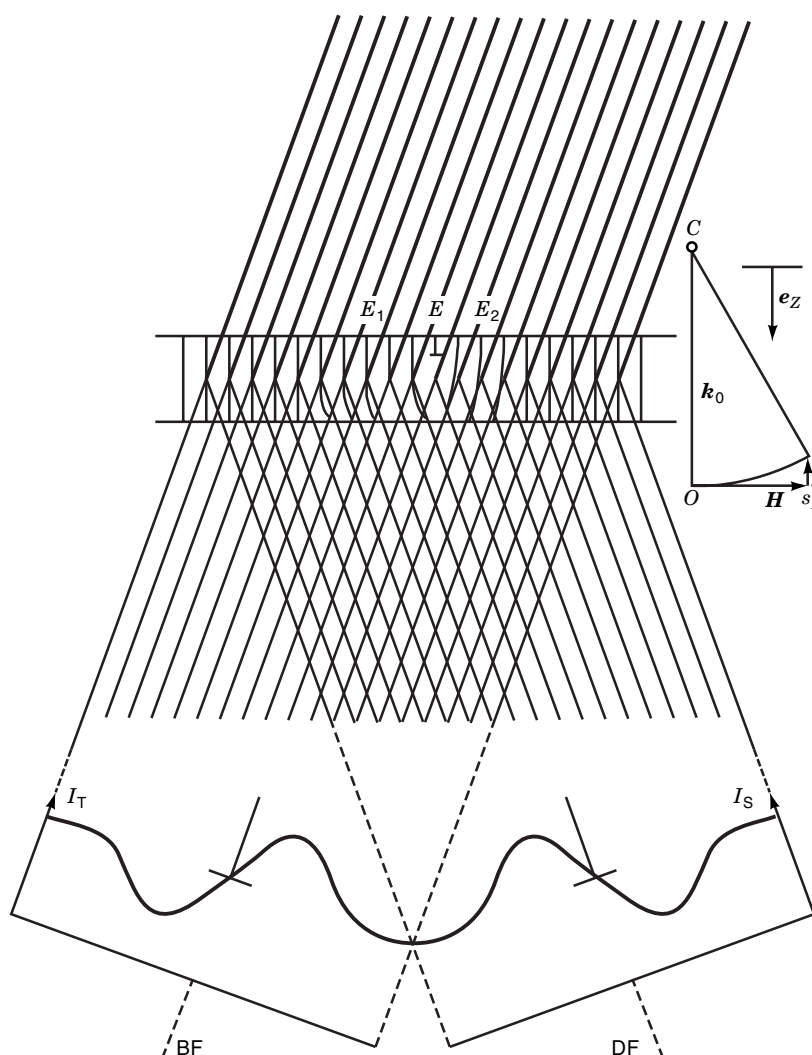


Figure 17. Image formation at an edge dislocation according to the kinematical approximation. The thickness of the lines is a measure for the intensity of the beams. I_T is the intensity of the transmitted beam and I_S of the scattered beam.

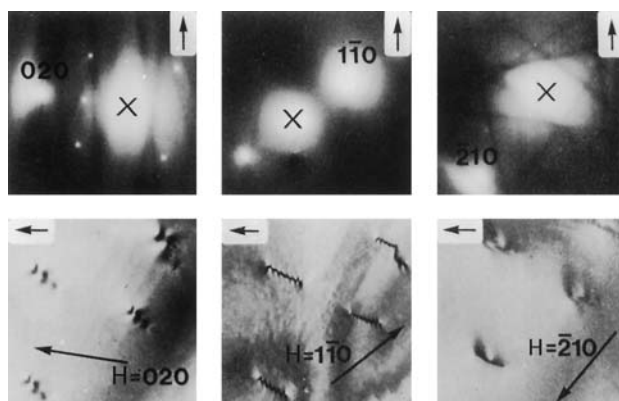


Figure 18. Three images of the same area made under two-beam conditions using three different diffraction vectors $H_1 = 020$, $H_2 = 1\bar{1}0$, $H_3 = \bar{2}10$. Note the extinctions.

Interpretation using Image Simulation. In most cases, however, the image cannot easily be decoded in terms of the object structure, making interpretation difficult, especially at very high resolution, where the image contrast can vary drastically with the focus distance. As a typical and historical example, structure images obtained by Iijima for the complex oxide

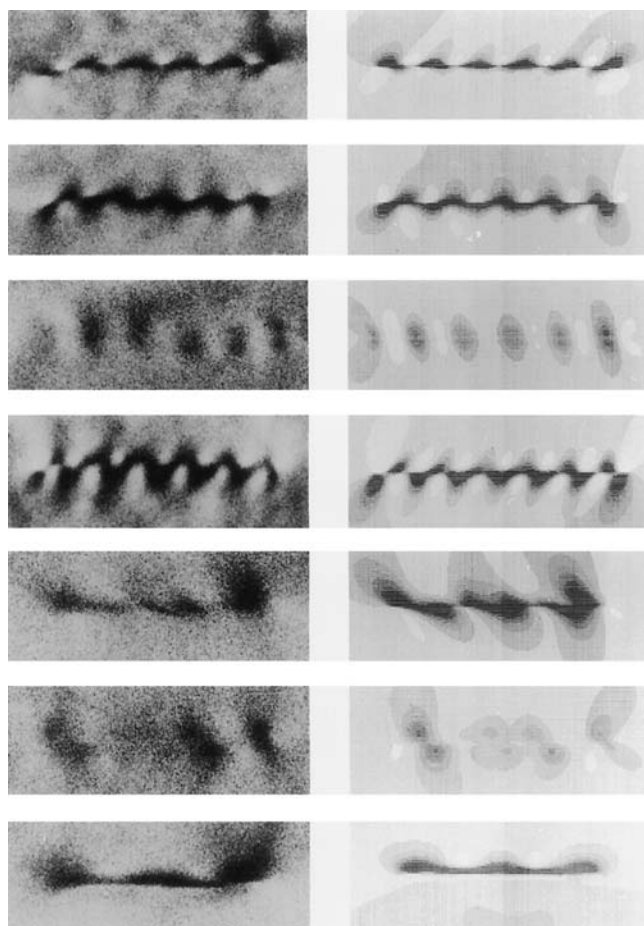


Figure 19. Comparison of observed (left) and computer simulated (right) images of dislocations. Note the excellent correspondence (after Ref. 9).

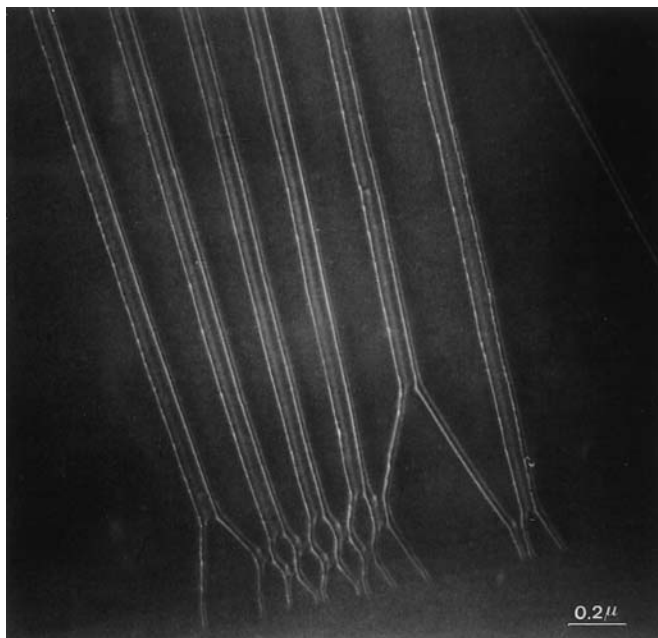


Figure 20. Weak beam image of dislocations in RuSe_2 .

$\text{Ti}_2\text{Nb}_{10}\text{O}_{25}$ with a point resolution of approximately 0.35 nm are shown in Fig. 22 (upper parts). The structure as reproduced schematically in Fig. 23 consists of a stacking of corner- or face-sharing NbO_6 octahedra with the titanium atoms in tetrahedral positions. High-resolution images are taken at different focus values, causing the contrast to change drastically. The best resemblance with the X-ray structure can be obtained near the optimum Scherzer defocus, which is -90 nm in this particular case. However, the interpretation of such high-resolution images never appears to be trivial. The only solution remains in the comparison of the experimental images with those calculated for various trial structures. The results of such a calculation using the model of Fig. 23 are also shown in Fig. 22 (lower parts) and show a close resemblance with the experimental

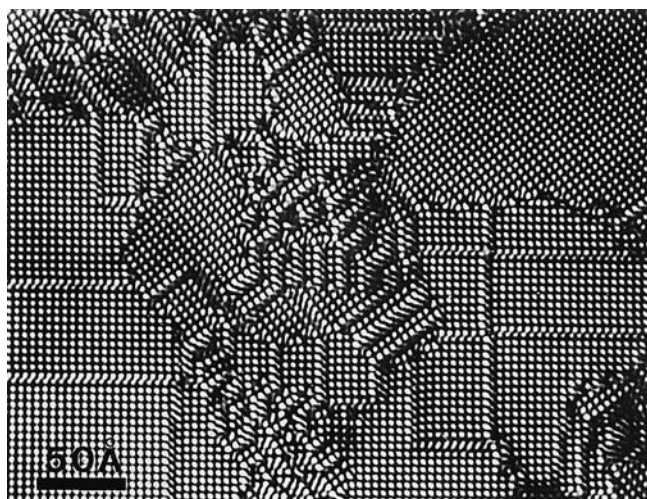


Figure 21. Dark-field superlattice image of Au_4Mn . Orientation and translation variants are revealed. (Courtesy of G. Van Tendeloo.)

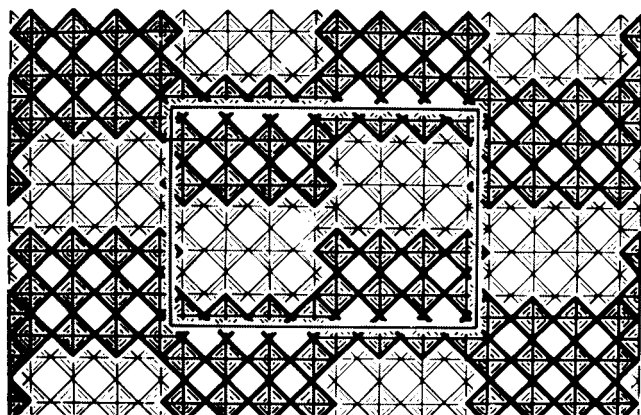


Figure 22. Schematic representation of the unit cell of $\text{Ti}_2\text{Nb}_{10}\text{O}_{29}$ consisting of corner-sharing NbO_6 octahedra with Ti atoms in tetrahedral sites.

images. Image simulation, however, is a very tedious technique that has to deal with a number of unknown parameters (specimen thickness, exact focus, beam convergence, etc.). Furthermore, the comparison is often done visually. As a consequence, the technique can only be used if the number of plausible models is very limited.

Holographic Reconstruction Methods. Due to the complexity of the imaging process, the information about the structure of the object is scrambled in the image. The structural information can be extracted directly from the images using holographic reconstruction methods. These methods aim at undoing the image process, that is, going back from image to object. Such a procedure consists of three steps. First one has to reconstruct the electron wave in the image plane. Then one has to reconstruct the exit wave at the object and from this one has to deduce the projected structure of the object. In a recorded image, which shows only intensities, the phase information is lost. Hence, the reconstruction of the whole image wave is a typical phase problem that can only be solved using holographic methods.

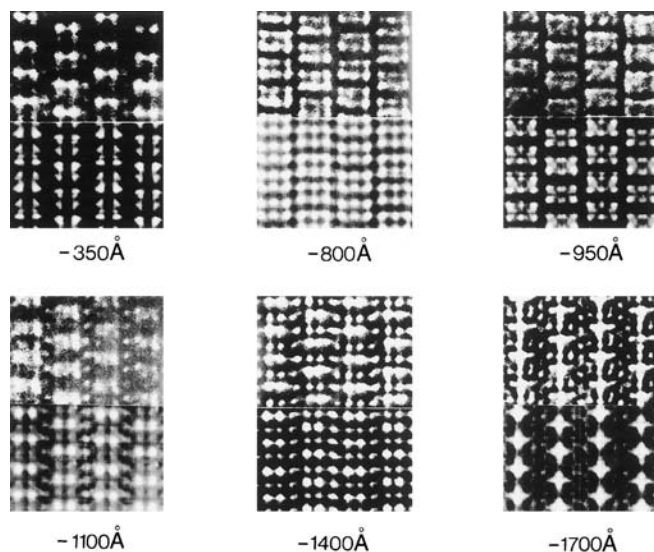


Figure 23. Comparison of experimental images (upper parts) and computer simulated images (lower parts) for $\text{Ti}_2\text{Nb}_{10}\text{O}_{29}$ as a function of defocus. (Courtesy S. Iijima.)

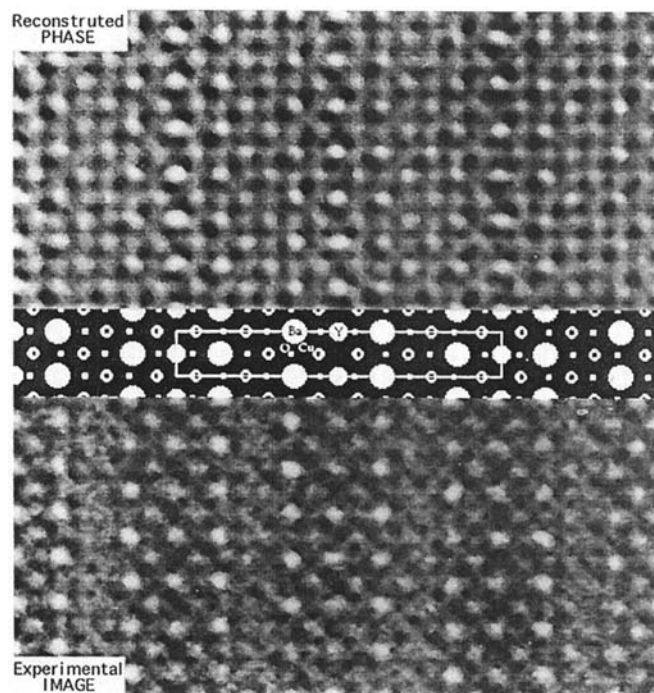


Figure 24. Experimentally reconstructed exit wave for $\text{YBa}_2\text{Cu}_4\text{O}_8$. Top, reconstructed phase; center: structure model; bottom: experimental image.

Basically two methods are workable. One method is *off-axis holography* (12). Here the electron beam is split in two waves by means of a biprism, which essentially is an electrostatically charged wire. One wave crosses the object so as to produce an enlarged image. The other wave (reference wave) passes by the object through the vacuum and interferes with the image wave in the image plane. In this way the high-resolution image is modulated by the interference fringes. From the position of the fringes one can then determine the phase of the electron wave. The other method is the *focus variation* method (1). The image wave is calculated by computer processing a series of images taken at different focus settings. Figure 24 shows an experimentally reconstructed exit wave for $\text{YBa}_2\text{Cu}_4\text{O}_8$. From this the structure of the object can be deduced.

Quantitative Structure Determination. Ideally quantitative extraction of information should be done as follows. One has a model for the object, for the electron object interaction, for the microscope transfer and for the detection, that is, all the ingredients needed to perform a computer simulation of the experiment. The object model that describes the interaction with electrons consists of the assembly of the electrostatic potentials of the constituting atoms. Also the imaging process is characterized by a number of parameters such as defocus, spherical aberration, and voltage. These parameters can either be known a priori with sufficient accuracy or not, in which case they have to be determined from the experiment. The model parameters can be estimated from the fit between the theoretical images and the experimental images. What one really wants is not only the best estimate for the model parameters but also their standard deviation (error bars), a criterion for the goodness of fit, and a suggestion for the best experimental setting. This requires a correct statistical analysis of the experimental data. The goodness of the fit between model and experiment has to

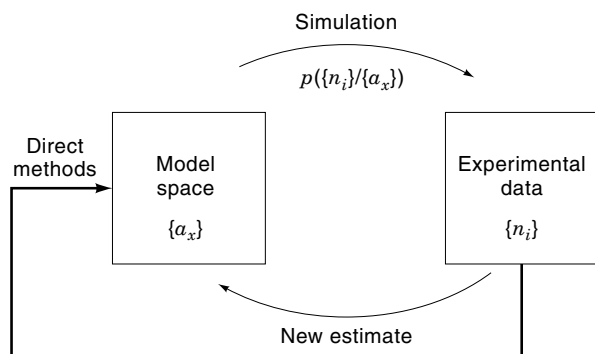


Figure 25. Scheme for the refinement procedure.

be evaluated using a criterion such as likelihood, mean square difference, or R factor (as with X-ray crystallography). For each set of parameters of the model, one can calculate this goodness of fit, so as to yield a fitness function in parameter space.

In principle, the search for the best parameter set is then reduced to the search for optimal fitness in parameter space. This search can only be done in an iterative way as given in the schematic in Fig. 25. First one has a starting model, that is, a starting value for the object and imaging parameters $\{a_n\}$. From these one can calculate the experimental images. This is a classical image simulation. (Note that the experimental data can also be a series of images and/or diffraction patterns.) From the mismatch between experimental and simulated images one can obtain a new estimate for the model parameters (for instance, using a gradient method), which can then be used for the next iteration. This procedure is repeated until the optimal fitness (i.e., optimal match) is reached.

The refinement procedure needs a good starting model to guarantee convergence. Such a model can be derived from holographic reconstruction. The refinement can also be done using experimental electron diffraction patterns.

An application of such refinement is shown in Figs. 26 and 27. Figure 26 (left) shows an HREM image of a Mg/Si precipitate in an Al matrix (13). Figure 26 (right) shows the phase of the exit wave, which is reconstructed experimentally using the focus variation method. From this an approximate structure model can be deduced. From different precipitates and different zones, electron diffraction patterns could be obtained, which were used simultaneously for a final fitting.

For each diffraction pattern the crystal thickness as well as the local orientation was also treated as a fittable parameter. An overview of the results is shown in Table 1. The obtained R factors are of the order of 5%, which is well below the R factors using kinematical refinement that do not account for the dynamical electron scattering. Figure 27 shows the structure obtained after refinement. Details of this study have been published by Zandbergen et al. (14).

SCANNING ELECTRON MICROSCOPY

The SEM is a mapping, rather than an imaging device (Fig. 28) and so is a member of the same class of instruments as the facsimile machine, the scanning probe microscope, and the confocal optical microscope (19). The sample is probed by a

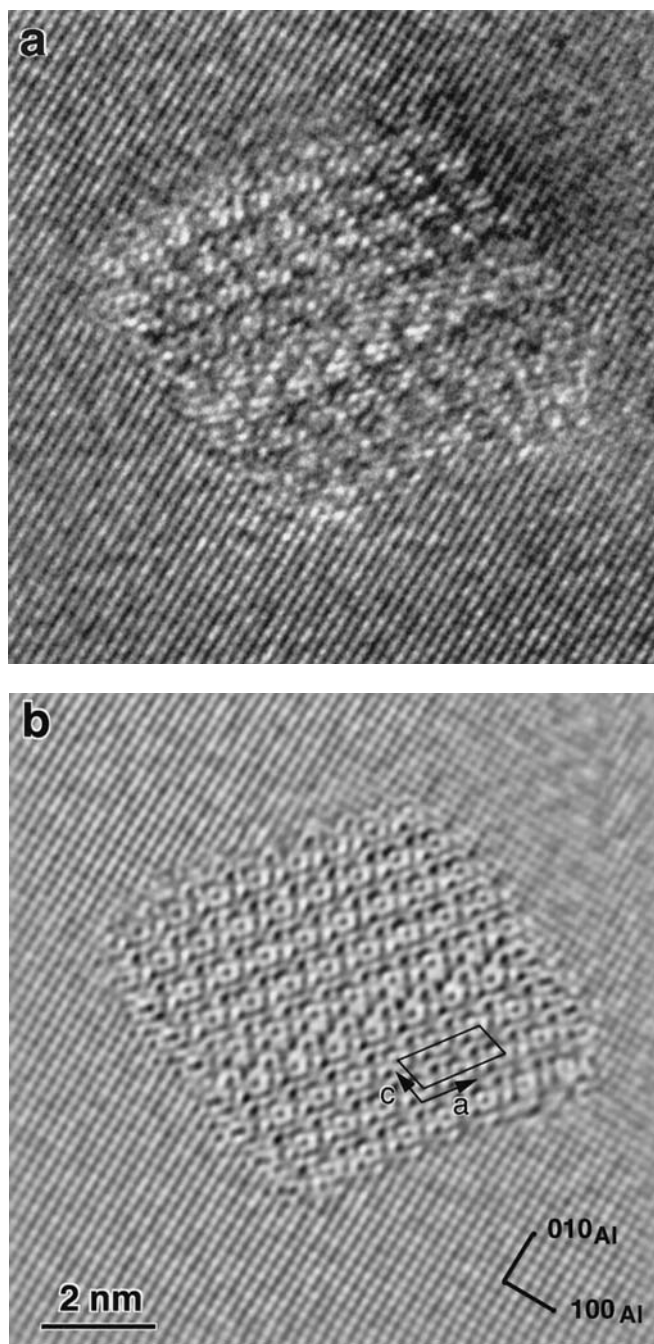


Figure 26. HREM image (a) and phase of the experimentally reconstructed exit wave (b) of an Mg/Si precipitate in an Al matrix. [Courtesy H. Zandbergen (14).]

beam of electrons scanned across the surface. Radiations from the specimen stimulated by the incident beam are detected, amplified, and used to modulate the brightness of a second beam of electrons scanned, synchronously with the first beam, across a cathode-ray-tube display. If the area scanned on the display tube is $A \times A$ and the corresponding area scanned on the sample is $B \times B$, then the linear magnification $M = A/B$. The magnification is therefore geometric in origin and may be changed by varying the area scanned on the sample. The arrangement makes it possible for a wide range of magnifications

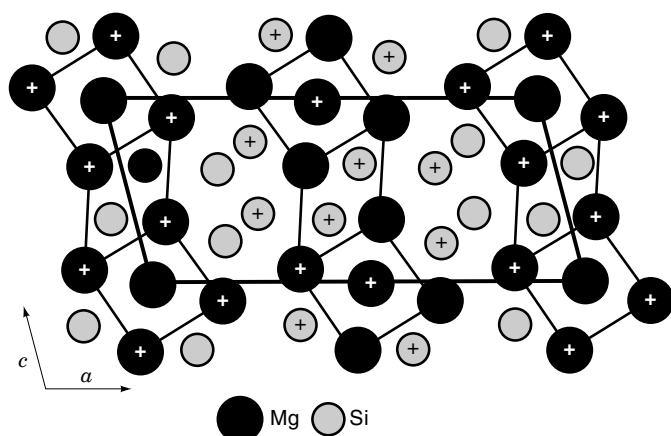


Figure 27. Structure model with MSLS from the fitting procedure described in the text. [Courtesy H. Zandbergen (14).]

to be obtained and allows rapid changes of magnification since no alterations to the electron-optical system are required. There is no rotation between object and image planes, and once the instrument has been focused on a given area the focus need not be changed when the magnification is varied. To a first approximation the size of the finest detail visible in the image will be set by the size of the probe scanning the specimen. Multiple detectors can be used to collect several signals simultaneously that can then be displayed individually or combined in perfect register with each other. It is this probability in particular that makes the SEM so useful a tool since multiple views of a sample, in different imaging modes, can be collected and compared in a single pass of the beam.

Figure 29 shows the basic components of a SEM. These can be divided into two main categories: the electron-optical and detector systems and the scanning processing and display systems. The electron-optical components are often described as being the “column” of the instrument while the other items are the “console” of the machine. The source of electrons is the gun, which produces electrons either by thermal emission, from tungsten or lanthanum hexaboride cathodes, or from a field-emission source. These electrons are then accelerated to an energy in the range from 500 eV to 30 keV. The beam of electrons leaving the gun is then focused onto the specimen by one or more condenser lenses. Although either electrostatic or electromagnetic lenses could be employed all modern SEMs use electromagnetic lenses. Typically, the final objective lens

has been of the pinhole design with the sample sitting outside the field of the lens since this arrangement gives good physical access to the specimen. However, in this arrangement the specimen is 10 mm to 20 mm away from the lens, which must therefore be of long focal length and correspondingly high aberration coefficients. In modern, high-performance instruments it is now common to use an immersion lens (15), in which the sample sits inside the lens at the center of the lens field, or a “snorkel” lens (16) in which the magnetic field extends outside of the lens to envelop the sample. Although the immersion lens gives very good performance and, by making the sample part of the lens structure, ensures mechanical stability, the amount of access to the specimen is limited. The snorkel lens, on the other hand, combines both good electron-optical characteristics with excellent access for detectors and stage mechanisms.

The coils that scan the beam are usually incorporated within the objective lens. A double-scan arrangement is often employed in which one sets of coils scans the beam through some angle θ from the axis of the microscope while a second set scans the beam through an angle 2θ in the opposite direction. In this way all scanned beams pass through a single point on the optic axis allowing for the placement of a defining aperture without any constriction of the scanned area. The scan pattern, produced on the specimen, is usually square in shape and is made up of 1000 horizontal lines, each containing 1000 individual scanned points or *pixels*. The final image frame thus contains 10^6 pixels, although for special activities such as focusing or alignment frames containing only 256×256 pixels may be used.

Increasingly the detector output is passed through an analog-to-digital convertor (ADC) and then handled digitally rather than as an analog video signal. This permits images to be stored, enhanced, combined, and analyzed using either an internal or an external computer. While the majority of the images are still recorded onto photographic film, digital images can be stored directly to magnetic or magneto-optic disks, and hard-copy output of the images can then be obtained using laser or dye sublimation printers. Typically scan repetition rates ranging from 15 or 20 frames/s (“TV rate”) to one frame in 30 s to 60 s (“photographic rate”) are provided. In addition individual pixels or arrays of pixels within an image field may be accessed if required.

In the case of the SEM the attainable resolution is determined by a number of factors, including the diameter of the electron-beam probe that can be generated, the current I_b contained in that probe, the magnification of the image, and the type of imaging mode that is being used. Over most of the

Table 1. Results of Structure Refinement Using Electron Diffraction Data

Zone	Number of Observed Reflections	Thickness (nm)	Crystal Misorientation			R Value (%)	
			<i>h</i>	<i>k</i>	<i>l</i>	MSLS	Kinematic
[010]	50	6.7(5)	8.3	0	−2.3	3.0	3.7
[010]	56	15.9(6)	2.6	0	−1.8	4.1	8.3
[010]	43	16.1(8)	−1.7	0	0.3	0.7	12.4
[010]	50	17.2(6)	−5.0	0	−1.0	1.4	21.6
[010]	54	22.2(7)	−5.9	0	2.5	5.3	37.3
[001]	72	3.7(3)	−3.9	4.5	0	4.1	4.5
[001]	52	4.9(6)	3.6	−1.9	0	6.8	9.3

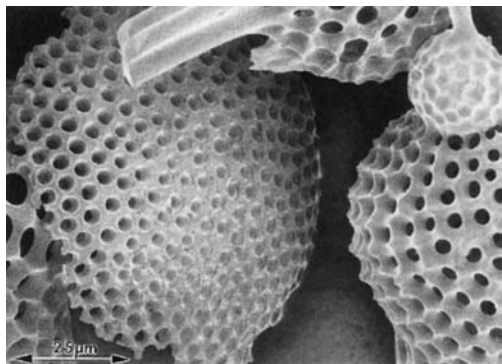


Figure 30. Secondary-electron images of radiolarium. Recorded in Hitachi S-4500 field emission SEM at 5 keV beam energy. Magnification: 800 $\times$. [Courtesy D. Joy (19).]

Other Imaging Modes. With a SEM it is possible to measure the current through the object as induced by the imaging electron beam [electron-beam-induced current (EBIC)]. This signal gives information about the electron-hole pair carriers in a semiconductor such as those at p - n junctions. In cathode luminescence, one detects the fluorescence radiation that is due to irradiation by the incident beam. This is a very sensitive technique that gives information about the impurities in semiconductors.

For more information we refer the reader to Ref. 19.

SCANNING TRANSMISSION ELECTRON MICROSCOPY

In principle a STEM can be considered as a SEM in which the object is transparent for the high energy electrons and in which the detector is placed behind the object. As in a STEM, a fine electron probe, formed by using a strong objective electron lens to demagnify a small source, is scanned over the specimen in a two-dimensional raster [Fig. 33(a)]. The electron probe is necessarily convergent: the convergence angle is, ideally, inversely proportional to the minimum probe size that determines the microscope resolution. On any plane after the specimen, a



Figure 31. High-resolution image of magnetic disk media surface recorded at 30 keV in a JEOL JSM 890 field-emission SEM. [Courtesy D. Joy (19).]

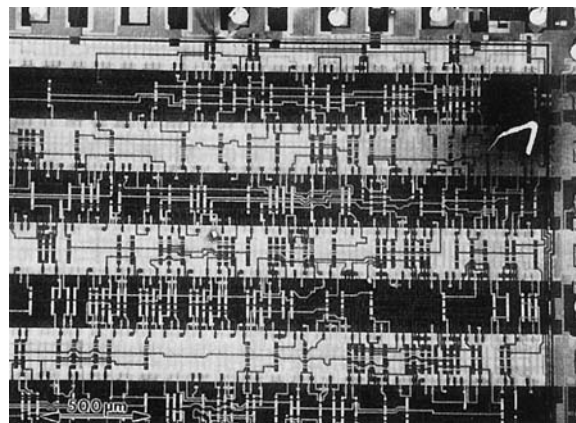


Figure 32. Voltage contrast from an integrated circuit. Recorded at 5 keV in a Hitachi S-800 field-emission SEM. [Courtesy D. Joy (19).]

convergent beam electron-diffraction pattern is formed. Some part of this diffraction pattern is collected in a detector, creating a signal, which is displayed on a cathode-ray-tube screen to form the image using a raster scan matched to that which deflects the incident electron beam (20,21).

Dark-field images, obtained with an annular detector in a STEM instrument, showed the first clear electron microscopy images of individual heavy atoms (22). From that time, STEM has developed as an important alternative to conventional, fixed-beam transmission electron microscopy (CTEM), with special advantages for many purposes.

The use of a field emission gun (FEG) for high-resolution STEM is necessary to provide sufficient signal strength for viewing or recording images in a convenient time period. Because the FEG source has a brightness that is a factor of 10^4 or 10^3 greater than that of a W hairpin filament, the total current in the electron beam is greater when beam diameters of less than about 10 nm are produced. The current in a beam of 1 nm diameter is typically about 1 nA. As suggested by Fig. 33(b), the essential components of a STEM imaging system are the same as those for a CTEM instrument, with the electrons traveling in the opposite direction. In this diagram condensor and projector lenses have been omitted, and only the essential objective lens, which determines the imaging characteristics, is included. The STEM detector replaces the CTEM electron source. The STEM gun is placed in the detector plane of the CTEM, and the scanning system effectively translates the STEM source to cover the CTEM recording plate. When one uses a detector with a hole to eliminate the unscattered electron beam, the imaging is effectively incoherent so that the image contrast can be interpreted directly in terms of the atomic number of the constituting atoms. This imaging mode is therefore called Z contrast imaging (20). Figure 34 shows a STEM image of a tilt boundary in silicon in which the local atomic configuration can be seen directly in the images.

The strength of STEM as compared to TEM is that a variety of signals may be obtained in addition to the bright-field or dark-field signals derived from the elastic scattering of electrons in the specimen. STEM instruments are visually fitted with an energy-loss spectrometer. Energy filtered images reveal compositional information. For more information we refer to Refs. 20 and 24.

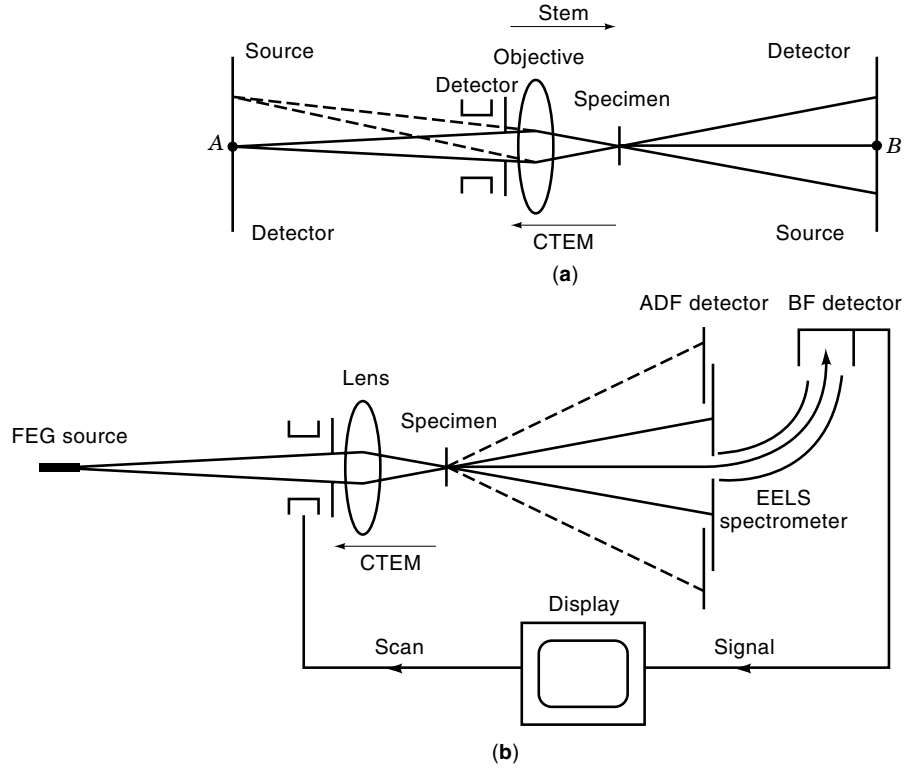


Figure 33. (a) Diagram of the essential components of a STEM instrument. (b) Diagram suggesting the reciprocity relationship between STEM (electrons going from left to right) and CTEM (electrons going from right to left). [Courtesy J. Cowley (20).]

APPENDIX A. ELECTRON-DIFFRACTION THEORIES

Phase Object Approximation

We will now follow a classical approach. The nonrelativistic expression for the wavelength of an electron accelerated by an electrostatic potential E is given by

$$\lambda = \frac{h}{\sqrt{2meE}} \quad (\text{A1})$$

with h the Planck constant, m the electron mass and e the electron charge.

During the motion through an object with local potential $V(x, y, z)$ the wavelength will vary with the position of the electron as

$$\lambda'(x, y, z) = \frac{h}{\sqrt{2me[E + V(x, y, z)]}} \quad (\text{A2})$$

For thin phase objects and large accelerating potentials the assumption can be made that the electron keeps traveling along the z direction so that by propagation through a slice dz the electron suffers a phase shift.

$$\begin{aligned} d_\phi(x, y, z) &= 2\pi \frac{dz}{\lambda'} - 2\pi \frac{dz}{\lambda} \\ &= 2\pi \frac{dz}{\lambda} \left(\frac{\sqrt{E + V(x, y, z)}}{\sqrt{E}} - 1 \right) \\ &\simeq \sigma V(x, y, z) dz \end{aligned}$$

with

$$\sigma = \pi/\lambda E \quad (\text{A3})$$

Therefore the total phase shift is given by

$$\phi(x, y) = \sigma \int V(x, y, z) dz = \sigma V_p(x, y) \quad (\text{A4})$$

where $V_p(x, y)$ represents the potential of the specimen projected along the z direction.

Under this assumption the specimen acts as a pure phase object with transmission function

$$\psi(x, y) = \exp[i\sigma V_p(x, y)] \quad (\text{A5})$$

In case the object is very thin, one has

$$\psi(x, y) \approx 1 + i\sigma V_p(x, y) \quad (\text{A6})$$

This is the weak-phase approximation.

The effect of all processes prohibiting the electrons from contributing to the image contrast, including the use of a finite aperture, can in a first approximation be represented by a projected absorption function in the exponent of Eq. (A.5) so that

$$\psi(x, y) = \exp[i\sigma V_p(x, y) - \mu(x, y)] \quad (\text{A7})$$

or

$$\psi(\mathbf{R}) = \exp[i\sigma V_p(\mathbf{R}) - \mu(\mathbf{R})] \quad (\text{A8})$$

with $\mathbf{R} = (x, y)$ the vector in the plane perpendicular to z .

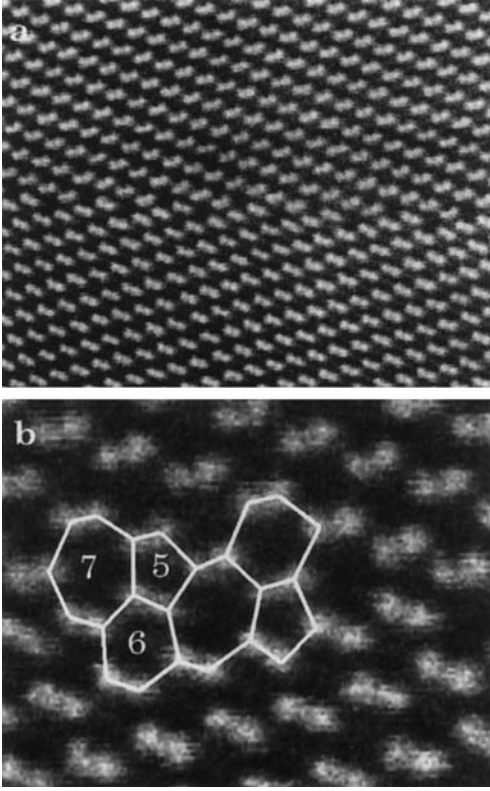


Figure 34. $\Sigma = 9, [221](100)$ symmetric tilt boundary in silicon viewed along the $[110]$ direction showing its five- and seven-membered ring structure. ADF = annular dark field; EELS = electron energy loss spectroscopy. [Courtesy S. Pennycook (21).]

Kinematical Theory

As follows from Eq. (1) a diffraction pattern can be calculated from the Fourier transforms of the exit wave $\psi(\mathbf{R})$. However, even for a simple approximation such as Eq. (A.8) the Fourier transform is not expressed in a simple analytical form. In order to derive a simpler, albeit approximated, expression for the diffraction pattern it is more convenient to describe the diffraction process directly in Fourier space.

According to the kinematical diffraction theory electrons are scattered in the specimen only and moreover the incident beam is not depleted by scattering. Each atom (scattering center) thus sees the same incident beam amplitude. This approximation is excellent in neutron diffraction, justified in X-ray diffraction, but it is poor in electron diffraction because the atomic scattering cross sections for electrons are relatively much larger than those for the other forms of radiation. The kinematical approximation is therefore only applicable to very thin crystals (a few nanometers for most materials) or for very large deviations from the exact Bragg condition (large s). It allows one to compute the amplitude of the diffracted beam only since the incident beam remains undepleted. Qualitative conclusions from the kinematical theory are nevertheless usually in agreement with the observations.

A crystal is made up of identical unit cells, regularly arranged at the basic lattice nodepoints given by

$$\mathbf{A}_L = l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3 \quad (\text{A9})$$

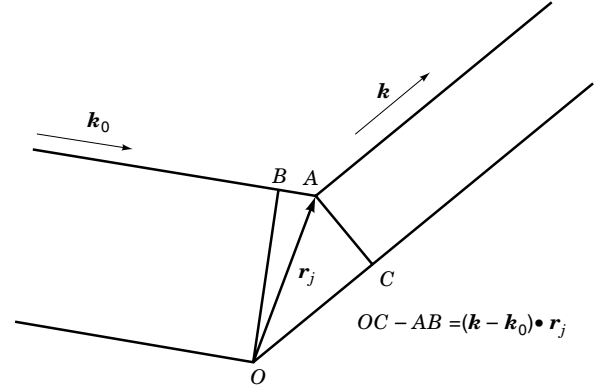


Figure 35. Illustrating the path difference $OC - AB$ between waves diffracted by an atom at the origin and an atom at $\mathbf{r}_j$.

(where l_j is an integer). In each unit cell, a number N of atoms is found at the relative positions ρ_k ($k = 1, \dots, N$). Mathematically speaking, the whole crystal is made up by convolution of one unit cell with the basic crystal lattice. Atom positions are thus $\mathbf{r}_j = \mathbf{A}_L + \rho_k$ and they depend on four indices l_1, l_2, l_3 , and k .

Let $\mathbf{k}_0$ represent the wave vector of the incident wave and $\mathbf{k}$ that of the diffracted wave; then at large distance, i.e., in the Fraunhofer approximation, the phase difference between a wave diffracted by an atom at the origin and an atom at $\mathbf{r}_j$ is given by $2\pi(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}_j$ and the scattered amplitude $A(\mathbf{k})$ along the direction of $\mathbf{k}$ (Fig. 35)

$$A(\mathbf{k}) = \sum_j f_j \exp[2\pi i(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}_j] \quad (\text{A10})$$

This amplitude will exhibit maxima if all exponents are integer multiples of $2\pi i$; maxima will thus occur if $(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}_L$ is integer, which implies that $\mathbf{k} - \mathbf{k}_0$ must be a reciprocal-lattice vector

$$\mathbf{k} - \mathbf{k}_0 = \mathbf{B}_H \equiv h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3 \quad (\text{A11})$$

(where h_j are integers and $\mathbf{b}_i$ are base vectors). This is Ewald's condition as discussed in the section on electron diffraction.

However, $A(\mathbf{k})$ will also be different from zero if the diffraction condition is not exactly satisfied, that is, if Ewald's sphere misses closely a reciprocal-lattice node by a vector $\mathbf{s}$, called the excitation error. This vector is parallel to the foil normal and connects the reciprocal-lattice node with the intersection point with Ewald's sphere; by convention s is positive when the reciprocal-lattice node is inside Ewald's sphere and negative when outside (Fig. 7). One can now set

$$\mathbf{k} - \mathbf{k}_0 = \mathbf{B}_H + \mathbf{s} \quad (\text{A12})$$

and

$$A_H = \sum_j f_j \exp[2\pi i(\mathbf{B}_H + \mathbf{s}) \cdot \mathbf{r}_j] \quad (\text{A13})$$

and with $\mathbf{r}_j = \mathbf{A}_L + \rho_k$.

$$A_H = \sum_L \sum_k f_k \exp[2\pi i(\mathbf{B}_H + \mathbf{s}) \cdot (\mathbf{A}_L + \rho_k)] \quad (\text{A14})$$

Neglecting $\mathbf{s} \cdot \boldsymbol{\rho}_k$ as compared to the other terms and noting that $\mathbf{B}_H \cdot \mathbf{A}_L$ is always an integer, this can be written as

$$A_H = F_H \sum_L \exp(2\pi i \mathbf{s} \cdot \mathbf{A}_L) \quad (\text{A15})$$

where the structure factor F_H is defined as

$$F_H = \sum_k f_k \exp[2\pi i (\mathbf{B}_H \cdot \boldsymbol{\rho}_k)] \quad (\text{A16})$$

Equation (A.15) is in fact a triple sum over the indices $\mathbf{L}(l_1, l_2, l_3)$. If $\mathbf{s}$ is written as a vector in terms of the reciprocal-space base vectors $\mathbf{b}_j$:

$$\mathbf{s} = s_1 \mathbf{b}_1 + s_2 \mathbf{b}_2 + s_3 \mathbf{b}_3 \quad (\text{A17})$$

one has

$$\mathbf{s} \cdot \mathbf{A}_L = l_1 s_1 + l_2 s_2 + l_3 s_3 \quad (\text{A18})$$

The triple sum can be expressed as the product of three single sums of geometrical progressions. Calling N_1 , N_2 , and N_3 the numbers of unit cells along the three lattice directions $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, one obtains, neglecting an irrelevant phase factor

$$\begin{aligned} A_H &= F_H \sum_{l_1=0}^{N_1-1} \sum_{l_2=0}^{N_2-1} \sum_{l_3=0}^{N_3-1} \exp[2\pi i (l_1 s_1 + l_2 s_2 + l_3 s_3)] \\ &= F_H \frac{\sin \pi s_1 N_1}{\sin \pi s_1} \frac{\sin \pi s_2 N_2}{\sin \pi s_2} \frac{\sin \pi s_3 N_3}{\sin \pi s_3} \end{aligned} \quad (\text{A19})$$

This is the well-known *von Laue interference function* (10) (Fig. 36), which describes the dependence of the scattered amplitude on the deviation parameter s . The sine functions in the denominators can be approximated by their arguments, since these are always small. We further note that for large N one has $\sin(\pi N s)/N \pi s \sim \delta(s)$ with $\delta(s) = 0$ for $s \neq 0$ and $\delta(s) = 1$ for $s = 0$. We can then write, neglecting irrelevant phase factors,

$$A_H = F_H \delta(s_1) \delta(s_2) \delta(s_3) (\Omega / V_a) \quad (\text{A20})$$

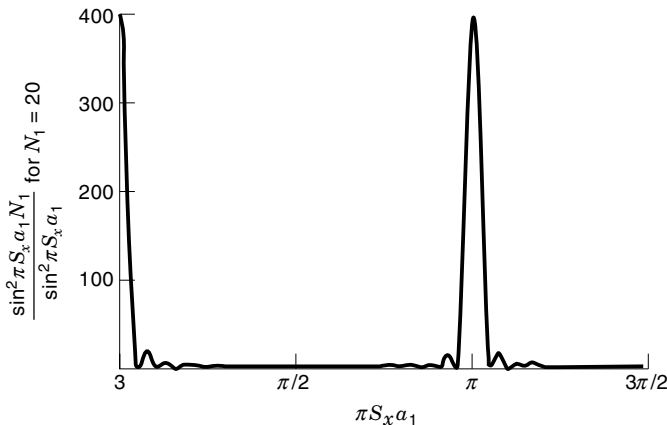


Figure 36. von Laue interference function describing the dependence of the scattered intensity on the excitation error s_x .

where Ω is the volume of the crystal and V_a the volume of the unit cell ($\Omega = N_1 N_2 N_3 a_1 a_2 a_3$; $V_a = a_1 a_2 a_3$). For a parallelepiped-shaped crystal block one often introduces the components of $\mathbf{s}$ along the three mutually perpendicular edges of the block with unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$:

$$\mathbf{s} = s_x \mathbf{e}_x + s_y \mathbf{e}_y + s_z \mathbf{e}_z \quad (\text{A21})$$

One can then rewrite Eq. (A.19) in terms of s_x , s_y , and s_z as

$$A_H = F_H \frac{\sin \pi s_x N_1 a_1}{\sin \pi s_x a_1} \frac{\sin \pi s_y N_2 a_2}{\sin \pi s_y a_2} \frac{\sin \pi s_z N_3 a_3}{\sin \pi s_z a_3} \quad (\text{A22})$$

Hereby use was made of the relations valid for a parallelepiped

$$s_1 = s_x a_1, \quad s_2 = s_y a_2, \quad s_3 = s_z a_3 \quad (\text{A23})$$

Equation (A.22) is only true if N_1 , N_2 , and N_3 are sufficiently large. However, the foils used in transmission electron microscopy are only large in two dimensions, that is, along x and y , the foil being only a small number N_3 of unit cells thick along z . In such a foil one thus obtains

$$A_H = F_H \frac{\Omega}{V_a} \delta(s_x) \delta(s_y) \frac{\sin \pi s_z N_3 a_3}{\pi s_z a_3} \quad (\text{A24})$$

Introducing the specimen thickness $N_3 a_3 = t$ and assuming s_x , $s_y = 0$ and calling $s_z = s$ one finds

$$A_H = \frac{\sin \pi s t}{s t_H} \quad (\text{A25})$$

per surface area where $t_H = \pi / F_H$; t_H is called the *extinction distance*. This result is interpreted as meaning that the sharp reciprocal-lattice nodes, characteristic of a large crystal, become rods in the case of a thin plate, as already mentioned before. These rods are perpendicular to the foil plane and have a weight profile given by $\sin(\pi s t) / s t_H$. The corresponding intensity is given by (Fig. 37)

$$I_H = \frac{\sin^2 \pi s t}{(s t_H)^2} \quad (\text{A26})$$

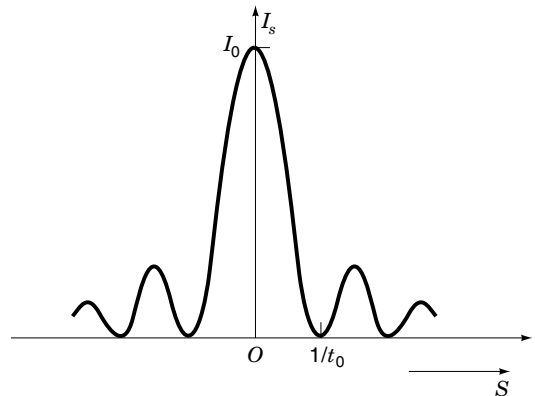


Figure 37. Rocking curve for a foil with thickness t_0 according to the kinematical theory.

it is called the *rocking curve*. An intensity can be associated with each intersection point of the Ewald sphere with this rod (called *relrod*), that is, with each s value, the intensity being given by the value of the function at the intersection point.

Another way to interpret these results is the following: In the Fraunhofer approximation, the diffracted wave is expressed by the Fourier transform of the electrostatic potential of the crystal.

A crystal can be considered as a convolution product of two factors: The convolution theorem then states that the diffracted wave is given by the product of the respective Fourier transforms of these two factors [e.g., Eq. (A.15)]. The Fourier transform of the lattice function yields delta functions at the reciprocal nodepoints, which describe the directions of the diffracted beams. The amplitudes of these beams are then given by the Fourier transforms of the potential of one unit cell, i.e., the structure factors [e.g., Eq. (A.20)]. The conservation of energy also requires that the wavevectors of the diffracted beams should all have constant length, or, that the reciprocal nodes should lie on a sphere, the Ewald sphere. In case the object is a thin crystal slab, it can be described as the product of an infinite crystal with a slab function that is equal to 1 inside the slab and 0 elsewhere. In that case, the diffraction pattern is given by the convolution product of the diffraction pattern of the infinite crystal with the Fourier transform of the slab function. Then each reciprocal node is smeared along a line perpendicular to the slab, with a factor given by a sinc function of the form in Eq. (A.26).

Note that if the Ewald sphere would be flat, the diffracted wave can be derived from the Fourier transform of Eq. (A.8) provided the phase is weak. This means that, apart from the Ewald sphere, the weak phase object theory and the kinematical theory are equivalent.

Equation (A.15) can also be understood intuitively in terms of the column approximation along the z direction. The amplitude of the wave diffracted by the volume element Δz_n at level z_n in the column (measured from the entrance face) is given by $\Delta A_H = F_H \Delta z_n \exp(2\pi i s z_n)$ or in differential form

$$dA_H = F_H \exp(2\pi i s z) dz \quad (\text{A27})$$

The amplitude at the exit face of the column is then given by the sum

$$A_H = F_H \sum_n \exp(2\pi i s z_n) \Delta z_n \quad (\text{A28})$$

which, if $s = \text{const}$, can be approximated by the integral

$$A_H = F_H \int_0^t \exp(2\pi i s z) dz \quad (\text{A29})$$

or

$$A_H = (F_H \sin 2\pi s t) / \pi s \quad (\text{A30})$$

which is consistent with Eq. (A.26), though not identical.

In the complex plane the sum equation (A.28) can be represented by an amplitude-phase diagram (7) (Fig. 38); it consists of the vector sum of elementary vectors, all of the same length, each representing the amplitude diffracted by a unit cell. Successive unit cells along a column in a perfect crystal

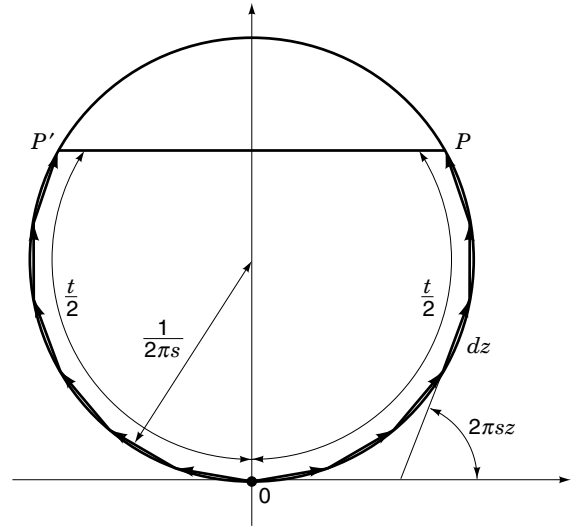


Figure 38. Complex plane construction of the amplitude-phase diagram for a perfect foil.

diffract with constant phase differences, that is, the corresponding vectors enclose constant angles. The diagram is a regular polygon that can be approximated by a circle with radius $F_H/2\pi s$. The length of the arc of the circle is equal to the column length. The amplitude diffracted by the column is given by the length of the segment connecting the endpoints P' and P of the arc. It is clear that the amplitude will be zero if the column length (i.e., the foil thickness t) is an integer number of complete circles long. The maximum amplitude is equal to the diameter of the circle, that is, to $F_H/\pi s$. Along deformed columns the amplitude-phase diagrams become curved (spiral shaped) since the angle between successive segments is no longer constant.

Two-Beam Dynamical Theory for Perfect Crystals

The dynamical theory takes into account that a scattered beam can act in turn as an incident beam and be scattered again in the interior of the crystal. The simplest case that can analytically be discussed and which moreover is relevant for image formation in the diffraction contrast mode is the two-beam case. Next to the incident beam only one beam is strongly excited (has small s). This scattered beam is then again an incident under the Bragg angle for the same set of lattice planes and can thus be scattered again. This interplay between incident and scattered beam tends to obliterate the destruction between incident and scattered beam in the interior of the crystal; it limits strongly the lateral spread of an incident Bragg diffracted electron beam and justifies the column approximation used in image calculations of defects according to the diffraction contrast mode. The dynamical theory is applicable to "thick" crystals provided also absorption is taken into account. It allows to compute the amplitudes of the transmitted beam as well as of the diffracted beam for a single Bragg reflection.

We will include all usual approximations in the model already from the onset. These approximations are as follows: (i) ideal two-beam situation, (ii) no absorption, and (iii) column approximation. Within a column along z , perpendicular to the

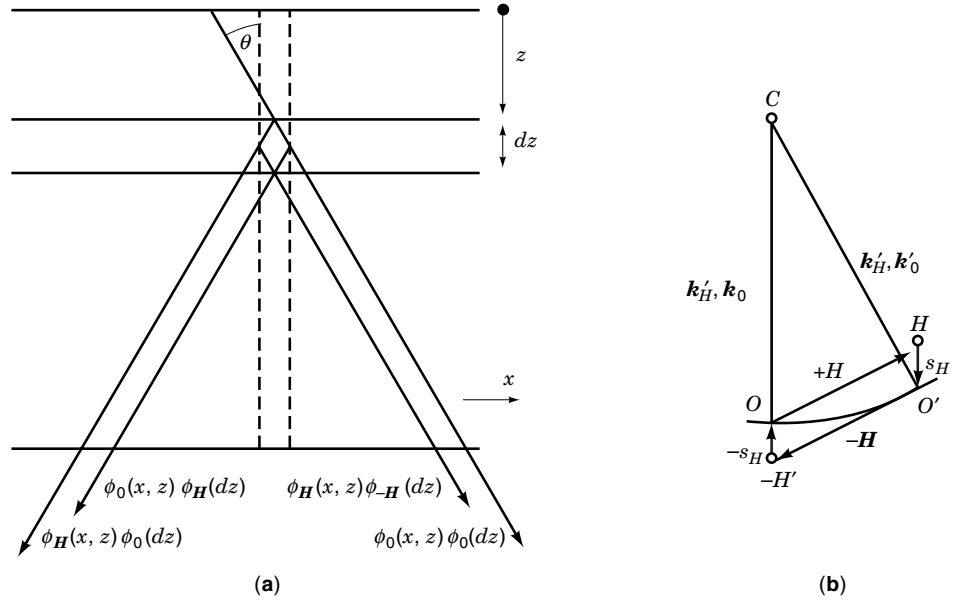


Figure 39. Schematic representation of the interfering waves during dynamical diffraction in the two-beam case. (a) Right: transmitted beam, left: scattered beam. (b) Changing $H \rightarrow -H$ changes also $s \rightarrow -s$.

foil surface, we describe the interplay between the transmitted beam represented by the plane wave $\phi_0(z) \exp(2\pi i \mathbf{k}_0 \cdot \mathbf{r})$ and the scattered beam represented by $\phi_H(z) \exp(2\pi i \mathbf{k} \cdot \mathbf{r})$ (two-beam approximation). The complex amplitudes ϕ_0 and ϕ_H depend on z only (column approximation). Within the slice dz at the level z behind the interface we express that the transmitted beam amplitude results from the interference between the twice transmitted beam with amplitude $\phi_0(z)\phi_0(dz)$ and the doubly scattered beam of which the amplitude is $\phi_H(z)\phi_{-H}(dz)$ [Fig. 39(a)]. The minus sign in $-H$ means that reflection takes place from the $-H$ side of the set of lattice planes. We thus obtain

$$\phi_0(z + dz) = \phi_0(z)\phi_0(dz) + \phi_H(z)\phi_{-H}(dz) \quad (\text{A31})$$

The slice dz being arbitrarily thin the kinematical approximation [e.g., Eq. (A.27)] can be applied rigorously to $\phi_0(dz)$ ($\equiv 1$, no beam depletion) and $\phi_{-H}(dz) = \pi i/t_H \exp(-2\pi i s z) dz$, where the factor i results from the phase change on scattering and where the structure amplitude F_H has been expressed in terms of the extinction distance t_H . Note that changing $H \rightarrow -H$ also changes the sign of s [Fig. 39(b)].

Similarly the scattered beam amplitude results from the interference between (1) the transmitted beam, which is subsequently scattered in dz , and (2) the scattered beam, which is subsequently transmitted through dz [Fig. 39(a)]. This leads to the relation

$$\phi_H(z + dz) = \phi_0(z)\phi_H(dz) + \phi_H(z)\phi_0(dz) \quad (\text{A32})$$

where again $\phi_0(dz) = 1$ and $\phi_H(dz) = (\pi i/t_H)[\exp(2\pi i s z)] dz$. The two equations (A.31) and (A.32) can be transformed into differential equations by noting that quite generally

$$\phi(z + dz) - \phi(z) = \frac{d\phi}{dz} dz \quad (\text{A33})$$

One thus obtains the following set of coupled differential

equations:

$$\begin{aligned} \frac{d\phi_0}{dz} &= [(\pi i/t_{-H}) \exp(2\pi i s z)] \phi_H(z) \\ \frac{d\phi_H}{dz} &= [(\pi i/t_H) \exp(-2\pi i s z)] \phi_0(z) \end{aligned} \quad (\text{A34})$$

in centro-symmetrical crystals $t_{-H} = t_H$. An alternative system is obtained by the substitution

$$\phi_0 = T, \quad \phi_H = S \exp(-2\pi i s z) \quad (\text{A35})$$

which only changes the phase of the amplitudes but not the resulting intensities. One obtains

$$\begin{aligned} \frac{dT}{dz} &= (\pi i/t_{-H}) S \\ \frac{dS}{dz} &= 2\pi i s S + (\pi i/t_H) T \end{aligned} \quad (\text{A36})$$

These are the *Darwin-Howie-Whelan equations* (10,11) of the two-beam dynamical diffraction theory.

The solution for a perfect crystal (i.e., s is constant) is easily obtained by the standard procedure used to solve systems of coupled first order differential equations; one finds

$$\begin{aligned} T &= [\cos(\pi \sigma_H z) - i(s_H/\sigma_H) \sin(\pi \sigma_H z)] \exp(\pi i s_H z) \\ S &= [(i/\sigma_H t_H) \sin \pi \sigma_H z] \exp(\pi i s_H z) \end{aligned} \quad (\text{A37})$$

where

$$\sigma_H^2 = (1 + s_H^2 t_H^2)/t_H^2 \quad (\text{A38})$$

The scattered intensity is thus given by the square modulus of S

$$I_S = SS^* = [\sin^2(\pi \sigma_H z)]/(\sigma_H t_H)^2 \quad (\text{A39})$$

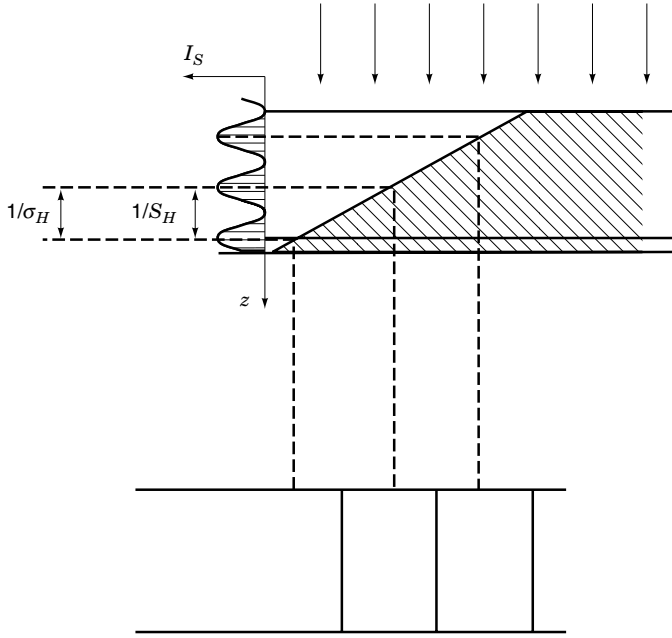


Figure 40. Illustration of the formation of thickness extinction contours.

where S^* denotes the complex conjugate of S and $I_T = 1 - I_S$ since absorption is neglected. Formula (A.39) is the homolog of formula (A.26), found in the kinematical approximation. Note that the depth period, which is $1/s_H$ in the kinematical case, now becomes $1/\sigma_H$. There is no longer a divergence for $s_H \rightarrow 0$.

Equations (A.34) describe the periodic transfer of electrons from the transmitted beam into the scattered beam and vice versa; this effect is called the *Pendellösung* effect because of its similarity with the behavior of two coupled pendulums or two coupled oscillating circuits.

Equations (A.37) describes the periodic depth variations of the diffracted and transmitted intensity, as well as the variation as a function of the excitation error s . Equation (A.39) is called the *rocking curve*. In an undeformed wedge-shaped specimen the depth variation gives rise to *thickness extinction contours*, which are parallel to the cutting edge of the wedge (Fig. 40). In a bent plane-parallel specimen the lines of constant s give rise to equi-inclination or *bent contours* (Fig. 41).

It can be shown that taking absorption into account the shape of the rocking curve becomes asymmetric in s for the transmitted beam (Fig. 42) whereas it remains symmetric for the scattered beam. [A similar effect occurs in x-ray diffraction: the *Bormann effect* (22).] The steep slope of the transmitted intensity in the vicinity of $s = 0$ is exploited in imaging strain fields due to dislocations and other defects.

Two-Beam Dynamical Theory for Faulted Crystals

Displacement Fields of Defects. In transmission electron microscopy defects are characterized by their displacement fields $R(\mathbf{r})$. The simplest example is the stacking fault, for which $R(\mathbf{r})$ is a step function, $R = 0$ for $z < z_1$, and $R = R_0$ for $z_1 < z < z_0$, z_1 being the level of the stacking fault plane in the foil and z_0 being the foil thickness. The exit part of the foil is displaced over a vector $\mathbf{R}_0$ with respect to the entrance part (Fig. 15). At the

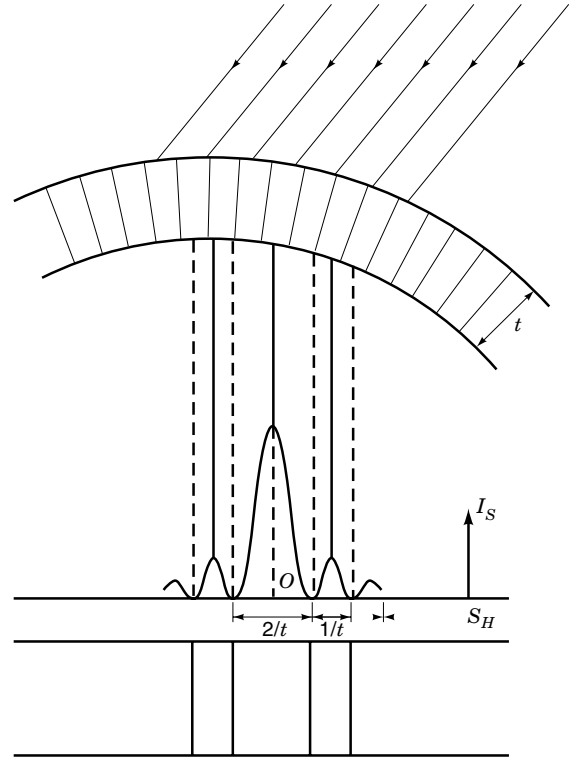


Figure 41. Illustration of the formation of equi-inclination (bent) contours.

level of the interface the diffracted beam undergoes a relative phase shift given by

$$\gamma = 2\pi \mathbf{H} \cdot \mathbf{R}_0 \quad (\text{A40})$$

whereas the transmitted beam is unaffected. The amplitude T_S of the transmitted beam for the foil containing a stacking fault

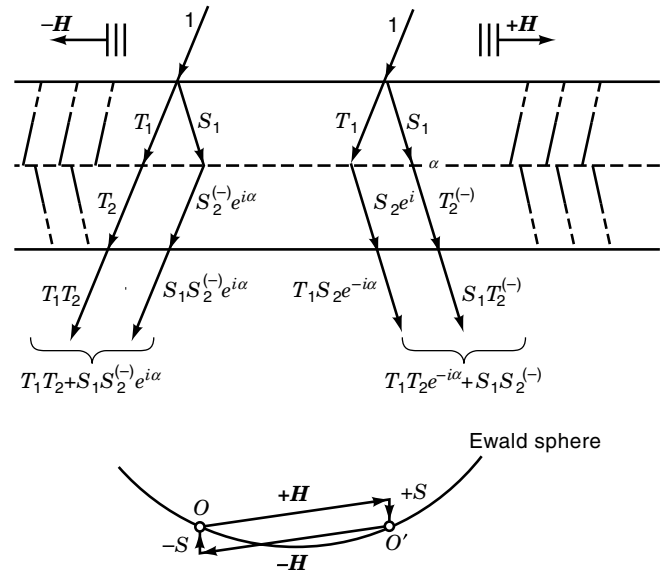


Figure 42. Schematic representation of the interfering waves in the case of a foil containing a planar interface. Left: transmitted amplitude; right: scattered amplitude.

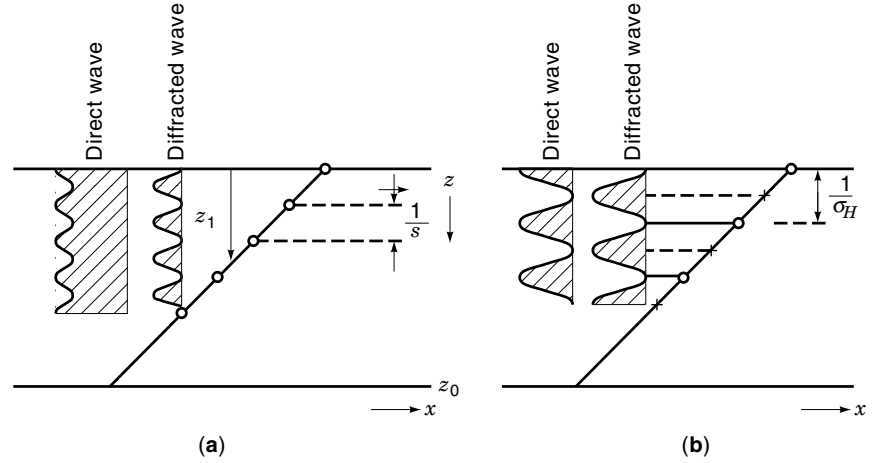


Figure 43. Cross section of foil containing a stacking fault in an inclined plane: illustrating the formation of stacking-fault fringes. (a) According to the kinematical theory $s \neq 0$; (b) according to the dynamical theory $s = 0$.

parallel to the foil plane can thus be formulated as (Fig. 42, left)

$$T_S = T_1 T_2 + S_1 S_2^- e^{i\gamma} \quad (\text{A41})$$

The expressions T_1 , T_2 , S_1 , S_2 refer to the amplitudes for perfect foils. The indices 1 and 2 refer to the entrance part and the exit part, respectively; the minus sign indicates that the excitation error is $-s$ in the corresponding expression because the diffraction vector is $-\mathbf{H}$.

Similarly the diffracted beam amplitude can be expressed as (Fig. 42, right)

$$S_S = T_1 S_2 e^{-i\gamma} + S_1 T_2^- \quad (\text{A42})$$

The meaning of Eq. (A.41) is obvious; it expresses that the transmitted beam results from the interference between the doubly transmitted beam and the doubly scattered beam. The factor $\exp(i\alpha)$ takes into account the phase shift over α of the beam scattered by the exit part. Equation (A.42) has a similar meaning. In Eq. (A.42) the phase shift is $-\alpha$ because the phase shifts are opposite in sign for S_2 and S_2^- . Detailed expressions can be obtained by replacing T_1 , T_2 , S_1 , S_2 by their explicit expressions in Eq. (A.37).

If the fault plane is inclined with respect to the foil plane, the phase change α takes place at a level z_1 , which now depends on position x along the foil. For instance, in Fig. 43, z_1 becomes a linear function of x . As a result T_S and S_S become quasiperiodic functions not only of z_1 , but also of x .

For $s = 0$ the depth period is equal to t_H ; for $s \neq 0$ it becomes $1/\sigma_H$, where σ_H is given by Eq. (A.38).

Strained Crystals. Strain fields and lattice defects are characterized by their displacement fields $\mathbf{R}(\mathbf{r})$: the atom that was at $\mathbf{r}$ before deformation will be found at $\mathbf{r} + \mathbf{R}(\mathbf{r})$ after deformation. A twin boundary with a small twinning vector (domain boundary) parallel to the foil plane at the level z_1 (Fig. 15) can, for instance, be represented by the displacement field $R = 0$ for $z < z_1$ and $R = kz$ for $z > z_1$.

A pure screw dislocation can be described by the function $\mathbf{R} = \mathbf{b}(\theta/2\pi)$, where θ is the azimuth angle measured in the plane perpendicular to $\mathbf{b}$; all displacements are parallel to $\mathbf{b}$.

The Darwin-Howie-Whelan equations (A.36) can be adapted to the case of a deformed crystal by the substitution

$$s \Rightarrow s_{\text{eff}} = s + H \frac{dR}{dz} \quad (\text{A43})$$

The Multislice Method

The two-beam dynamical treatment is insufficient for the general situation in which high resolution images are taken with the incident beam along a zone axis where many diffracted beams are involved. Therefore the multislice method was developed as a numerical method to compute the exit wave of an object.

Although the multislice formula can be derived from quantum-mechanical principles, we follow a simplified version (23) of the more intuitive original optical approach (24). A more rigorous treatment is given in the next section.

Consider a plane wave, incident on a thin specimen foil and nearly perpendicular to the incident beam direction z . If the specimen is sufficiently thin, we can assume the electron to move approximately parallel to z so that the specimen acts a pure phase object with transmission function equation (A.5)

$$\psi(x, y) = \exp[i\sigma V_p(x, y)] \quad (\text{A44})$$

A thick specimen can now be subdivided into thin slices, perpendicular to the incident beam direction. The potential of each slice is projected into a plane that acts as a two-dimensional phase object. Each point (x, y) of the exit plane of the first slice can be considered as a Huyghens source for a secondary spherical wave with amplitude $\psi(x, y)$ (Fig. 44).

Now the amplitude $\psi(x', y')$ at the point (x', y') of the next slice can be found by the superposition of all spherical waves of the first slice, that is, by integration over x and y , yielding

$$\psi(x', y') = \int \exp[i\sigma V_p(x, y)] \times \frac{\exp(2\pi ikr)}{r} dx dy \quad (\text{A45})$$

When $|x - x'| < \epsilon$, $|y - y'| < \epsilon$, with ϵ the slice thickness, the

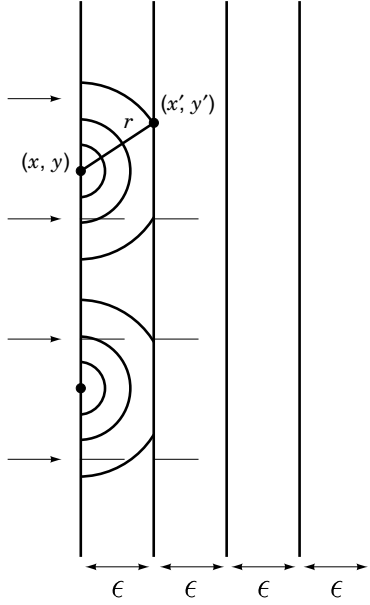


Figure 44. Schematic representation of the propagation effect of electrons between successive slices of thickness ϵ .

Fresnel approximation can be used, that is,

$$\begin{aligned} r &= \sqrt{(x - x')^2 + (y - y')^2 + \epsilon^2} \\ &\approx \epsilon \left(1 + \frac{(x - x')^2}{2\epsilon^2} + \frac{(y - y')^2}{2\epsilon^2} \right) \end{aligned} \quad (\text{A46})$$

so that

$$\begin{aligned} \psi(x', y') &\approx \frac{\exp(2\pi i k \epsilon)}{\epsilon} \int \exp[i\sigma V_p(x, y)] \\ &\times \exp\left(i \frac{\pi k}{\epsilon} [(x - x')^2 + (y - y')^2]\right) dx dy \end{aligned} \quad (\text{A47})$$

which, apart from constant factors, can be written as a convolution product:

$$\psi(x, y) = \epsilon [i\sigma V_p(x, y)] \exp\left(\frac{i\pi k(x^2 + y^2)}{\epsilon}\right) \quad (\text{A48})$$

where the convolution product of two functions is defined as (in one dimension)

$$f(x) * g(x) = \int f(x') g(x - x') dx' \quad (\text{A49})$$

If the wave function at the entrance face is $\psi(x, y, 0)$, instead of a plane wave one has for the wave function at the exit face

$$\psi(x, y, \epsilon) = \{\psi(x, y, 0) \exp[i\sigma V_p(x, y)]\} \exp\left(\frac{i\pi k(x^2 + y^2)}{\epsilon}\right) \quad (\text{A50})$$

This is the Fresnel approximation in which the emerging spherical wavefront is approximated by a paraboloidal wavefront.

The propagation through the vacuum gap from one slice to the next is thus described by a convolution product in which each point source of the previous slice contributes to the wave function in each point of the next slice. The motion of an electron through the whole specimen can now be described by an alternation of phase object transmissions (multiplications) and vacuum propagations (convolutions). In the limit of the slice thickness ϵ tending to zero, this multislice expression converges to the exact solution of the nonrelativistic Schrödinger equation in the forward-scattering approximation.

In the original multislice method one used the Fourier transform of Eq. (A.50) where the real space points (x, y) are transformed into diffracted beams $\mathbf{g}$ and where convolution and normal products are interchanged, that is,

$$\psi(\mathbf{g}, \epsilon) = [\psi(\mathbf{g}, 0) \exp(i\sigma V_g)] \exp\left(\frac{i\pi \mathbf{g}^2 \epsilon}{k}\right) \quad (\text{A51})$$

where V_g are the structure factors (Fourier transforms of the unit-cell potential).

The wave function at the exit face of the crystal can now be obtained by successive application of Eq. (A.50) or (A.51). This can either be done in real space [Eq. (A.50)] or in reciprocal space [Eq. (A.51)]. The major part of the computing time is required for the calculation of the convolution product, which is proportional to N^2 [N is the number of sampling points (real space) or beams (reciprocal space)].

Since the Fourier transform of a convolution product yields a normal product (with calculation time proportional to N) a large gain in speed can be obtained by alternately performing the propagation in reciprocal space and the phase object transmission in real space (23). In this way the computing time is devoted to the Fourier transforms and is proportional to $N \log_2 N$.

Another way of increasing the speed is in the so-called real-space method (24). Here the whole calculation is done in real space using Eq. (A.50) but the forward scattering of the electrons is exploited so as to calculate the convolution effect of the propagation only in a limited number of adjacent sampling points. In this way, the calculation time is proportional to N . This method does not require a periodic crystal and is thus suitable for calculation of crystal defects.

Electron Channeling

The multislice method is an efficient method to compute numerically the exit wave of an object. However it observes interesting physical aspects of dynamical electron scattering. The channelling theory is more approximate (1,25) (although improvements are currently being made) but it is simple and it gives much physical insight.

Electron Wave. Consider an isolated column of atoms, parallel to the electron beam. If we assume that the fast electron in the direction of propagation (z axis) behaves as a classical particle with velocity $v = \hbar k/m$ we can consider the z axis as a time axis with

$$t = \frac{mz}{\hbar k} \quad (\text{A52})$$

Hence we can start from the time-dependent Schrödinger equation

$$-\frac{\hbar}{i} \frac{\partial \psi}{\partial t}(\mathbf{R}, t) = H\psi(\mathbf{R}, t) \quad (\text{A53})$$

with

$$H = -\frac{\hbar^2}{2m} \Delta_{\mathbf{R}} - eU(\mathbf{R}, t) \quad (\text{A54})$$

with $U(\mathbf{R}, t)$ the electrostatic crystal potential, m and $\hbar$ the relativistic electron mass and wavelength, and $\Delta_{\mathbf{R}}$ the Laplacian operator acting in the plane $(\mathbf{R})$ perpendicular to z .

Using Eq. (A.52) we then have

$$\frac{\partial \psi(\mathbf{R}, z)}{\partial z} = \frac{i}{4\pi k} [\Delta_{\mathbf{R}} + V(\mathbf{R}, z)] \psi(\mathbf{R}, z) \quad (\text{A55})$$

with

$$V(\mathbf{R}, z) = \frac{2me}{\hbar^2} U(\mathbf{R}, z) \quad (\text{A56})$$

This is the well-known high-energy equation in real space, which can also be derived from the stationary Schrödinger equation in the forward-scattering approximation (22).

If we now consider the depth proportional to the time, the dynamical equation (A.55) represents the walk of an electron in the two-dimensional projected potential of the columns. The solution can be expanded in eigenfunctions (eigenstates) of the Hamiltonian

$$\psi(\mathbf{R}, z) = \sum_n C_n \phi_n(\mathbf{R}) \exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) \quad (\text{A57})$$

where

$$H\phi_n(\mathbf{R}) = E_n \phi_n(\mathbf{R}) \quad (\text{A58})$$

with the Hamiltonian

$$H = -\frac{\hbar^2}{2m} \Delta_{\mathbf{R}} - eU(\mathbf{R}) \quad (\text{A59})$$

$U(\mathbf{R})$ is the projected potential of the column

$$E = \frac{\hbar^2 k^2}{2m} \quad (\text{A60})$$

E is the incident electron energy, and λ is the electron wavelength. For $E_n < 0$ the eigenstates are bound to the column. We now rewrite Eq. (A.57) as

$$\psi(\mathbf{R}, z) = \sum_n C_n \phi_n(\mathbf{R}) + \sum_n C_n \phi_n(\mathbf{R}) \left[\exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) - 1 \right] \quad (\text{A61})$$

The coefficients C_n are determined from the boundary condition

$$\sum_n C_n \phi_n(\mathbf{R}) = \psi(\mathbf{R}, 0) \quad (\text{A62})$$

In case of plane-wave incidence one thus has

$$\sum_n C_n \phi_n(\mathbf{R}) = 1 \quad (\text{A63})$$

so that

$$\psi(\mathbf{R}, z) = 1 + \sum_n C_n \phi_n(\mathbf{R}) \left[\exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) - 1 \right]$$

Only states will appear in the summation for which

$$|E_n| \geq \frac{E\lambda}{z} \quad (\text{A64})$$

These are bound states with deep energy levels that are localized near the column cores. In practice if the column does not consist of heavy atoms and the distance between columns is not too close (e.g., larger than 0.1 nm) only one eigenstate will appear, which can be compared to the 1s state of an atom.

We then have

$$\psi(\mathbf{R}, z) = 1 + C\phi(\mathbf{R}) \left[\exp\left(-i\pi \frac{E}{E_0} \frac{z}{\lambda}\right) - 1 \right] \quad (\text{A65})$$

A very interesting consequence of this description is that, since the state ϕ is very localized at the atom core, the wave function for the total object can be expressed as a superposition of the individual column functions ϕ_i so that Eq. (A.65) in that case becomes

$$\psi(\mathbf{R}, z) = 1 + \sum_i C_i \phi_i(\mathbf{R} - \mathbf{R}_i) \left[\exp\left(-i\pi \frac{E}{E_0} \frac{z}{\lambda}\right) - 1 \right] \quad (\text{A66})$$

where the summation runs over all the atomic columns of the object parallel to the electron beam.

The interpretation of Eq. (A.66) is simple. Each column i acts as a channel in which the wave function oscillates periodically with depth. The periodicity is related to the “weight” of the column, that is, proportional to the atomic number of the atoms in the column and inversely proportional to their distance along the column. The importance of these results lies in the fact that they describe the dynamical diffraction for larger thicknesses than the usual phase-grating approximation and that they require only the knowledge of one function ϕ_i per column (which can be set in tabular form similar to atom scattering factors or potentials). Furthermore, even in the presence of dynamical scattering, the wave function at the exit face still retains a one-to-one relation with the configuration of columns for perfect crystals as well as for defective crystals provided they consist of columns parallel to the electron beam. Hence this description is very useful for interpreting high-resolution images.

Diffraction Pattern. Fourier transforming the wave function equation (A.66) at the exit face of the object yields the wave function in the diffraction plane, which can be written as

$$\psi(\mathbf{g}, z) = \sum_i \sum_i \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}_i) F_i(\mathbf{g}, t) \quad (\text{A67})$$

In a sense the simple kinematical expression for the diffraction amplitude holds, provided the scattering factor for the atoms is replaced by a dynamical scattering factor for the columns, which is defined by

$$F_i(\mathbf{g}, z) = \left[\exp\left(\frac{-i\pi E_i}{E} \frac{z}{\lambda}\right) - 1 \right] C_i f_i(\mathbf{g}) \quad (\text{A68})$$

with $f_i(\mathbf{g})$ the Fourier transform of $\phi_i(\mathbf{R})$. It is clear that the dynamical scattering factor varies periodically with depth. This periodicity may be different for different columns.

In case of a monatomic crystal, all F_i are identical. Hence $\psi(\mathbf{g}, z)$ varies perfectly periodically with depth. In a sense the electrons are periodically transferred from the central beam to the diffracted beams and back. The periodicity of this dynamical oscillation (which can be compared with the Pendelösung effect) is called the dynamical extinction distance. It has, for instance, been observed in Si(111). An important consequence of Eq. (A.67) is the fact that the diffraction pattern can still be described by a kinematical type of expression so that existing results and techniques (e.g., extinction rules) that have been based on the kinematical theory remain valid to some extent for thicker crystals in zone orientation.

BIBLIOGRAPHY

1. S. Amelinckx et al. (eds.), Electron microscopy, in *Handbook of Microscopy*, Weinheim: VCH, 1997, vol. 1, chap. IV.
2. D. W. Robards, A. J. Wilson, *Procedures in Electron Microscopy*, Chichester: Wiley, 1993.
3. W. L. Bragg, *Nature*, **124**: 125, 1929.
4. P. P. Ewald, *Ann. Phys.*, **54**: 519, 1917.
5. J. Steeds, Convergent beam electron diffraction, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods I*, Weinheim: VCH, 1997, vol. 1, chap. IV.1.5.
6. S. Takagi, *Acta Crystallogr.*, **15**: 1311, 1962.
7. P. B. Hirsch, A. Howie, M. J. Whelan, *Philos. Trans. R. Soc.*, **A 252**: 499, 1960.
8. S. Amelinckx, J. Van Landuyt, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science*, Amsterdam: North Holland, 1978, p. 107.
9. P. B. Hirsch et al., *Electron Microscopy of Thin Crystals*, London: Butterworths, 1965.
10. P. Humble, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science*, Amsterdam: North Holland, 1978, p. 315.
11. D. J. H. Cockayne, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science*, Amsterdam: North Holland, 1978.
12. H. Lichte, Electron holography methods, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods I*, Weinheim: VCH, 1997, chap. IV.1.8.
13. J. Jansen et al., *Acta Crystallogr.*, **A 54**: 91, 1998.
14. H. W. Zandbergen, S. Anderson, J. Jansen, *Science*, **12**: 1221, 1997.
15. T. Nagatani et al., *Scan. Microsc.*, **1**: 901, 1987.
16. T. E. Mulvey, C. D. Newman, *Inst. Phys. Conf. Ser.*, **18**: 16, 1973.
17. K. C. A. Smith, in O. Johari (ed.), *Proc. 5th Annu. SEM Symp.*, Chicago: IITRI, 1972, p. 1.
18. J. I. Goldstein et al., *Scanning Electron Microscopy and X-ray Microanalysis*, New York: Plenum, 1992, Chap. 2.
19. S. Amelinckx et al. (eds.), *Scanning Electron Microscopy*, D. C. Joy, Scanning reflection electron microscopy, in *Handbook of Microscopy, Methods II*, Weinheim: VCH, 1997, chapter IV.2.1.
20. J. Cowley, Scanning transmission electron microscopy, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods II*, Weinheim: VCH, 1997, vol. 1, chap. IV.2.2.
21. S. Pennycook, Scanning transmission electron microscopy, Z contrast, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods II*, Weinheim: VCH, 1997, Vol. 1, Chap. IV.2.3.
22. A. V. Crewe, J. Wall, J. Langmore, *Science*, **168**: 1333, 1970.
23. D. Van Dyck, in P. Hawkes (ed.), *Advances in Electronics and Electron Physics*, New York: Academic, 1985.
24. J. M. Cowley, A. F. Moodie, *Acta Crystallogr.*, **10**: 609, 1957.
25. K. Ishizuka, N. Uyeda, *Acta Crystallogr. A* **33**: 740, 1977.
26. K. Kambe, G. Lempfuhr, F. Fujimoto, *Z. Naturforsch.*, **29a**: 1034, 1974.

D. VAN DYCK
S. AMELINCKX
University of Antwerp

X-RAY MICROSCOPY

X-ray microscopy is not a new field, with the first work dating back to the beginning of the twentieth century. The first instruments were simple indirect transmission microscopes. A common technique was contact microradiography and involved placing a photographic film in contact with a sample and then illuminating the sample with X rays to produce a shadowgraph. The image was then magnified using an optical microscope. This technique, when modified to allow projection of the object onto the film instead of having the film in contact with the sample, produced a 0.1 μm resolution by the 1950s. Contact microscopy, in which the film is replaced by a photosensitive material, was the next natural development. Both classical X-ray microscopy techniques continue to be used in medical and biological research and have produced some spectacular images of samples ranging from chromosomes to brain tissue (1,60,61).

The question may arise then why a 0.1 μm resolution for modern day X-ray microscopes (XRMs) should be counted as an important achievement. The reason is that today's XRMs are capable of doing many more types of experiments than the previous generation microscopes could. Although they lost ground to the high resolution scanning electron microscope (SEM) in the 1960s and later to the atomic resolution scanning tunneling microscope (STM) in the 1980s, XRMs have recaptured the interest of the scientific community because they are capable of unique experiments not possible with other microscopes. High intensity broad-band X rays generated as synchrotron radiation (62,63), improved optics and instrumentation, the ability to study radiation-sensitive and wet samples with soft X rays, and the ability to do spectroscopic imaging make today's X-ray microscopes extremely versatile (64–67). The 0.1 μm resolution record has also been broken by several X-ray microscopes.

Many interesting effects—both basic and technological—are observed in the dimensional range of approximately 50 nm to 5 μm . This region bridges the nanostructure domain with the macroscopic world and contains aspects of both. A typical

case is that of integrated circuit (IC) development, where today the critical dimension (CD) is approaching $0.1\ \mu\text{m}$. X-ray microscopes are ideally suited to the study of systems in that size range, from untreated biological samples to semiconductor heterostructures. They can yield information about local variations in chemical composition and electronic structure through spatially resolved absorption, fluorescence, and photoelectron spectra. One speaks often of *spectromicroscopes*, not just microscopes. The word *spectromicroscopy* combines aspects of spectroscopy and microscopy as does the word *microspectroscopy*. However, it is the last four syllables of either expression that describe the thrust of either technique. Thus, spectromicroscopic measurements entail taking images containing partial spectral information of small areas. Typically a larger energy band pass is selected to allow for a sufficiently high count rate to allow image acquisition in reasonable time. Microspectroscopic measurements involve acquiring spectroscopic data with partial spatial information. In this incarnation it is common to collect a high resolution spectrum from an area of the image by either counting for a longer time or using a larger area. The two techniques are frequently used in unison. Given a high X ray flux, it is even possible to use both techniques simultaneously, collecting an energy spectrum for each pixel of the image.

In 1994, Moewes catalogued the world's VUV and X-ray microscopes, listing 42 such instruments, 25 of which use synchrotron radiation (2). The extreme vacuum ultraviolet (VUV) region extends from $2\ \text{eV}$ to $1000\ \text{eV}$ ($12\ \text{eV}$ to $6200\ \text{eV}$). Of this, the range from $20\ \text{eV}$ to $500\ \text{eV}$ ($25\ \text{eV}$ to $620\ \text{eV}$) is often called the soft X-ray region. In the hard X-ray region, the wavelength ranges from $0.5\ \text{nm}$ to $2\ \text{nm}$ (the energy from $6.2\ \text{keV}$ to $25\ \text{keV}$) (3). Since then, however, more X-ray microscopes have come into existence, in particular at new third-generation synchrotrons such as the Advanced Light Source of the Lawrence Berkeley National Laboratory and the Sincrotrone Trieste in Trieste, Italy. Spectromicroscopy programs are in the genesis stage at budding synchrotrons such as the Advanced Photon Source in Illinois and the Synchrotron Radiation Research Center in Hsinchu, Taiwan. Any of these light sources is a site of choice for building a X-ray spectromicroscope because of the high brightness available. There are also efforts to use plasma X-ray sources for microscopy (4). These sources are inherently limited in light output, but they do allow experiments to be done away from synchrotrons.

Most modern X-ray microscopes, though varying widely in their optics and signal collected, can be divided into two broad categories: *imaging* and *scanning*. Both types, in their transmission and emission incarnations, will be described in this article. Indirect techniques such as Fourier-transform X-ray holography will not be discussed.

INTERACTION OF X RAYS WITH MATTER

It would not be appropriate to treat in depth here a topic on which books have been written. We will confine our discussion of the interaction of X rays with matter to the cases that directly concern the operation of X-ray spectromicroscopes, with an emphasis on photoemission in solids.

Surface Sensitivity

Nondiffractive optics for X rays only work at glancing angles. This fact follows from the calculation of the reflectivity of light

at an interface using the Fresnel equations. Only a glancing angle produces a significant reflectivity for X-ray wavelengths. X rays are not simply reflected off the top surface of the material under study. In fact, it is important to know how far the X rays penetrate into the sample in order to understand what fraction of the sample contributes to the signal collected, and thus, what part of the sample we are really studying. This can be arrived at through consideration of the absorption coefficients of the materials composing the sample. The absorption coefficient α is a function of the photon energy [small h , stroke] ω , and the intensity at a distance z inside the material can be written as

$$I = I_0 e^{-\alpha z} \quad (1)$$

where I_0 is the intensity incident on the surface. The photon attenuation length λ_{ph} (also called the photon mean free path) is $1/\alpha$ and is commonly used as a measure of how far the photons penetrate into the sample (although the intensity at that distance is about 37% of the incident intensity). λ_{ph} is a function of photon energy and varies for different substances because of differences in material characteristics. For instance, at a photon energy of $100\ \text{eV}$, λ_{ph} is $0.18\ \mu\text{m}$ for carbon and $0.11\ \mu\text{m}$ for molybdenum, yet at $1000\ \text{eV}$ it is $1.9\ \mu\text{m}$ for carbon and $0.19\ \mu\text{m}$ for molybdenum, and at $5000\ \text{eV}$ it is $234.1\ \mu\text{m}$ for carbon and $1.8\ \mu\text{m}$ for molybdenum. Clearly, for transmission X-ray microscopes rather thick samples may be used, with local differences in the absorption coefficient of the sample providing a contrast mechanism. Photoemission microscopes rely on the photoelectrons ejected from the sample for their signal. As illustrated in Fig. 1, photoelectrons (and Auger electrons) are excited everywhere the X rays can reach in the sample, but because of electron scattering, only those electrons excited within the (isotropic) electron mean free path λ_e of the sample surface can escape from the sample. This is why photoemission spectroscopy (PES) is a *surface-sensitive* technique. λ_e , a function of the electron kinetic energy (approximately proportional to the

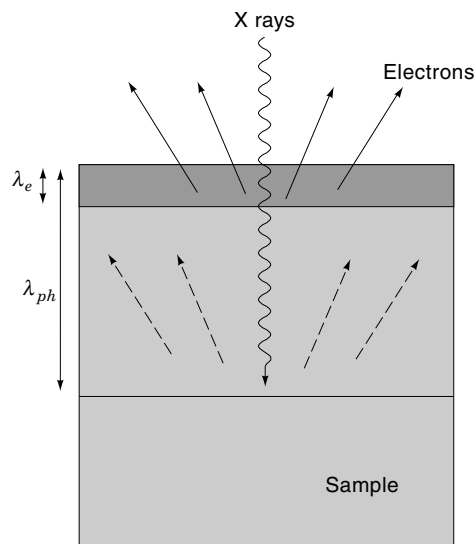


Figure 1. Schematic showing the surface sensitivity of photoemission. Although X rays penetrate to a depth of λ_{ph} in the sample, only those photoelectrons excited within the electron escape depth λ_e are emitted from the sample.

square root of the kinetic energy) that also varies for different materials, varies from about 5 eV to 100 eV for kinetic energies from 2 eV to 130 eV , (5), a range of interest for the University of Wisconsin spectromicroscope MAXIMUM. Because of this ability to sample only the top layers of a material, PES is an ideal technique for the study of surfaces and interfaces. It is also possible to design a microscope that collects the photoelectrons emitted from the opposite side of a thin sample that is illuminated with X rays, thus combining transmission and emission X-ray microscopy (6).

Photoemission

To explain the photoemission process, it is useful to examine the band structure of a solid. A photon of energy $[h\omega]$ excites an electron in a valence band state of energy E_i into an unoccupied state of energy E_f in the conduction band and leaves behind a hole. We have

$$E_f = E_i + \hbar\omega \quad (2)$$

If E_f is sufficient to overcome the material's photoionization potential or work function ϕ , the electron energy is above the vacuum level E_{vac} given by $E_v + \phi$, where E_v is the energy of the top of the valence band (and is identical to the Fermi energy E_F in a metal), the photoelectron can escape into vacuum. This electron is said to have been *photoemitted*. Subtracting E_{vac} from Eq. (2), we obtain

$$E_f - E_{\text{vac}} = \hbar\omega - (E_v + \phi) \quad (3)$$

$$= \hbar\omega - (E_v - E_i) - \phi \quad (4)$$

Because $E_f - E_{\text{vac}}$ is the photoelectron kinetic energy E_{kin} and $E_v - E_i$ is the binding energy E_b of the electron in the solid, we may rewrite Eq. (4) as

$$E_{\text{kin}} = \hbar\omega - E_b - \phi \quad (5)$$

E_{kin} is the value measured in a photoemission experiment. In an experimental system, assuming that the electron energy analyzer and the sample are in equilibrium, ϕ in Eq. (5) will be the work function of the analyzer, which is typically higher than the sample work function. (Because the analyzer's vacuum level is not the same as the sample's (although the Fermi level is), it can only detect electrons with a kinetic energy above its own threshold). This difference also leads to the secondary electron peak being cut off at its lower energy end.

Not all transitions from E_i to E_f are allowed. In the absence of spin-orbit interaction, only those transitions that obey the dipole selection rules

$$\Delta l = \pm 1 \quad (6)$$

$$\Delta m = 0 \pm 1 \quad (7)$$

are permitted, where l is the azimuthal or orbital angular momentum quantum number and can take on the values $0, 1, 2, \dots, n-1$ (where n is the principal quantum number), and m is the magnetic quantum number and ranges from $-l$ to l . If a spin-orbit intersection exists, as it does for metals, the total angular momentum quantum number $j = l \pm 1/2$ gives rise to a splitting in the binding energy. Such a split for the

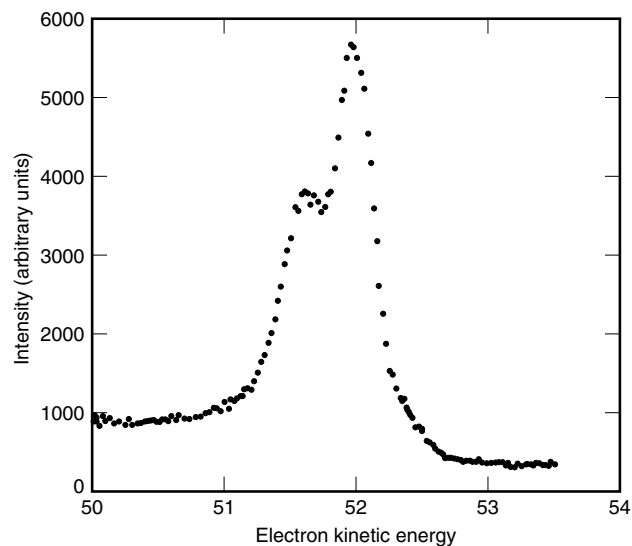


Figure 2. EDC from MAXIMUM showing spin-orbit splitting of the Al $2p$ core level.

Al $2p$ core level can be seen in the electron energy distribution curve (EDC) in Fig. 2. In EDCs there is also a large diffuse peak near 0 eV kinetic energy, which is the secondary electron background and results from electrons that have lost energy through inelastic scattering from other electrons. Core levels of low kinetic energy will be seen riding on top of this background.

See also the article Photoemission in this Encyclopedia.

Photoemission: Applications. There are various experimental techniques that use photoemission. Examples include angle-resolved photoemission spectroscopy (ARPES), extended X-ray absorption fine structure (EXAFS) and its cousins SEXAFS (S for surface) and NEXAFS (N for near edge), and, of course, electron spectroscopy for chemical analysis (ESCA) and X-ray photoelectron spectroscopy (XPS). ESCA and XPS are essentially the same physical technique, as are EXAFS and NEXAFS, with the difference in nomenclature arising from the application. We shall refer to them as *X-ray photoemission* and *X-ray absorption*, respectively. In the former, the occupied valence band and deeper states are probed by using a fixed photon energy and examining the distribution of photoelectron kinetic energies (the EDC). In the latter, unoccupied conduction band states are explored by scanning the photon energy and collecting the photoelectrons in total yield mode as a function of photon energy. In both photoemission and photoabsorption modes the aim of spectromicroscopy is to distinguish between chemical states on a microscopic scale. This can be done by studying chemical shifts of electronic levels corresponding to specific materials comprising the sample.

Chemical Shifts. Core level electrons are tightly bound to their atoms and thus serve well as a means of elemental identification. Variations in local chemistry induce local charging and hence *shifts* in the core level binding energies, which are reflected in the photoelectron spectra. This happens because of changes in the screening of the nucleus by the valence electrons. For example, if the atom of interest is in a chemical environment that results in additional valence electrons being drawn to the atom, Coulomb screening will increase and the binding

energy of the core level will be decreased. If, however, the atom is bonded to a highly electronegative material, its oxidation state will change and the number of valence electrons screening the core will be decreased, resulting in a corresponding increase in the core level binding energy. The effects on binding energy are more complicated in the case of nonionic bonding. Shifts from pure elemental values have been observed for several compound materials. The amount and sign of the shift can be quantified in a straightforward way for the chemical shift $\Delta E_n(A, B)$ for atom n in two different chemical environments A and B by the following equation (7):

$$\Delta E_n(A, B) = \left(\frac{q_n^A}{\bar{r}_{v,n}^A} - \frac{q_n^B}{\bar{r}_{v,n}^B} \right) + (V_n^A - V_n^B) \quad (8)$$

$\bar{r}_{v,n}$ is the average radius of the valence shell of the n th atom, and the q_n are screening charges. Thus the first term is the difference in screening potential between environments A and B . The second term is the difference in the Madelung potentials of A and B , where for V_n we may write

$$V_n = \sum_{n' \neq n} \frac{q_n}{R_{nn'}} \quad (9)$$

Here we sum over all surrounding ionic charges q_n centered at positions $R_{nn'}$ measured with respect to the atom n . Equation (8) has been shown to be a simple but good predictor of chemical shifts in solids.

Other Spectral Features. Photoelectron spectra exhibit features not solely ascribable to primary and secondary electrons. The lineshape, for example, is not a delta function, but is broadened. Taking into account the finite lifetime of the hole created by photoexcitation of a core level electron, the theoretical lineshape should be a Lorentzian with width inversely proportional to the core lifetime. There is also a further broadening caused by electron-phonon scattering. In real life, the shape is not a Lorentzian, but is convoluted with a Gaussian arising from instrumental limits such as detector energy resolution, monochromator resolution, the Doppler effect, and thermal broadening. The resulting lineshape is called a Voigt function and is extensively used in spectroscopy.

The hole can decay via radiative or Auger recombination. In the former case, an electron from a higher energy level drops down into the vacancy, and radiation with an energy corresponding to the difference between the levels is emitted and is called fluorescence. Luminescence refers to visible or UV light emitted from valence transitions. Both kinds of light can be used as chemical signatures for spectromicroscopy, as they are for two of the X-ray microscopes discussed in this article. This photon emission that accompanies photoemission, though it is not as surface-sensitive, can be exploited in a manner analogous to energy dispersive analysis of X rays (EDAX) in a scanning electron microscope (SEM) to acquire images mapping elements on the sample.

In the Auger process, the energy difference is used to ionize a higher-level electron. This Auger electron is ejected with a kinetic energy characteristic of the material in question. There is a finite set of permitted Auger transitions for a given element, and Auger peaks are independent of photon energy under non-resonance conditions.

Plasmons occur in metals in a series, with the plasmon peaks shifted from the core peak to lower kinetic (higher binding) energies by multiples of [small h, stroke] ω_p , the plasma frequency, if $\omega > \omega_p$. Thus plasmons are a form of energy loss for primary electrons. The conduction band electrons in a metal form an electron gas that can be set into collective oscillation by an excited photoelectron. A plasmon represents one quantum of this oscillation.

Excitons can be formed in the bandgap of semiconductors when a core hole and an electron pair up. For example, if the $2p$ level is excited in silicon, an exciton is formed that has a binding energy of about 0.9 eV that places it in the gap. It shows up in the photoabsorption spectra.

IMAGING MICROSCOPES

Imaging X-ray microscopes are somewhat similar in optical design to traditional visible-light microscopes in that an image of an object is formed solely with the aid of lenses operating on the detector side. Depending on the detection method, these instruments operate in either (X-ray) transmission or (electron) emission mode. Both types will be described below. Some use specialized X-ray optics, whereas others use electron optics to image the sample.

Transmission

The sample is illuminated with X rays at the region of interest, and the transmitted X rays are used to produce a magnified image of the sample. An example of such an instrument is that developed by Meyer-Ilse at the Advanced Light Source (ALS) of the Lawrence Berkeley National Laboratory (LBNL) (8). Shown in Fig. 3, it uses a condensing Fresnel zone plate (FZP), a diffractive optic similar to a circular diffraction grating whose pitch increased with radial distance from the center. This condensing FZP focuses the X rays onto a 15 μm pinhole, after which they emerge from vacuum into a small space filled with helium gas, where the sample is mounted. Helium is used instead of air or another gas because it has low absorption in the X-ray regime. The X rays pass through the samples (which are typically cells, tissues, or microorganisms encased in a SiNH window in their growing solution), where they are scattered, and then enter another vacuum chamber, where an objective FZP forms an enlarged image of the sample and projects it onto

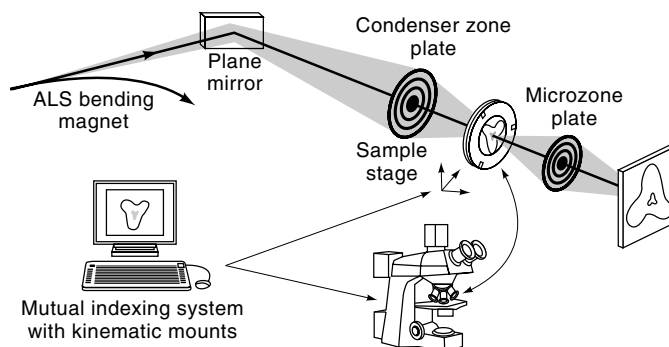


Figure 3. Schematic of the XM-1 transmission X-ray microscope at the ALS of LBNL. (Courtesy of W. Meyer-Ilse.)

a CCD camera. This instrument has demonstrated a resolution of 43 nm (9). A similar X-ray microscope at the Aarhus electron storage ring in Denmark has been used to study systems as diverse as human spermatozoa at various stages of maturation and iron-precipitating bacteria used for water treatment (10).

Emission

Emission mode microscopes directly image the electrons (primary, secondary, or Auger electrons) ejected from the sample to produce a magnified image of it. The photon energy can be varied to take advantage of variations in the absorption spectra of materials. The incident X rays are coarsely focused onto the sample, chiefly to uniformly illuminate the region of interest and to minimize flux loss and its resultant degradation of the signal-to-noise ratio. The spatial resolution is limited by the electron optics and inherent aberrations.

Two imaging photoelectron microscopes will be discussed below. One uses electrostatic optics, and the other uses electromagnetic optics.

Photoemission Electron Microscope. The photoemission electron microscope (PEEM) uses electrostatic lenses to form an image of the sample with photoelectrons. An immersion lens is used as an objective to focus the electrons. Typically, it consists of a primary electrode at high voltage to accelerate the electrons, and a three-electrode lens that forms a real image. The sample itself serves as the cathode. The real image can then be magnified further using an intermediate objective lens. Next, it can be projected onto an assembly of a microchannel plate (MCP) and a phosphor screen, using a projector lens made up of another three-electrode lens and a two-stage decelerating lens designed to lower the kinetic energy of the electrons to the sensitivity range of the MCP (11,12).

The resolution of the objective lens is approximated by

$$\delta = \epsilon K_0 / |E| \quad (10)$$

where K_0 is the mean photoelectron kinetic energy and E is the electric field at the sample surface, and ϵ is a dimensionless constant of order unity (13). A typical resolution for a traditional PEEM (UV-PEEM) is several micrometers, but a resolution of 0.9 μm has been achieved (14). With very soft X rays, a resolution of 10 nm has been obtained using very low-kinetic-energy electrons (15). The theoretical total resolution is limited by chromatic aberration (since photoelectrons of different energies are focused at different positions) and residual spherical aberration in the objective lens itself, as well as by the aberrations arising from the high accelerating potential between the sample and the objective lens (16).

Improvements in the electron optics design can overcome some of these limitations. For example, a large accelerating potential of ≈ 20 kV can be used to decrease the angular spread of the photoelectrons, and to thus reduce the spherical aberration (which is proportional to the cube of the spread angle) and improve the resolution (17). The insertion of an aperture at the back focal plane of the objective lens also reduces chromatic aberration by blocking electrons that would not be focused exactly at the image plane. Smaller apertures improve spatial resolution more than larger apertures, but also reduce the signal level. Using a 20 μm aperture, De Stasio and coworkers

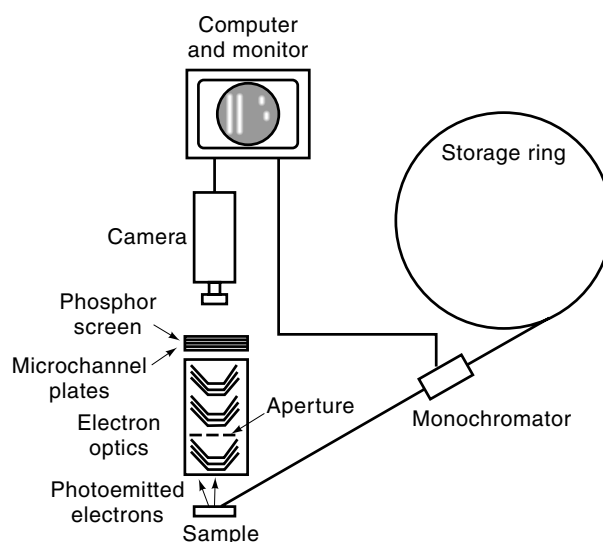


Figure 4. Schematic of the MEPHISTO photoemission electron microscope. (Courtesy of G. De Stasio.)

have demonstrated a 50 nm resolution using the Rayleigh criterion on a knife-edge (12% to 88% of full intensity) with their microscope MEPHISTO (18), which is shown in Fig. 4. This resolution is being improved by several microscopy groups even as this article is being written. It is also possible to add an energy-filtering detector to further reduce the chromatic aberration and to allow the extraction of more spectroscopic information by implementing photoelectron energy selection and analysis. In the case of the instrument PRISM, this was done by inserting an electrostatic "prism" energy filter between the intermediate and projector lenses. This "prism" consists of a sector of a hemispherical capacitor used in conjunction with circular apertures and coupling lenses at the input and output of the capacitor. The lenses preserve the image, and the capacitor behaves as a spherical mirror analyzer to provide the bandpass electron energy filtering. The ability to select the photoelectron energy also makes possible the acquisition of photoelectron energy distribution curves (EDCs). PRISM has achieved a spatial resolution of 0.6 μm in energy-filtering mode on a 25 μm pitch gold grid; the Au 4*f* core level could be measured with an energy resolution of 1.0 eV and was used for chemical imaging. Currently, PRISM is operating without the energy filter, that is, as a standard X-ray PEEM (X-PEEM). To date, PRISM has achieved a resolution of 0.25 μm in X-PEEM mode (19).

PRISM in X-PEEM mode was used to study the formation of titanium silicide (TiSi₂) in a patterned sample with microstructures ranging from 100 μm to 0.1 μm . TiSi₂ is an important material used for interconnects in ultra large-scale integration (ULSI) devices. It occurs in two phases: the high-resistivity (60 $\mu\Omega \cdot \text{cm}$ to 90 $\mu\Omega \cdot \text{cm}$) metastable C49 phase, and the desirable low-resistivity (12 $\mu\Omega \cdot \text{cm}$ to 15 $\mu\Omega \cdot \text{cm}$) stable C54 phase, into which C49 is transformed with higher-temperature annealing. The formation of C54 is, however, inhibited for sub-micron lines, leading to a rise in resistivity. This poses a problem since TiSi₂ is used to reduce contact resistance in modern MOS (metal oxide semiconductor) devices, which are now being fabricated with sub 0.25 μm linewidths. Since the trend is to

shrink linewidths (and, hence, devices) even further, the inhibition of the formation of the low resistivity phase of TiSi_2 at small linewidths is an important issue.

C54 has a face-centered orthorhombic crystal structure, while C49 has a body-centered orthorhombic crystal structure with different lattice constants. This leads to different bonding between Ti and Si in the two phases, and hence we should expect the chemical shift between core electron binding energies in C49 and C54 to be reflected in different photoelectron spectra. As seen in Fig. 5, photoelectron images acquired on the Ti $L_{2,3}$ absorption edge show dramatic differences at different photon energies, with the spoke tips and edges appearing brighter than the spoke centers at a photon energy of 455 eV. This suggests a local difference in chemical status of titanium silicide. In fact, absorption spectra taken on regions 1–4 as defined in Fig. 6(b) show that spectra acquired on a spoke tip and a spoke edge (regions 2 and 3) have a higher intensity in the 454 eV to 457 eV energy range than the spectra acquired on blanket silicide or on a spoke center (regions 1 and 4). (See Fig. 7.) This difference is consistent with C54 TiSi_2 being formed in spoke centers and C49 TiSi_2 remaining untransformed not only at spoke tips but also at spoke edges (20,21). This was the first successful application of spectromicroscopy to the study of TiSi_2 formation and revealed that C54 formation is inhibited not only in small features but at feature edges as well, suggesting a dependence on geometry that has been borne out by other experiments (22).

Other experiments facilitated by PEEM include mapping of domains on magnetic recording disks using magnetic circular X-ray dichroism (MCXD) (23,24) and spatial and temporal waves of oxidizing CO on platinum (25,26).

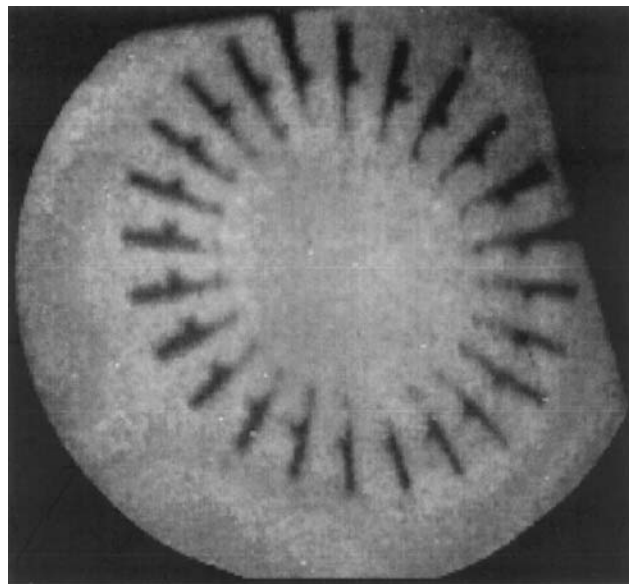
Magnetic Imaging Photoelectron Microscope. The MicroESCA manufactured by Surface Science Instruments is a commercial version of this type of X-ray microscope.

A superconducting (typically NbTi) solenoid is used to bathe the sample in a uniform magnetic field that diverges from the sample normal. The photoemitted electrons spiral about the diverging magnetic field lines and are projected on a resistive anode detector to form an image magnified by a factor of about 70. A supplementary magnification or demagnification is attained by using two field coils around the detector to modify the solenoid's magnetic field near the detector. This feature results in a possible magnification ranging from $20\times$ to $140\times$. An electron-retarding-grid energy analyzer placed before the detector acts as a high-pass filter and allows the acquisition of photoelectron spectra. Because the electrons are collimated along the field lines, there is little loss of signal at the analyzer as long as it intercepts enough of the field. This collimation results in the additional benefit of a high (0.1 eV) energy resolution for 100 eV photoelectrons. Such a resolution is well suited to the detection of subtle chemical shifts (27).

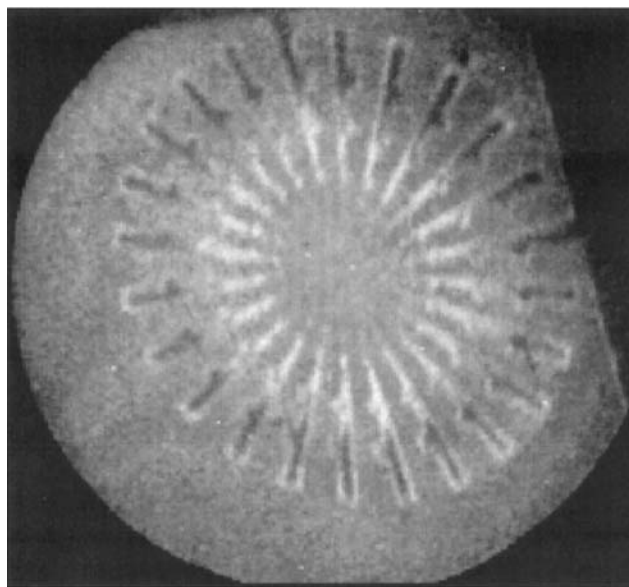
The spatial resolution is proportional to the radius of the helical path traveled by the electrons. For off-axis photoelectrons, this is the cyclotron radius:

$$R_{\text{max}} = (2mE_0)^{1/2} / eB_0 \quad (11)$$

where E_0 is the photoelectron kinetic energy and B_0 is the magnetic field strength at the sample. B_0 can range up to 8 T. The



(a)



(b)

Figure 5. Photoelectron images of silicided-Ti-on-Si resolution test pattern. In (a), acquired below the Ti $L_{2,3}$ edge at 445.5 eV, only the centers of the spokes show up, and these are dark (absorbing). In (b), acquired at 455 eV on a shoulder on the edge, the spoke tips and edges emit much more strongly than the spoke centers, indicating a change in bonding.

spatial resolution is estimated to range between $0.1R_{\text{max}}$ and $2R_{\text{max}}$. For electrons of kinetic energies from 10 eV to 1000 eV, the practical resolution ranges from $3\text{ }\mu\text{m}$ to $30\text{ }\mu\text{m}$ (28,29).

The commercial MicroESCA has seen various applications, ranging from geological samples to heterostructures. One experiment with Au/GaAs showed variations in the surface segregation of As as a function of gold distribution (29). One of the original research instruments is currently operating at

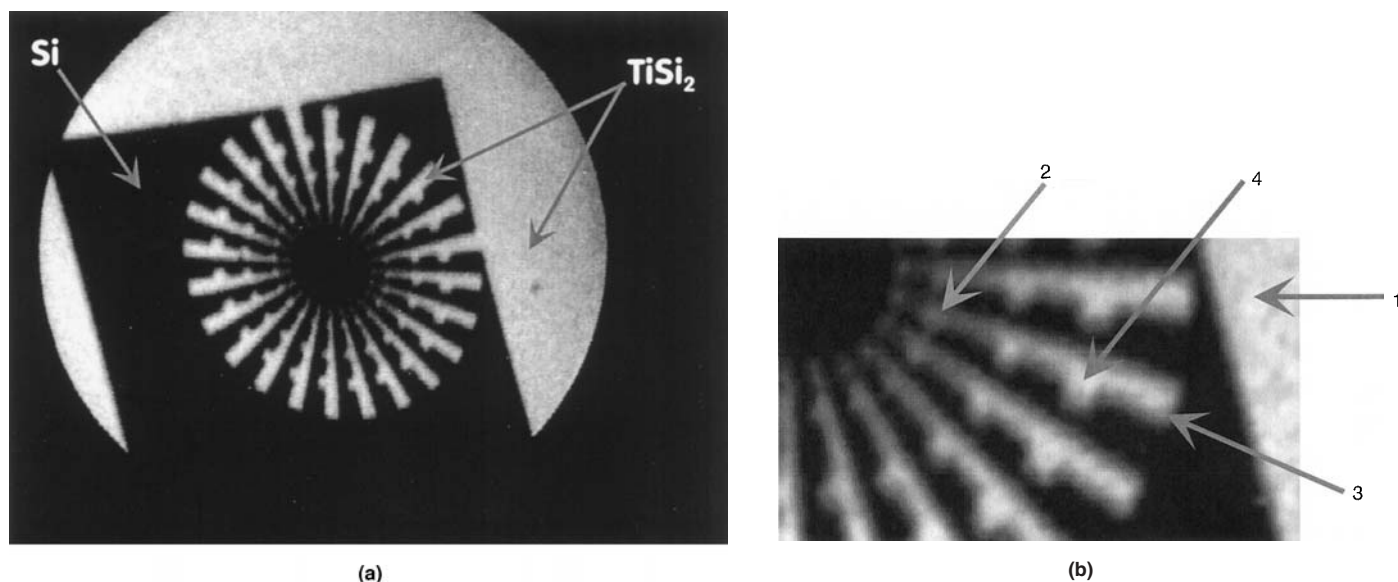


Figure 6. Photoelectron image of silicided STAR acquired at 464 eV. The width of the entire spoke pattern in (a) is 80 μm , and the spoke width at the points where it touches the edge of the square is 5 μm . A magnified view is shown in (b), where region 1 is on a blanket TiSi₂ area outside the pattern, 2 is on a spoke tip, 3 is on a spoke edge, and 4 is in the center of a spoke.

the Synchrotron Radiation Center (SRC) of the University of Wisconsin—Madison and has been used for spectromicroscopic study of spatial variation in the oxidation of Bi in a high- T_c Bi—Sr—Ca—CuO superconductor (30).

SCANNING MICROSCOPES

Scanning X-ray microscopes finely focus the X rays to a micrometer- to submicrometer-sized spot on the sample. This spot is then rastered across the sample, with a signal collected for each (x, y) point. The scanning is most simply achieved

by rastering the sample. The success of scanning X-ray microscopy is dependent on the quality and efficiency of the X-ray optics used for focusing the X rays to a point. Scanning X-ray microscopes exist in both transmission and emission flavors.

Transmission

The scanning transmission X-ray microscope (STXM) is the primary microscope of this type in existence. Developed by Schmahl (31) for use on the ACO ring in France, this type of microscope was also implemented by Kirz (32) at the NSLS of

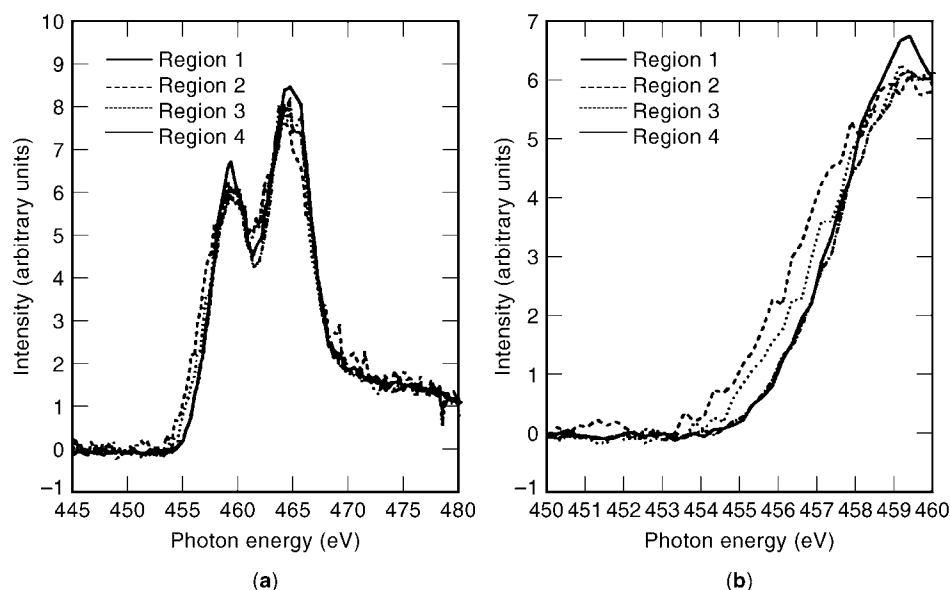


Figure 7. Ti $L_{2,3}$ absorption spectra acquired on regions 1–4 as defined in Fig. 6(b) are shown in (a). A closeup of the Ti L_3 edge in (b) shows that the tips (as in region 2) and the edges (as in region 3) have a higher intensity in the range between 454 eV and 457 eV that is not present in the blanket silicide or at the centers of the spokes, suggesting a difference in bonding in narrow features and feature edges.

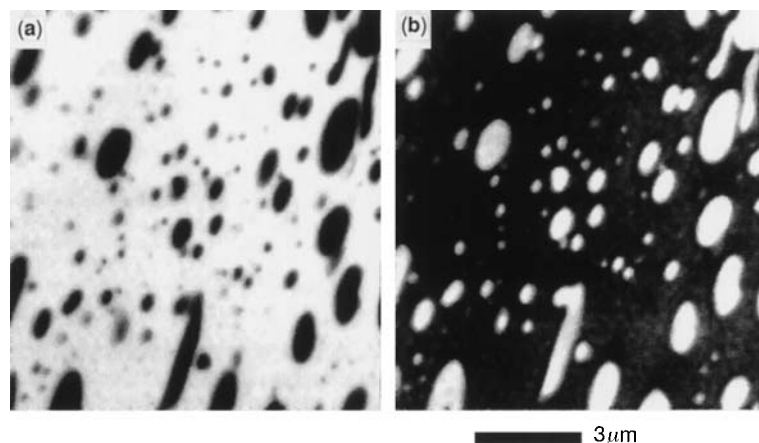


Figure 8. Transmission X-ray micrographs of a PC-PET polymer blend. (a) was acquired at a photon energy of 284.8 eV (energy A), and (b) at 285.1 eV (energy B). PET appears bright in (a), and PC in (b). (Courtesy of H. Ade.)

Brookhaven National Laboratory. It too uses a FZP to focus the X rays onto the sample, which is rastered using a scanning stage to obtain an image. The FZP was fabricated at IBM and has a focal length of 1 mm at a wavelength of 3.6 nm. Using the Raleigh criterion discussed below under “Diffraction-Limited Optical System,” the resolution δ of the first order (which is the most intense) is given by

$$\delta = 1.22 \Delta r \quad (12)$$

where Δr is the width of the outermost zone of the FZP. A central stop is used to block higher diffraction orders, and a pinhole is placed between the FZP and the sample to eliminate unfocused light. A wire proportional counter is used to measure the transmitted X-ray signal. The instrument operates in an atmosphere of helium (33). Biological studies, including those of live and wet specimens, are the primary application of this microscope, as they are for Meyer–Ilse’s imaging microscope described above in the subsection “Photoemission.”

Energies in the water window (between the carbon and oxygen K edges) are well suited because of the transparency of water and the absorptivity of carbon (in biological structures) in this range. Several experiments on chromosomes have been done (32,34). The STXM was later applied to chemical imaging using NEXAFS signals of other systems such as diagonally sectioned Kevlar fibers, which exhibit linear dichroism (35). This instrument has also been used to study a polycarbonate–polyethyleneterephthalate (PC–PET) polymer blend. By varying the incident photon energy, it is possible to track an electronic resonance between carbonyl groups and the phenylene in the polymers. Figure 8 shows X-ray micrographs of the sample acquired at 284.8 eV and 285.1 eV (36). The contrast is reversed in the two images, with the PET bright and PC dark at 284.8 eV and the PC bright and PET dark at 285.1 eV. The shift in the NEXAFS microspectra shown in Fig. 9, which were acquired on both PC and PET areas of the sample is responsible for this contrast. The NSLS STXM has achieved submicrometer resolution. A newer instrument installed at the ALS has attained a resolution of 0.12 μm (37).

Emission

There are several advantages the scanning emission microscope has over the imaging emission microscope, including (1)

the possibility of detecting transmitted X rays and luminescence signals in addition to photoelectron signals, (2) decoupling of the imaging mechanism from the signal collected, (3) a single optical element whose aberrations can be minimized, (4) a standard detector and its accompanying ease of obtaining spectra, and (5) a higher theoretical spatial resolution. Several different approaches will be discussed below.

Scanning Photoemission Microscope. The FZP-based design for the scanning photoemission microscope (SPEM) arose from the STXM discussed above in the subsection “Transmission.” It is a UHV instrument and is typically operated in the energy range of 650 eV to 680 eV with a resolution of 2 eV in order to maximize flux at the sample. At these wavelengths (1.8 nm to 1.9 nm) the focal length is about 5 mm, but the pinhole needed to block unfocused light needs to be placed about three-fourths of the focal length away from the FZP before the sample. The working distance, therefore, is less than 1.5 mm, and part of the microscope must be placed in the cylindrical mirror analyzer (CMA) used for electron energy analysis. The CMA has been modified to accommodate the FZP, the pinhole, and a tilt assembly to align the two. Both the sample current and the CMA signal can be used to acquire images, and the CMA can be used to obtain EDC as well. This instrument has reached submicrometer resolution and has been

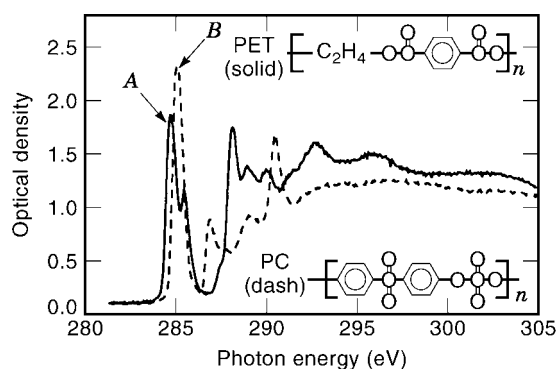


Figure 9. NEXAFS spectra from 0.1 μm^2 areas of PC and PET from a 70:30 PC:PET blend. PET and PC spectra have peaks at energies A and B, respectively. This explains the contrast observed in (a) and (b) of the previous figure. (Courtesy of H. Ade.)

used in test experiments on a commercial device at the NSLS (38). A SPEM, which achieves scanning by moving the zone plate instead of the sample stage, thus facilitating in situ sample treatment such as heating or cooling, is operational at the ALS (68).

Another SPEM is already working at the third-generation ring in Trieste in the energy range of 300 eV to 400 eV. This microscope has submicrometer resolution and is being used for surface studies of metals and semiconductors (39). A high-energy SPEM is in place at the Advanced Photon Source (APS) of Argonne National Laboratory. It uses blazed phase-shifting FZPs, which can theoretically achieve an efficiency of 80%, compared to 10% for ordinary FZPs (40). This microscope can focus hard X rays down to a 0.25 micrometer spot. This spot can then be scanned across the sample while collecting the X-ray fluorescence signal, thus obtaining a chemical map of elements present in very dilute concentrations in the sample. Because of the high-energy X rays used, the experiments can be carried out in air. Biological samples such as plant roots from *Plantago* and *Arabidopsis* have been studied. The X-ray microprobe has also been used for strain measurements using microdiffraction techniques (41–43).

Hamburg Microscope. Designed by Christoph Kunz, this instrument uses a ring ellipsoid mirror to focus monochromatized X rays from wiggler radiation onto the sample. Unlike the FZP, the ellipsoid mirror's focal position is independent of the wavelength, allowing easy switching between wavelengths of interest without moving the sample or the optics to stay at the focus. Its efficiency is also higher because it operates at grazing incidence. In practice, the flux is limited because of problems in illuminating the mirror properly without loss. It is installed on the FLIPPER monochromator at HASYLAB of DESY, Hamburg. The source is situated at one focus of the ellipsoid, and the sample is placed at the other focus. A resolution of 0.4 μm has been achieved at 200 eV. The instrument can also operate at lower resolutions over the entire range of 20 eV to 1300 eV. Mirror quality is the obstacle in reaching higher resolutions. Since this is a grazing-incidence optical system, figure error is very important. The ellipsoidal mirror, custom manufactured by Carl Zeiss, is, like other aspheric mirrors, difficult to make. Because the tolerances for microscopy are so tight,

this limitation becomes significant. Surface finish is important too, since surface roughness leads to scattering and hence loss of useful flux at the sample.

EDCs can be acquired with this instrument, as can photoelectron images at different energies. In addition, the total luminescence signal can be collected, thus allowing a luminescence image to be acquired simultaneously with a photoelectron image. This technique has been applied to the study of transition metal compounds, including variously doped LaOBr powders and chemical Tl^{1+} -doped CsI (3).

Wolter Scanning Photoemission Microscope. The Wolter mirror is a glancing-optics objective consisting of an ellipsoidal section next to a section of a hyperboloid of revolution. The addition of the hyperboloid section corrects the aberrations of a single ellipsoid surface and increases the field of view. The application of these optics to X-ray microscopy was led by Hitachi, Ltd. and the University of Tsukuba, Japan. Like the ellipsoid mirror used in the Hamburg microscope, the Wolter mirror has a focus independent of the wavelength. It has less aberration than other grazing-incidence mirrors, but, since it too is aspheric, it also suffers from the same difficulties in accurate fabrication as the ellipsoid mirror, although the manufacturing techniques are different. (For the Wolter mirror, a vacuum replication process using epoxy resin and a precision-ground and -polished mandrel is used (44,45), whereas a direct grinding and polishing technique is used for Hamburg's ellipsoid.) The Wolter SPEM has been installed on a bending-magnet beamline at the Photon Factory synchrotron of the National Laboratory for High Energy Physics of Japan. Figure 10(a) shows details of the mirror design, while the actual experimental setup is depicted in Figure 10(b) (46). The mirror demagnifies the pinhole-defined source by a factor of 29.8. At a photon energy of 150 eV, the reflectivity is $\approx 77\%$ for a Au-surfaced mirror, and the energy resolution is about 1 eV (47). The microscope has demonstrated a 0.3 μm resolution in total photoyield mode (in which all of the photoelectrons emitted are collected with the microchannel plate) and 2.0 μm in energy-selected mode (when a single electron energy is selected by the electron energy analyzer) (48). The latter value can be greatly improved by increasing the signal-to-noise ratio by operating with higher X-ray flux on an undulator beamline.

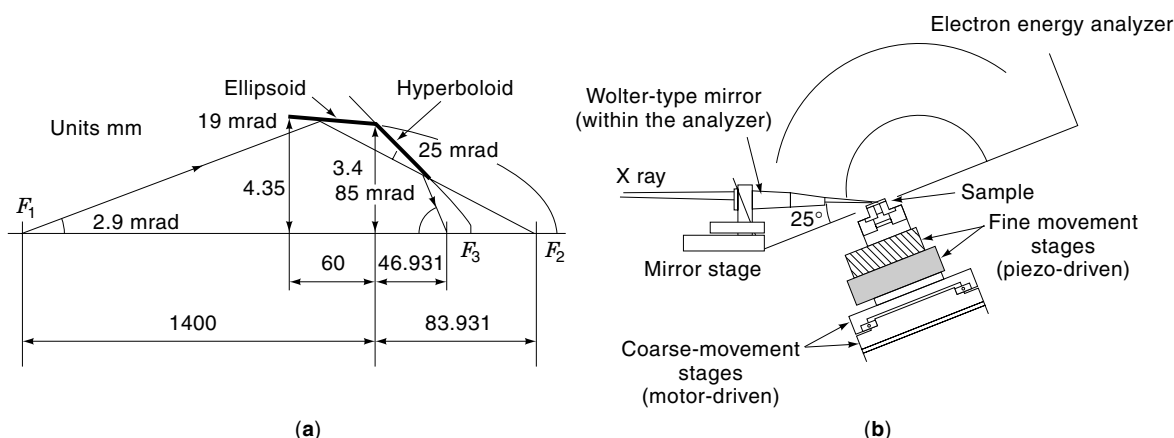


Figure 10. (a) Specifications of Wolter optics. (b) Layout of Wolter microscope developed by Hitachi. [Courtesy of M. Hasegawa (46).]

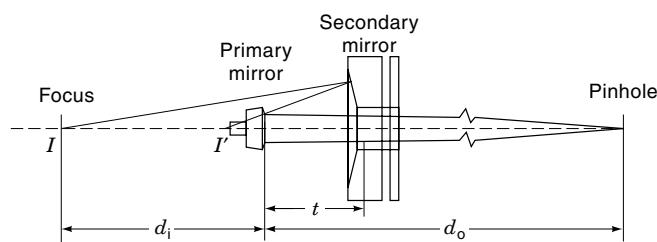


Figure 11. SO geometry and image formation.

The relationship between the signal-to-noise ratio and resolution is discussed below in the subsection "Resolution: Limiting Factors."

Schwarzschild Objective. The Schwarzschild objective (SO) was described by Schwarzschild in 1905 and has been used in telescopes for a long time, but was applied to X-ray microscopy only some 20 years ago by Haelrich (49), following the suggestion of Eberhard Spiller that multilayers could be used as high-energy reflectors. A commercial SO from Ealing was coated with some of the first X-ray multilayers with sufficiently high reflectivity in normal incidence. The first synchrotron-based microscope using these optics was constructed at DESY's HASYLAB. The resolution at a photon energy of 62 eV was 1.5 μm , 25 times the diffraction limit. This was due to the poor alignment tolerances of the Schwarzschild objective used (49). Lovas, Spiller, and coworkers built a modified instrument with a SO designed to minimize aberration (comatic aberration in particular) and make diffraction-limited performance attainable. They forecasted a resolution of 0.1 μm in 1981 (50).

This did not prove so simple, however, and MAXIMUM, a soft-X-ray spectromicroscope developed at the University of Wisconsin—Madison, was the first and only such SO-based instrument to reach that resolution, more than a decade later (51). The design optical setup and design parameters of the SO used in MAXIMUM are shown in Fig. 11 and Table 1, respectively. Photographs of the mirrors coated with Ru/B₄C multilayers for a peak reflectivity at 130 eV are shown in Fig. 12. As shown schematically in Fig. 13, X rays emitted by an undulator are monochromatized to 130 eV and focused on a pinhole. The pinhole defines a source that is demagnified 20 times to form an X-ray microspot on the sample. The sample is scanned using a piezo-driven flexural hinge stage, and the photoemitted

electrons are detected by an electron energy analyzer (a cylindrical mirror analyzer or CMA) to produce a two-dimensional photoelectron image. In another mode of operation, the sample is held fixed, and the CMA is used to collect an EDC of the photoelectrons.

In a TiSi₂ formation study done with MAXIMUM on beamline 6.3.2 of the Advanced Light Source of the Lawrence Berkeley National Laboratory, static random access memory (SRAM) wafers from IBM, East Fishkill, were subjected to Ti deposition and rapid thermal anneal (RTA) steps in an ultrahigh-vacuum (UHV) chamber attached to the main microscope chamber. This was done to avoid exposure to air, which would cause surface contamination and oxidation, and allowed a clean photoemission signal to be obtained from the TiSi₂ formed. TiSi₂ was expected to form on top of polysilicon features overlying a silicon dioxide base layer. The polysilicon pads and lines of the SRAM sample were silicided by depositing ≈ 30 nm Ti *in situ* after subjecting the wafer to a light chemical clean, then annealing at $\approx 800^\circ\text{C}$ for 20 s and flashing to $\approx 1000^\circ\text{C}$. Figure 14(a) and (b) show two 100 $\mu\text{m} \times 100 \mu\text{m}$ photoelectron micrographs acquired at 25.74 eV and 20.48 eV, respectively, of the same area of a SRAM sample. These energies correspond in turn to the unshifted and oxide-shifted, respectively, Si 2*p* core levels shown in the EDCs in Fig. 14(c). The two EDCs were taken on a pad (region A) and on an area near the pad edge (region B) that appears bright in Fig. 14(b). The photon energy was 129.3 eV.

The lines and pads appear to be unaffected in Fig. 14(a), but exhibit a different morphology in Fig. 14(b). Images acquired with secondary photoelectrons were similar to Fig. 14(b) as well. However, secondary-electron images taken *before* the (unusually high) formation anneal resembled Fig. 14(a), as did images taken on samples silicided with the normal 700°C to 900°C sequence. This suggests that the differences observed in Fig. 14 and are due to the high-temperature processing, namely the resulting agglomeration of the silicide in the lines and also lateral growth of the TiSi₂ onto the SiO₂ at the edges of the silicided polysilicon regions. Because Fig. 14(b) is tuned to the SiO₂ signal, we can conclude that the bright regions in this image represent a larger amount of SiO₂ at the surface. This is because some regions of the SiO₂ are exposed by agglomeration of the overlying laterally grown TiSi₂. The laterally grown silicide is also thinner on the SiO₂ than the main silicide formed on the polysilicon, whence the higher oxide signal there. Neither effect is evident in Fig. 14(a), which is tuned to the unshifted Si 2*p* because the Si concentration is always higher in the main silicide formed on the polysilicon than on the Ti-covered silicon oxide or the thin laterally grown silicide. SEM analysis showed a similar morphology to that observed in Fig. 14(b), reflecting the differences in emissive characteristics of the materials and corroborating the presence of agglomeration.

MAXIMUM has also been applied to the study of another type of interconnect used in the semiconductor industry: Al—Cu alloy. Electromigration in interconnect lines is the transport of material along with electron flow when a current is passed through the line. This can lead to the formation of voids and, eventually, an open circuit. The presence of Cu is thought to retard the electromigration process. Cu is, however, not very soluble in Al and hence may segregate to different phases such as θ -Al₂Cu as well as to different locations in the line such as grain surfaces and interfaces.

Table 1. Design Parameters of Schwarzschild Objective

Object distance d_o	1298 mm
Image distance d_i	120.78 mm
Distance between mirrors, t	55.498 mm
Primary mirror:	
Radius of curvature	66.336 mm
Outer diameter	25.5 mm
Inner diameter	12.5 mm
Secondary mirror:	
Radius of curvature	117.253 mm
Outer diameter	80 mm
Inner diameter	25.5 mm
Demagnification	20×
Numerical aperture	0.2

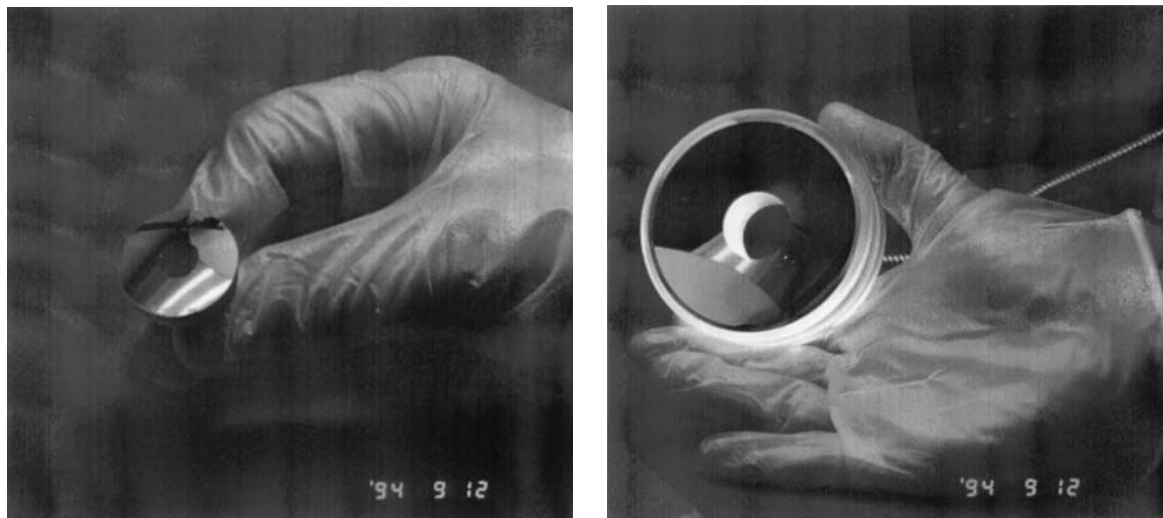


Figure 12. Primary and secondary mirrors of the Schwarzschild objective coated with Ru/B_4C multilayers. Note the pinholes in the coating that allow light through.

A $5\text{ }\mu\text{m}$ wide line patterned in a 4% Cu Al–Cu alloy on thermal oxide grown on Si was stressed (by passing current and heating *in situ* in the microscope chamber) to induce electromigration. The line was examined with MAXIMUM before and after electromigration. Images acquired at different photoelectron energies before and after stressing are shown in Fig. 15. Already before electromigration, the image acquired at 122 eV (the Cu 3d core level) shows bright spots corresponding to Cu-rich precipitates. A comparison between spectra acquired on the precipitate labeled P and a uniform, “ordinary” area O is shown in Fig. 16(a). P exhibits a stronger Cu 3d signal than O, as expected, in Fig. 16(a). The presence of Cu in the Al film at P shifts the Al 2p core level to lower binding energy (higher kinetic energy) with respect to O, as can be seen in Fig. 16(b). This explains why the image in Fig. 15(a), acquired using 50 eV photoelectrons, shows the precipitates brighter than the uniform film area, while that acquired using 47 eV photoelectrons shows them darker. The chemical shift to higher kinetic energy results in more emission from the precipitates at 50 eV and more from the O area at 47 eV.

After electromigration, it is seen from Fig. 15(b) that the Cu-containing precipitates have not moved and that voids

have formed in the line. The void areas in the photoelectron image acquired at 42 eV appear bright because of surface charging of an Al_2O_3 skin left behind over the void. This charging causes the Al 2p core level to be shifted to higher

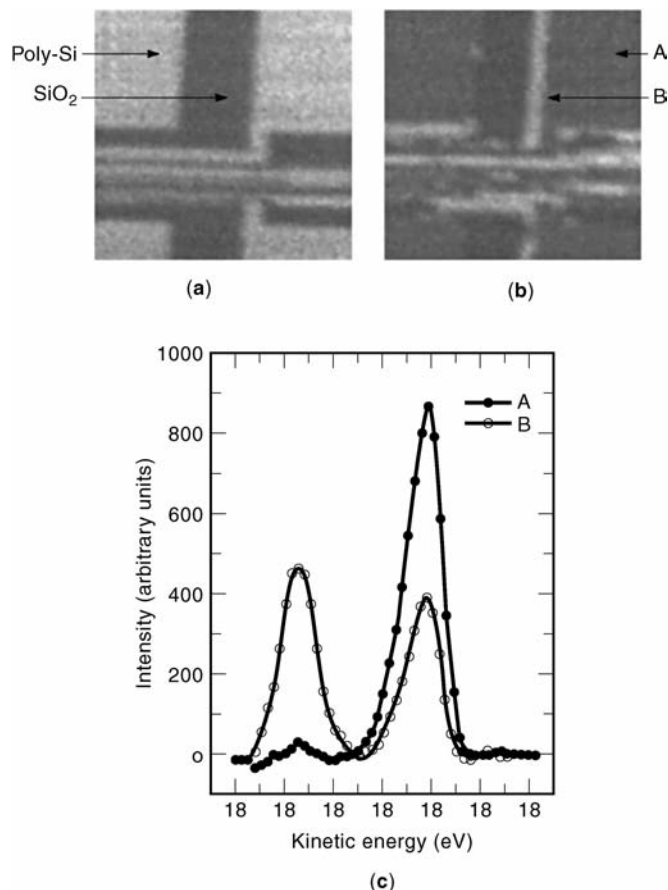


Figure 14. $100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$, $1\text{ }\mu\text{m}$ step photoelectron maps of the same silicided area taken at the unshifted (25.75 eV) and oxide-shifted (20.48 eV) peaks of the Si 2p core level shown in the EDC.

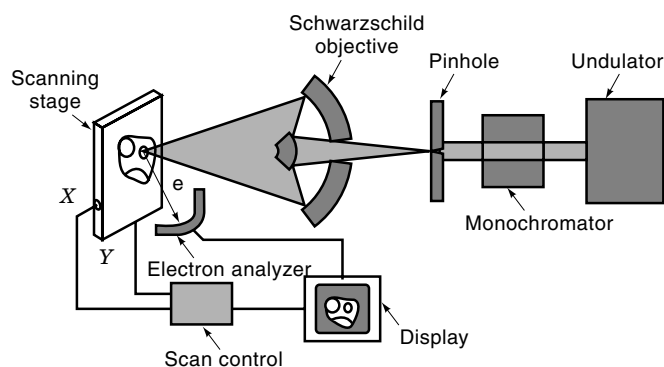


Figure 13. Schematic of the spectromicroscope MAXIMUM.

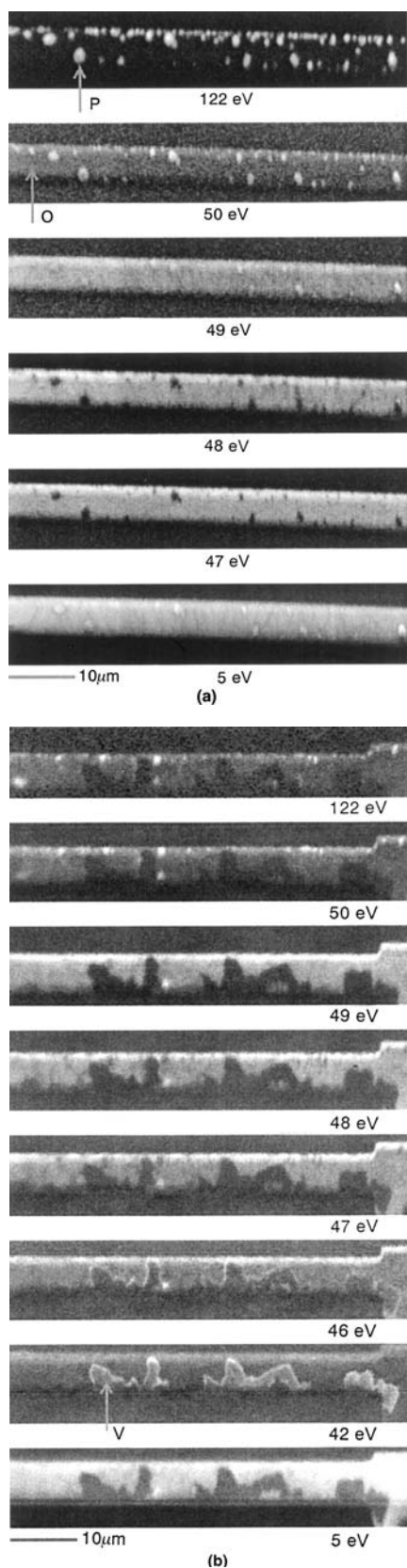


Figure 15. (a) Micrographs of 5 μm wide Al-4% Cu line before electromigration. P labels a Cu-containing precipitate, and O labels an "ordinary" point on the line. (b) Micrographs of 5 μm wide Al-4% Cu line after electromigration. V marks a void. Images are acquired at the indicated photoelectron energies. (Courtesy of M. Solak.)

binding energies (since the sample becomes positively charged during photoemission because the void prevents a good earth connection to supply electrons, thus causing electrons to be more tightly bound) and hence lower kinetic energies, resulting in the void appearing bright at the shifted Al 2p peak at 42 eV (69).

There are other implementations of the SO microscope as well. Another microscope has been designed to run the SO with C $K\alpha$ radiation (277 eV). This relatively high energy (for normal-incidence optics) is made possible by the use of NiCr/C multilayers. The objective's resolution has been measured to be better than 0.5 μm (52).

IMAGE FORMATION

So far we have discussed many different approaches to X-ray microscopy and have compared the resolutions achieved and have listed experiments performed, but we have not yet done a detailed analysis of the image formation itself. We will do so here for microscopes utilizing X-ray optics and synchrotron radiation.

Synchrotron radiation can be generated by three types of devices: bending magnet, wiggler, and undulator. The first two types are similar in their broad energy spectrum and their large opening angles of radiation and can be treated together. In both a bending magnet and a wiggler, the radiation from one electron at one point is independent of that at a different point in the synchrotron, and photons are emitted incoherently along the tangent at each point of the electron trajectory. However, when the emittance of the electron beam becomes very small (as it does for third-generation synchrotron sources such as the ALS), diffraction effects arise along the electron trajectories (53). The interference effect is very strong in undulators because of the small opening angle of the radiation and results in a spectrum with harmonic peaks at certain wavelengths. This spectral purity implies a high degree of inherent temporal coherence in undulator radiation, unlike bending-magnet and wiggler radiation. The photon field and the trajectory remain in phase over a large distance. Again, when the emittance of the electron beam is very small, the radiation cannot be assumed to be spatially incoherent.

It is convenient to define the following phase-space relationship for quasimonochromatic sources such as undulators:

$$d \cdot \theta \geq \lambda/2\pi \quad (13)$$

where d is the source (electron beam) diameter, and θ is half the opening angle of the radiation, which is determined solely by the electron trajectory. If the source has an opening angle θ at a wavelength λ , any points separated by less than $\delta = \lambda/\sin \theta \approx \lambda/\theta$ cannot be resolved, that is, the source is *spatially coherent*, or *diffraction-limited*. The source size d is larger than δ in practice, however. Currently the ALS has undulators with a phase-space at least 5 to 50 times larger than the diffraction limit (54).

As stated earlier, even the radiation from bending magnets and wigglers has some spatial coherence and, although not temporally coherent, is typically monochromatized before entering the X-ray imaging system. Thus we can speak in general of the quasimonochromatic synchrotron radiation used in X-ray microscopes as being *partially coherent*.

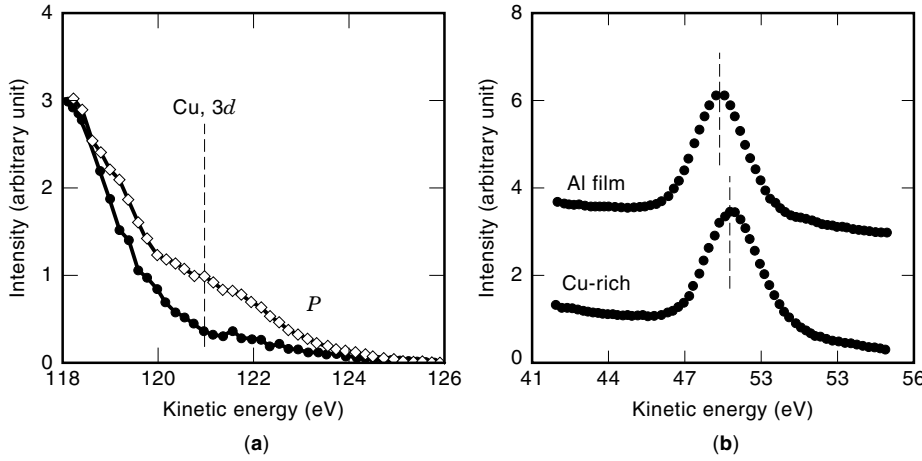


Figure 16. (a) EDCs acquired at points P (precipitate) and O (ordinary) of Fig. 15(a) showing higher Cu 3d signal at P. (b) EDCs acquired on an Al film area O and a Cu-rich area P, showing that the presence of Cu shifts the Al 2p to higher kinetic energy (lower binding energy). (Courtesy of M. Solak.)

The Frequency Response of an Optical System

Spatial-Frequency Cutoff. Following the treatment of Born and Wolf (55), let us consider the two limiting cases of coherent and incoherent illumination. $\mathcal{K}$, the *frequency response function* in the coherent case, is actually equivalent to the pupil function G with a change of variables, and we may write

$$\mathcal{K}(f, g) = G(\lambda Rf, \lambda Rg) \quad (14)$$

where R is the radius of the Gaussian sphere of reference (which is centered on the geometric image point and passes through the center of the exit pupil). Because the pupil function has a value of zero outside the exit pupil itself, frequencies above a certain value are not transmitted. In particular, if the exit pupil is approximated by a circle of radius a , then the pupil function will be zero for points outside the circle and only those frequencies will be transmitted for which

$$(\lambda Rf)^2 + (\lambda Rg)^2 < a^2 \quad (15)$$

holds. For an optical system with a small image-side angular aperture that obeys the sine condition

$$M = \frac{y_1}{y_0} = \frac{n_0 \sin \theta_0}{n_1 \sin \theta_1} \quad (16)$$

where M is the magnification and n_0 and n_1 the indices of refraction in the object and image spaces, respectively—we may make the approximation

$$\frac{a}{R} = \tan \theta_1 \approx \sin \theta_1 \quad (17)$$

where θ_1 is the half angle. For a one-dimensional (1-D) object, if $\Delta y_0 = 1/g$ is the period corresponding to the spatial frequency g , we may write that only those frequencies corresponding to the spatial periods that satisfy

$$P_y > \frac{\lambda_0}{A_1} \quad (18)$$

where $\lambda_0 = n_1 \lambda$ is the vacuum wavelength and $A_1 = n_1 \sin \theta_1$ is the image-side numerical aperture, are transmitted.

In the incoherent case, $\mathcal{L}$ is the frequency response function and, because of the connection between $\mathcal{K}$ and the pupil function G shown in Eq. (14), is also the autocorrelation function of the pupil function. If we again consider a circular exit pupil of radius a , the autocorrelation function will have a non-zero value only if there is some overlap between two displaced pupil functions, for example, if the maximum displacement between the two pupil functions is less than the pupil diameter $2a$. Thus only those frequencies will be transmitted that satisfy

$$(\lambda Rf)^2 + (\lambda Rg)^2 < (2a)^2 \quad (19)$$

Then, for a 1-D object, only those spatial frequencies will be transmitted whose spatial periods satisfy

$$P_y > \frac{\lambda_0}{2A_1} \quad (20)$$

The spatial frequency cutoff for the partially coherent case falls between the cutoffs for the coherent and incoherent cases. In general, the minimum spatial period resolvable, Δ , is the reciprocal of the cutoff frequency. For

$$\Delta = d \frac{\lambda_0}{A_1} \quad (21)$$

we have

$$\begin{aligned} d = 1 & \quad \text{coherent} \\ d = \frac{1}{2} & \quad \text{incoherent} \\ \frac{1}{2} < d < 1 & \quad \text{partially coherent} \end{aligned} \quad (22)$$

Accordingly, incoherent illumination can resolve the smallest spatial period, and hence can produce the highest spatial resolution, followed by partially coherent, and then by coherent illumination. It is important to note, however, that the results obtained here apply rigorously only for an unobstructed circular exit pupil in an optical system that has a small angular aperture and satisfies the sine condition. Other pupil illumination functions may increase or decrease the resolution.

Diffraction-Limited Optical System. All axially symmetric X-ray scanning systems (e.g., SO-based microscopes, the SPEM, and the STXM) have a central obstruction. All microscopes, both scanning and imaging, also have aberrations. These and other effects in real systems can be taken into account in the formulation of the complex transmission function K and are reflected in its inverse Fourier transform $\mathcal{K}$, which is directly related to the frequency response function in the coherent, incoherent, and partially coherent cases. We will not treat the complex topic of aberrations here. Rather, we will treat the opposite, ideal case of the *diffraction-limited optical system*.

The term “diffraction-limited” is often bandied about in X-ray microscopy. We have used it ourselves when talking about third-generation synchrotron sources, but it is different for photons. An imaging optical system is called diffraction-limited if it converts a spherical wave emanating from a point source into another spherical wave converging to the Gaussian image point in the imaging plane (56). If the wavefront leaving the exit pupil is not perfectly spherical, it is an indication of aberrations in the system. In a diffraction-limited system, then, only the functional form of K is important in determining the system’s imaging properties. We consider an optical system of small enough angular aperture to justify the paraxial approximation and the lens equation

$$\frac{1}{d_0} + \frac{1}{d_1} = \frac{1}{F} \quad (23)$$

where d_0 is the object distance, d_1 is the image distance, and F is the focal length of the lens.

For such a system and coherent illumination, the complex transmission function K is given by the Fraunhofer diffraction pattern of the lens aperture centered on the geometrical image point (56). Then for a lens with a circular aperture and a central circular obstruction we may write

$$K(x, y) = \frac{1}{1 - \epsilon} \left(\frac{J_1(\nu)}{\nu} - \frac{\epsilon J_1(\epsilon\nu)}{\nu} \right) \quad (24)$$

where J_1 is a Bessel function of the first order, $\nu = 2\pi, x^2 + y^2$, with x, y measured in units of λ/A_1 , and ϵ represents the fraction of aperture obstructed and ranges from 0 to 1. It can be seen that if $\epsilon = 0$ Eq. (24) reduces to the Fraunhofer diffraction pattern from a circular aperture. $K(x, y)$ is also called the amplitude point spread function and is the impulse response of the optical system. $\mathcal{K}(f, g)$, the inverse Fourier transform of $K(x, y)$, is the frequency response of the system, as stated earlier.]

For incoherent illumination, the frequency response of the system is (f, g) , and the impulse response (also called the intensity point spread function) is given, for an annular aperture, by the square of K as defined in Eq. (24). For a circular aperture, $\epsilon = 0$ and the impulse response is the Airy pattern.

Scanning emission microscopes like MAXIMUM are *incoherent* imaging systems in that, even though the illumination may be partially coherent, the photoelectron signal collected per pixel has no relation to the phase of the incident light and hence is incoherent with that of neighboring pixels.

What is the Meaning of Resolution?

There are several criteria than can be used to determine the resolution of an imaging optical system. One commonly used criterion is the Rayleigh criterion. Strictly speaking, it applies only to an aberration-free, unobstructed optical system. According to this treatment, two monochromatic and mutually incoherent point sources can just barely be resolved if their separation is such that the maximum of the Airy pattern of one point falls on the first minimum of the other’s Airy pattern, resulting in a saddle point between the two maxima of the summed pattern. This separation δ is given by

$$\delta = \frac{1.22\lambda}{2A_1} \quad (25)$$

and is also the diameter of the Airy disk. For a scanning microscope, the two point sources are replaced by two small pinholes with a diameter much smaller than their separation, but the same analysis applies. Often, δ is quoted for a microscope although the preconditions of incoherence and lack of aberration may not be satisfied. This is permissible only as an approximation. In addition, strictly speaking, $\lambda/2A_1$ is the diffraction-limited resolution of an *incoherent* imaging system. (λ/A_1 is the diffraction-limited resolution of a *coherent* imaging system.) However, since this is close to the Rayleigh criterion [Eq. (25)], the two terms are sometimes used interchangeably, and the factor of 1.22 crops up.

Although, the diffraction-limited resolution has long been thought to be the ultimate resolution for an optical system, it is in principle possible to resolve certain types of objects that are smaller. Goodman (57) states two theorems that apply to bounded objects. The first theorem states that the Fourier transform of a function bounded in the (x, y) plane is an analytic function in the (f, g) plane. The second theorem states that if an analytic function in the (f, g) plane is known exactly in an arbitrarily small finite region, then the whole function can be found uniquely by analytic continuation. For an optical system and a bounded object, this means that the entire object frequency spectrum can be calculated by analytic continuation if a finite portion of it can be determined exactly (in the absence of noise) from the image frequency spectrum, which is determined by the frequencies passed by the point spread function of the optical system. Thus, because the entire object frequency spectrum is then known, the object itself can be reconstructed exactly, thus allowing a kind of *superresolution*. This may be possible, for instance, in the ubiquitous case of the circular aperture. If the surrounding material is not completely opaque to the incident light, but rather induces a phase shift, it is possible to narrow the diameter of the Airy disk. Since the diameter of the Airy disk is a good estimate of the resolution, this will improve the resolution too.

Actually, assuming that a finite portion of the object frequency spectrum is an analytic function is a highly restrictive hypothesis, since it is difficult to have a priori knowledge of the object. Normally we have an unknown object to which we apply criteria such as the Rayleigh criterion to determine the resolution. But in that case we have available only the information that is transmitted by the optical system, and not intrinsic information about the object. Thus it appears that the amount of information that can be transmitted by an optical system is fixed, and if the resolution is to be improved, it must be at the expense of something else. In theory superresolution should be

feasible in some controlled cases, but in the real world there is significant noise, which limits the accuracy with which the frequency spectrum of the bounded object can be measured within the spatial-frequency bandpass of the optical system. This effect, together with the effect of aberrations, can preclude not only superresolution, but diffraction-limited resolution as well.

Resolution: Limiting Factors. Often the first question asked about a particular X-ray microscope is “What’s its resolution?” The answer, once extracted, must include three values: the everyday resolution, the best resolution achieved, and the theoretical best resolution. The everyday resolution is the poorest of the three and arises from a compromise between spatial resolution, signal level, noise level, and other experimental parameters such as acquisition time and energy resolution. The best resolution is a much touted number and is typically obtained only with painstaking alignment and fine tuning followed by a lengthy data acquisition period. The theoretical best resolution depends on the image formation under different types of illumination, as discussed earlier. It is instructive to discuss the practical limitations on resolution in a quantitative way in order to understand and predict experimental results.

Effect of Noise. The signal measured in an image acquired with an X-ray spectromicroscope is a combination of the true signal and the background signal. The background signal can be a function of the photon energy or the photoelectron energy as well as of the spatial frequency. There is also noise accompanying both signals, and it reduces the image contrast. What is actually measured in the image is always intensity, regardless of the illumination. Thus we need to know the frequency spectrum of the intensity. For an incoherent system, which is linear in intensity, it is straightforwardly given by Eq. (26). For partially coherent and coherent systems, the intensity cannot be expressed in terms of a frequency transfer function acting as a linear filter to transform the object intensity $\mathcal{I}_0$ into the image intensity $\mathcal{I}_1$. As mentioned earlier, scanning emission microscopes are incoherent systems. Even transmission X-ray microscopes rely on the incoherence in partially coherent light to deliver higher resolution because of the higher spatial-frequency cutoff. It is useful, then, to examine the effect of noise in the mathematically simpler incoherent case:

$$\mathcal{I}_1(f, g) = \mathcal{I}_0(f, g) \mathcal{L}(f, g) \quad (26)$$

where $\mathcal{L}$ is the frequency response function for incoherent illumination. Another name for the normalized $\mathcal{L}(f, g)$ is the *optical transfer function* (OTF). The *modulation transfer function* (MTF) is simply $|\mathcal{L}(f, g)|$, the modulus of the OTF, and is a useful quantity for describing the resolving capability of an optical system under incoherent illumination. The MTF $\mathcal{T} = |\mathcal{L}(f, g)|/|(0, 0)|$ for the simple case of a diffraction-limited optical system with an unobstructed circular aperture can be obtained with K as in Eq. (24) and $\epsilon = 0$. Because of circular symmetry, radial coordinates may be used and the MTF may be computed to be

$$\mathcal{T}(f) = \begin{cases} 0, & f > f_0 \\ \frac{2}{\pi} \left[\cos^{-1} \left(\frac{f}{2f_0} \right) - \frac{f}{2f_0} \sqrt{1 - \left(\frac{f}{2f_0} \right)^2} \right], & f \leq f_0 \end{cases} \quad (27)$$

where f_0 is the incoherent cutoff spatial frequency $2A_1/\lambda$ (58). Using a linear approximation for the transmitted frequencies, we may write

$$\mathcal{T}(f) = 1 - \frac{f}{f_0} \quad (28)$$

and for a one-dimensional object we may rewrite Eq. (26) as

$$\mathcal{I}'_1(f) = \mathcal{I}'_0(f) \mathcal{T}(f) \quad (29)$$

where the primed quantities indicate normalization. Following Cerrina’s treatment (59) of the limiting effect of statistical noise on spatial resolution, we study the signal levels in the image. Since the intensity is measured as a count rate, the signal $\mathcal{I}$ is obtained by multiplying $\mathcal{I}'$ by the counting time Δt . Then, the background can be taken into account in the same fashion, and the total signal may be written as

$$\mathcal{I}_T = \mathcal{I}_R + \mathcal{I}_B \quad (30)$$

$$= (\mathcal{I}'_{0,R} \mathcal{T} + \mathcal{I}'_{0,B} \mathcal{T}) \Delta t \quad (31)$$

where subscripts R and B refer to the “real” and background signals, respectively. In order to resolve a feature, $[\text{Sscr}]_R$ must be higher than the noise level. This can be quantified by

$$\mathcal{I}_R = \mathcal{I}_T - \mathcal{I}_B \geq n(\sigma_T + \sigma_B) \quad (32)$$

where n is a positive nonzero integer indicating the level of confidence and σ_T and σ_B are the noise levels in the total and background signals, respectively.

Because Poisson’s distribution describes statistical noise, for N counts a random noise of $\sqrt{N}$ will be observed. Then we may substitute $\sqrt{\mathcal{I}_T}$ and $\sqrt{\mathcal{I}_B}$ for σ_T and σ_B , respectively. After defining the signal-to-noise ratio R as $\mathcal{I}'_{0,R}/\mathcal{I}'_{0,B}$ we may write

$$\mathcal{I}_R \geq \frac{n^2}{\mathcal{T}} \left(\sqrt{1 + \frac{1}{R}} + \sqrt{\frac{1}{R}} \right)^2 \quad (33)$$

For a small signal-to-noise ratio ($R < 1$), then, we may write the real object frequency distribution $\mathcal{I}'_{0,R}$ as

$$\mathcal{I}'_{0,R}(f) \geq \frac{4n^2}{\mathcal{T}(f) \Delta t R} \quad (34)$$

With the help of Eq. (28), we may solve this for the frequencies transmitted by an idealized incoherent optical system of circular aperture in the presence of statistical noise. We obtain

$$f \leq f_0 \left(1 - \frac{4n^2}{\mathcal{I}'_{0,R} \Delta t R} \right) \quad (35)$$

which shows that the maximum distinguishable frequency passed by a system with statistical noise is reduced, thus increasing the minimum period resolvable and hence decreasing the resolution. If the counting time Δt is sufficiently long for the desired level of confidence n , this frequency will approach the pure incoherent cutoff frequency for a diffraction-limited system. The counting time is directly related to the

absolute intensity level of the image, which in turn is related to the brightness. This is true not just for incoherent scanning systems, but for X-ray microscopes in general, as long as there is no significant temporal oscillation in the true signal or the noise.

Importance of Brightness. Brightness is often defined for synchrotrons and insertion devices as the number of photons per unit time per unit angle per unit area per unit fraction of bandwidth (typical units are photons/s.mrad².mm².0.1%). Without deriving expressions for the brightness in specific cases, we simply note that it is a measure of the ability of a source to illuminate a small area with small angular divergence. Clearly a high-brightness source (such as a third-generation synchrotron and its insertion devices) results in more incident flux per unit area on the sample in a X-ray microscope using that source. In a sequential imaging system such as a scanning microscope, this higher flux automatically means that the intensity level goes up, and thus, for a given counting time Δt , the signal goes up. Hence, better resolution can be obtained for the same optical system by counting for the same amount of time if all other parameters remain the same. In a parallel-imaging X-ray microscope such as the PEEM, although the X rays are not focused to a microspot, the electrostatic optics accept only those photoelectrons from a small area of the sample whose size is determined by the field of view for different magnifications and potentials of the optical column, the signal also increases.

Thus, the brighter the source is, the greater is the incident X-ray intensity on the area of interest, the higher the signal level in the photoelectron image, and the greater the capability to acquire spectra from small regions in reasonable real time. This time is longer for nonsynchrotron X-ray sources such as plasma point sources, which have a lower-flux output. The actual intensity at the image depends on various factors, including detector efficiency, sample quantum efficiency, acquisition time, and source brightness. Brightness can overcome shortcomings in the efficiencies and dramatically reduce the acquisition time, thus enabling higher spectromicroscopic resolution. This includes the ability to acquire higher-energy-resolution spectra from the samples, with a concurrent high spatial resolution. With fourth-generation synchrotrons like SOLEIL (France) and SLS (Switzerland) in the offing, the future of X-ray microscopy shines only brighter.

BIBLIOGRAPHY

1. T. A. Hall, H. O. E. Rockert, and R. L. deC. H. Saunders, *X-ray Microscopy in Clinical and Experimental Medicine*, Springfield, IL: Charles C. Thomas Publisher, 1972.
2. A. Moewes, *Lumineszenz-Rastermikroskopie im Bereich weicher Röntgenstrahlung*, Ph.D. thesis, Universität Hamburg, 1994.
3. B. K. Agarwal, *X-Ray Spectroscopy: An Introduction*, 2nd ed., New York: Springer-Verlag, 1991.
4. J. B. Vukovic and R. Antanasijevic, X-ray microscopy today, *Radiat. Meas.*, **28**: 817, 1997.
5. J. Stöhr, *NEXAFS Spectroscopy*, New York: Springer-Verlag, 1992.
6. R. N. Watts et al., A transmission x-ray microscope based on secondary-electron imaging, *Rev. Sci. Instrum.*, **68**: 3464, 1997.
7. M. Cardona and L. Ley (eds.), *Photoemission in Solids I: General Principles*, Vol. 1, New York: Springer-Verlag, 1978.
8. W. Meyer-Ilse et al., New high-resolution zone plate microscope at Beamline 6.1 of the ALS, *Synchrotron Radiat. News*, **8**: 29, 1995.
9. J. M. Heck et al., Resolution determination in x-ray microscopy: An analysis of the effects of partial coherence and illumination spectrum, *J. X-Ray Sci. Technol.*, 1998.
10. F. Vollrath, E. Sogaard, and J. T. Moller, X-ray microscopy in Aarhus, in J. Thieme et al. (eds.), *X-Ray Microscopy and Spectromicroscopy*, New York: Springer-Verlag, 1997.
11. M. Mundschau, Photoelectron emission microscopy, *Synchrotron Radiat. News*, **4**: 29, 1991.
12. W. Engel et al., A UHV-compatible photoelectron emission microscope for applications in surface science, *Ultramicroscopy*, **36** (1–3): 148, 1991.
13. B. P. Tonner and G. R. Harp, Photoelectron microscopy with synchrotron radiation, *Rev. Sci. Instrum.*, **59**: 853, 1988.
14. B. P. Tonner and D. Dunham, Sub-micron spatial resolution of a micro-XAFS electrostatic microscope with bending magnet radiation: Performance assessments and prospects for aberration correction, *Nucl. Instrum. Methods A*, **347**: 436, 1994.
15. B. P. Tonner et al., The development of electron spectromicroscopy, *J. Electron Spectrosc. Relat. Phenom.*, **75**: 309, 1995.
16. D. Dunham et al., Evaluation of aberration coefficients of practical electrostatic lenses for x-ray absorption micro-spectroscopy and imaging, *Nucl. Instrum. Methods A*, **347**: 441, 1994.
17. B. P. Tonner, Energy-filtered imaging with electrostatic optics for photoelectron microscopy, *Nucl. Instrum. Methods A*, **231**: 60, 1990.
18. G. De Stasio et al., MEPHISTO: Performance tests of a novel synchrotron imaging photoelectron spectromicroscope, *Rev. Sci. Instrum.*, **69**: 2062, 1998.
19. B. P. Tenner et al., A photoemission microscope with a hemispherical capacitor energy filter, *J. Electron Spectrosc. Relat. Phenom.*, **84**: 211–229, 1997.
20. S. Singh, X-ray photoemission spectromicroscopy and its application to the study of patterned titanium silicide, Ph.D. thesis, Univ. of Wisconsin, Madison, 1996.
21. S. Singh et al., An x-ray spectromicroscopic study of the local structure of patterned titanium silicide, *Appl. Phys. Lett.*, **71**: 55, 1997.
22. D.-B. Kao et al., Two-dimensional effects in titanium silicidation, *IEEE Trans. Electron Devices*, **45**: 187, 1998.
23. J. Stöhr et al., Element-specific magnetic microscopy with circularly polarized x-rays, *Science*, **259**: 658, 1993.
24. B. P. Tonner et al., Imaging magnetic domains with the x-ray dichroism photoemission microscope, *Nucl. Instrum. Methods A*, **347**: 142, 1994.
25. S. Jakubith et al., Spatiotemporal concentration patterns in a surface reaction: Propagating and standing waves, rotating spirals, and turbulence, *Phys. Rev. Lett.*, **65**: 3013, 1990.
26. G. Ertl, Oscillatory kinetics and spatio-temporal self-organization in reactions at solid surfaces, *Science*, **254**: 1750, 1991.
27. P. Pianetta et al., Core level photoelectron microscopy, *J. Electron Spectrosc. Relat. Phenom.*, **52**: 797, 1990.
28. P. Pianetta et al., Core level photoelectron microscopy with synchrotron radiation, *Rev. Sci. Instrum.*, **60**: 1687, 1989.
29. G. D. Waddill et al., Photoelectron microscopy and spectroscopy using synchrotron radiation, *J. Vac. Sci. Technol. A*, **9**: 1634, 1991.
30. T. Komeda et al., Photoelectron microscopy and spectroscopy of Bi₂Sr_{2-x}Ca_{1-x}Cu₂O_{8+y}(100), *Phys. Rev. B*, **43**: 8713, 1991.
31. G. Schmahl et al., X-ray microscopy of biological specimens with a zone plate microscope, *Ann. N.Y. Acad. Sci.*, **342**: 368, 1980.
32. J. Kirz, X-ray microscopy, *Synchrotron Radiat. News*, **4**: 17, 1991.
33. H. Rarback et al., The scanning transmission microscope at the NSLS, *Nucl. Instrum. Methods A*, **291**: 54, 1990.

34. J. Kirz, C. Jacobsen, and M. Howells, Soft x-ray microscopes and their biological applications, *Q. Rev. Biophys.*, **28**: 33, 1995.
35. H. Ade, NEXAFS microscopy of polymeric samples, *Synchrotron Radiat. News*, **7**: 11, 1994.
36. H. Ade et al., X-ray microscopy in polymer science: Prospects of a "new" imaging technique, *Polymer*, **36**: 1843, 1995.
37. H.-J. Shin, Personal communication, 1996.
38. H. Ade et al., Images of a microelectronic device with the X1-SPeM, a first generation scanning photoemission microscope at the National Synchrotron Light Source, *J. Vac. Sci. Technol. A*, **9**: 1902, 1991.
39. M. P. Kiskinova, *ESCA Microscopy of Surfaces and Interfaces with Submicron Spatial Resolution*. New York: Materials Research Society, 1998.
40. E. Di Fabrizio et al., High-performance multilevel blazed x-ray microscopy Fresnel zone plates: Fabricated using x-ray lithography, *J. Vac. Sci. Technol. B*, **12**: 3979, 1994.
41. W. Yun et al., Design of a dedicated beamline for x-ray microfocusing- and coherence-based techniques at the Advanced Photon Source, *Rev. Sci. Instrum.*, **67**: 3373, 1996.
42. W. B. Yun, *X-ray Microscopy Applications to Materials Science*, New York: Materials Research Society, 1998.
43. W. Rodrigues et al., X-ray microdiffraction studies of an integrated laser-modulator system, **417**: 161, 1997.
44. K. Ninomiya et al., Fabrication of an axisymmetric Wolter type I mirror with a gold deposited surface, *Jpn. J. Appl. Phys.*, **28**: 2303, 1989.
45. M. Hasegawa et al., Fabrication of Wolter-type x-ray focusing mirror using epoxy resin, *Rev. Sci. Instrum.*, **65**: 2568, 1994.
46. Photon Factory activity report, The Photon Factory synchrotron, Tsukuba, Japan, 1996, p. c11.
47. K. Ninomiya and M. Hasegawa, Scanning photoelectron microscope with submicron lateral resolution using a Wolter-type x-ray focusing mirror, *J. Vac. Sci. Technol.*, **13**: 1224–1228, 1995.
48. M. Hasegawa, A. Yoneyama, and K. Ninomiya, Soft x-ray scanning photoelectron microscope using Wolter-type focusing mirror, *J. Electron Spectrosc. Relat. Phenom.*, **80**: 361, 1996.
49. R.-P. Haelbich, Untersuchung der Eigenschaften von Mehrschicht-interferenzspiegeln aus absorbierenden Materialien und ihre Verwendung in einem Raster-mikroskop für weiche Röntgenstrahlung. Ph.D. thesis, Universität Hamburg, 1980.
50. I. Lovas et al., Design and assembly of a high-resolution Schwarzschild microscope for soft x-rays. *High Resolution Soft X-ray Optics, Proc. SPIE* **316**: 90, 1981.
51. W. Ng et al., High resolution spectromicroscopy reaches the 1000 Å scale, *Nucl. Instrum. Methods A*, **347**: 422, 1994.
52. K. Murakami et al., Schwarzschild microscope for carbon $K\alpha$ radiation, *Appl. Opt.*, **32**: 7057, 1993.
53. K.-J. Kim, Brightness, coherence and propagation characteristics of synchrotron radiation, *Nucl. Instrum. Methods A*, **246**: 71, 1986.
54. D. Attwood, New opportunities at soft-x-ray wavelengths, *Phys. Today*, **45** (8): 24–31, 1992.
55. M. Born and E. Wolf, *Principles of Optics*, New York: Pergamon, 1959, Chaps. 9, 10.
56. J. W. Goodman, *Introduction to Fourier Optics*, San Francisco, CA: McGraw-Hill, 1968, Chap. 6, p. 103.
57. J. W. Goodman, *Introduction to Fourier Optics*, San Francisco, CA: McGraw-Hill, 1968, Chap. 6, pp. 134–135.
58. J. W. Goodman, *Introduction to Fourier Optics*, San Francisco, CA: McGraw-Hill, 1968, Chap. 6, pp. 119–120.
59. F. Cerrina, X-ray photo-electron spectromicroscopy, *J. Electron Spectrosc. Relat. Phenom.*, **76**: 9, 1995.
60. K. Shinohara et al. (eds.), *X-ray Microscopy in Biology and Medicine*, Tokyo: Springer Verlag, 1990.
61. P.-C. Cheng and G.-J. Chen (eds.), *X-ray Microscopy: Instrumentation and Biological Applications*, Heidelberg: Springer Verlag, 1987.
62. G. Margaritondo, *Introduction to Synchrotron Radiation*, New York: Oxford University Press, 1988.
63. E.-E. Koch (ed.), *Handbook of Synchrotron Radiation*, New York: North Holland, Vols. 1–4, 1983–1991.
64. D. Sayre et al. (eds.), *X-ray Microscopy II*, Berlin: Springer Verlag, 1988.
65. C. J. Jacobsen and J. E. Trebes (eds.), Soft X-ray microscopy, *Proc. SPIE*, 1741, 1993.
66. W. Yun (ed.), X-ray microbeam technology and applications, *Proc. SPIE*, 2516, 1995.
67. W. Meyer-Ilse, X-ray microscopy and microanalysis, *Encyclopedia of Applied Physics*, VCH Publishers, Inc., in press, 1998.
68. F. Gozzo, Personal communication, 1998.
69. H. Solak et al., An x-ray spectromicroscopic study of electromigration in patterned Al(Cu) lines, *Appl. Phys. Lett.*, in press, 1998.

SANGEET SINGH-GASSON
University of Wisconsin—Madison

DATA ACQUISITION AND RECORDING

RADIOTELEMETRY

Radiotelemetry, by definition, is the science and technology of automatically measuring and transmitting data by radio from remote sources, as from space vehicles, to receiving stations for recording and analysis. It involves preparing of data generated at remote sources, transmitting data through radio channels, and processing data at the receiving stations. Behind these three tasks, how to convey the information in a reliable and timely manner through a communication link is the most fundamental problem in this field.

There are many ways to define a communication link, depending on where the link starts and ends. Customarily, a communication link is the transmission path through which information-bearing waveform flows from one point to another. A generic link includes the modulator, the transmitter, the channel, the receiver, and the demodulator. Another popular view of a link is to include error-correction coding and encoding as part of the link because, as we see in a later discussion, link performance is a function of both modulation and coding. Yet, for a communication system that uses data compression, one might want to take an end-to-end view and include the error propagative effect of data compression in link design and analysis, because the error propagative property affects the integrity and throughput of the link. In the following sections, we briefly discuss some important elements of a communication link: data compression, forward error-correction coding, modulation, demodulation, and synchronization. However, before moving on to these individual topics, we first discuss a higher level consideration of the overall link reliability, namely, the link-budget analysis.

The main objective of link-budget analysis is to maximize the data return for a communication link subject to (1) the available communication resources and (2) the required data quality. Power and bandwidth are the two primary resources in communication, and data quality is usually expressed in terms of the frequency of errors in the received data, that is, the bit-error rate (BER). To design a communication system, the first step is to understand what the available resources are. Whether a system is power-limited (e.g., deep space communications) or bandwidth-limited (e.g., near-Earth satellite communications) determines the available options for modulation and error-correction coding schemes.

COMMUNICATION LINK-BUDGET ANALYSIS

The link budget is a balance sheet of the gains and losses in decibels (dB) of various parameters in the communication path. Many of these parameters are statistical or time-varying or both. The required BER is a direct function of the bit signal-to-noise ratio (SNR), denoted as E_b/N_0 , which in turn is a function of the modulation and error-correction coding used. Because of the statistical nature (uncertainty) of these parameters, a safety margin M is built in to guarantee the transmission data quality at any given time.

To put link-budget analysis into perspective, more parameters need to be defined and they are given as follows. To begin with, the total received signal-power-to-noise-power-spectral-

density ratio, P_T/N_0 , which encompasses all the power gains and losses in the communication path, is conveniently expressed as (1)

$$\frac{P_T}{N_0} = \frac{\text{EIRP}(G/T)}{kL_sL_0} \quad (1)$$

where EIRP is the effective isotropic radiated power of the spacecraft, G/T is the receiver sensitivity of the ground system, k is Boltzmann's constant, $L_s = (4\pi d/\lambda)^2$ is the space loss where d is the distance between the transmitter and receiver, λ is the wavelength, and L_0 denotes all other losses and degradation factors not specifically addressed in the above equation. Note that all gains and losses are statistical quantities. Then, let R_s be the symbol rate and R_c be the code rate expressed as the ratio of the number of information bits in a code word to the number of coded bits in a code word. The data rate or bit rate, denoted as R_b , is related to the symbol and code rates via $R_b = R_s R_c$. The bit SNR is written as

$$\frac{E_b}{N_0} = \frac{P_T/N_0}{R_s R_c} \quad (2)$$

Finally, let $(E_b/N_0)_{\text{req}}$ be the required E_b/N_0 to guarantee a certain BER. To ensure a reliable link, the constraining dependency is expressed in dB as follows:

$$\left(\frac{E_b}{N_0}\right)_{\text{req}} + M \leq \frac{P_T}{N_0} - R_s - R_c \quad (3)$$

The essence of link-budget analysis is to determine how much link margin M is required and this, in turn, is determined by how well the statistics of various link parameters are known. If M is too small, one might get a large data return but the data might have frequent errors. If M is too large, one might get continuous clean data but the throughput might be low. The analyst needs to make the right tradeoff on quantity versus quality, depending on the requirements and objectives of the service the link supports. Also, in an operational scenario, one usually starts out with a large M (a conservative approach) for a newly deployed system, and gradually reduces M (an aggressive approach) as more experience and confidence accumulates.

DATA COMPRESSION

In the midst of the Information Age, when new data—texts, images, sensor data, and many other forms of knowledge—are generated at a lightning pace, how these data are efficiently communicated and/or archived becomes increasingly important. Data obtained directly from sources usually contains much redundancy, and data compression is the process of applying “pressure” to remove the redundancy. The processes of compression and decompression add complexity to the overall system. The main issue in the use of data compression is the tradeoff between the efficiency of data handling and the complexity of signal processing to retrieve the information.

In the English language, letters like “a” and “e” are used more frequently than “y” and “z”, and words like “and” and “the” are much more often than “data” and “compression”. For images and sensor data, the differences between adjacent samples are small rather than large. This nonuniform frequency distribution of data associated with most information sources allows one to assign shorter bit patterns to represent the more frequent elements from the source and longer bit patterns to represent the less frequent elements from the source. This is the simplest view of data compression.

Data compression is broadly classified in two categories: lossless and lossy compression. Lossless compression denotes compression in which the decompressed or reconstructed data exactly match the original. Lossy compression represents compression methods where some degradation of quality is tolerable if a more compact but approximate representation is achieved. Next, we give brief surveys on text compression (lossless) and image compression (lossless and lossy). However, the compression of audio, video, and facsimile is beyond the scope of this article. Interested readers are encouraged to consult specialized books and technical journals in these areas.

Textual data is probably the first area where people applied compression. In 1832 Samuel Morse developed a code using dots, dashes, and spaces to represent letters, numbers, and punctuation for telegraph transmission. Each dot or dash is delimited by a space. A dot and a space take the same time. A dash is three times longer. Morse assigned fewer time units to more commonly occurring letters (e.g., “e” is a dot), and more time units to ones (e.g., “q” is dash-dash-dot-dash) that rarely occur. Today, most popular text compression schemes are different variations of the Lempel–Ziv (LZ) scheme developed in 1977 (2) and 1978 (3). LZ schemes use a dictionary approach. Dictionary coding operates by replacing groups of consecutive characters with indices in some dictionary. It exploits the frequent reoccurrence of certain exact patterns that are very typical of textual data. Other than compressing text, LZ schemes are reasonably good at compressing images and other sources. LZ schemes are generally considered universal compression algorithms. The UNIX command “compress” and the modem standard V42.bis are examples of applications using LZ schemes.

Image compression and other waveform compression is lossless or lossy. For scientific and medical applications that demand further processing and analysis, one might prefer lossless compression to ensure data accuracy. In other applications, such as image database, electronic photography, and desktop publishing, where transmission and storage bandwidths are limited, lossy compression is usually employed to reduce the data size at the expense of image quality.

Lossless image compression consists of two steps, modeling and coding. A predictive model predicts a pixel value based on a previously transmitted one, and compares it with the current value. The error value is coded instead of the original pixel value. On the receiver side, because the errors and the predictive scheme are known, the receiver recovers the value of the original pixel. An example of lossless image compression is the Rice code used in space and satellite communications.

Lossy image compression involves an irreversible quantization step that causes distortion between the original and the reconstructed image. Popular techniques in lossy image

compression include differential pulse code modulation (DPCM), transform coding, subband coding, vector quantization, and some hybrid techniques. The Joint Photographic Expert Group (JPEG) established an international standard on still image compression that uses the discrete cosine transform along with scalar quantization and entropy coding. Currently, JPEG encoder and decoder chips are available from many semiconductor manufacturers.

FORWARD ERROR-CORRECTION CODING

In 1948, the landmark paper “A Mathematical Theory of Communication” (4) by Claude Shannon provided a mathematical framework for analyzing communication, the process of sending information from one point to another through a noisy media, by using the concept of entropy. He provided an existence proof which shows that communication systems can be made arbitrarily reliable as long as the fraction of redundant signals in the signal stream exceeds a certain value, which is a function of the signal and noise statistics. The process of introducing redundancy in the signal stream to combat noise, so that retransmission is not needed, is called forward error-correction coding. Shannon did not show how this could be done in his famous theorem. Later on, much research effort has yielded a variety of interesting theoretical and practical results in this area.

Because redundant signals are deliberately introduced in the signal stream as error-correction coding, bandwidth expansion is expected. Encoding and decoding processes inevitably also add complexity to the overall system. However, it can be shown that, with a proper choice of coding scheme, the resulting E_b/N_0 can be considerably lower compared with an uncoded system. Thus error-correction coding allows one to trade power for bandwidth and complexity. In this article, the application of two popular error-correction coding schemes, block codes and convolutional codes are explained, without providing the detailed performance analysis of these codes. Other more advanced coding schemes, like trellis codes and turbo codes, are not covered here. Interested readers are encouraged to consult specialized books and technical journals in this field.

In block coding, a block of k bits is encoded into a block of n bits, where $n > k$. There are 2^k code words in a code space of 2^n . The code rate R_c , which is the ratio k/n , is a measure of the amount of added redundancy. The minimum distance $d_{\min}$, which is the Hamming distance between two closest code words, is a measure of the error-correction capability of the code. It can be shown that $d_{\min} \leq n - k + 1$ (Singleton bound). For bounded-distance decoding, the number of errors t that the code can correct is related to $d_{\min}$ by

$$2t + 1 \leq d_{\min} \quad (4)$$

In convolutional coding, sequences of k information bits long are encoded as continuous sequences using a kK -stage shift register, where K is called the constraint length. n output sequences are produced, each of which is generated by a selected set of combinations of these K input sequences. The n output sequences are multiplexed into a single-bit stream for transmission. Similar to a block code, the

ratio k/n is a measure of the redundancy of the convolutional code. The minimum distance d_{free} between two closest code sequences, is a measure of the error correction capability of the code.

The most popular block code used today is the class of Reed–Solomon (RS) codes. RS codes are maximum-distance-separable codes that achieve maximum error-correction capability in the bounded-distance decoding sense. Well-known encoding and decoding algorithms exist that are reliable in software and hardware. RS codes are flexible. Long, powerful RS codes exist for demanding applications, and short RS codes for simple applications. Deep space communications use the (255, 223) RS code as the outer code of the concatenated coding system. This code is block-interleaved and is used in conjunction with an inner convolutional code to provide reliable communications between the spacecraft and the ground stations. The concatenated coding system is designed to combat the additive white Gaussian noise of the deep space communication links. Simpler RS codes are used in compact disks (CD), a mass-marketed consumer product. The coding system used, called the cross-interleaved Reed–Solomon code (CIRC), is composed of two RS codes with (n,k) values (32, 28) and (28, 24). In between the two encoders are sets of delay lines which scramble the data stream. The CIRC coding system is designed to correct the burst-like errors, typical of a CD channel.

The most popular decoding scheme for convolutional codes is the Viterbi algorithm. The complexity of Viterbi decoding grows exponentially with the constraint length of the code, which is the number of memory elements in the encoder shift register. As mentioned earlier, convolutional codes are used in deep space communications. In the past, short constraint-length codes (7 or 9) are used because of the limitations in hardware technologies. Today, long constraint-length Viterbi decoders (up to 15) are designed and built. Convolutional codes are also popular in satellite communications and wireless communications. Presently, only short constraint-length convolutional codes are used in these areas.

INTERACTION BETWEEN DATA COMPRESSION AND ERROR CONTROL IN A COMMUNICATION SYSTEM

This section addresses the interaction between data compression and error control (containment/detection/correction) processes, two indispensable building blocks in modern digital communication systems. Data compression conserves transmission and storage bandwidth by removing redundancy in the source data. Error correction introduces controlled redundancy to the data to eliminate channel errors. A combination of both techniques ensures efficient and reliable transmission of information from one point to another. The famous separation principle indicates that data compression and error control processes can be completely separated, that is, the task of transmitting the output of a source through a channel can be separated without loss into the task of forming a binary representation of the source output and the task of sending a binary sequence through the channel. This is not quite true in practice, however. In evaluating an end-to-end communication system, we have to consider effects, like instrument failures, adverse weather conditions, synchronization loss, and other

unpredictable phenomena in the communication system and the channel. The error-control process cannot handle all possible kinds of errors. Thus the effect of error propagation should be considered when evaluating a data compression scheme in a communication system.

There are two issues to be dealt with: how to prevent errors from occurring and, if errors occur, how to contain the errors. The former issue can be dealt with easily (at least in the conceptual sense) by using higher transmission power or more powerful error-correction coding techniques on the channel and better fault-tolerance techniques on various system components. The second issue of how to contain errors is more intricate and has commonly been overlooked. There are mainly two approaches (or a combination of both) to deal with the problem. The first approach is to add extra redundancy to the compressed data to detect and confine errors. Some simple ways are adding a delimiter pattern at regular intervals and forcing fixed block-size transmission by appending filling 0's, etc. These techniques reduce compression performance. The second approach is to choose data compression schemes which are less susceptible to error propagation. These schemes usually transmit the compressed data in fixed block size or inherently have delimiter patterns in their compressed data streams. We illustrate the interaction between data compression and error-correction coding with an example in a later section.

MODULATION AND DEMODULATION

Modulation

Modulation is the process of transforming signals into waveforms suitable for transmission through a certain type of medium or channel. For example, when transmitting electronic signals through free space with an antenna which converts electronic signals into electromagnetic (EM) waves, the physical dimensions of the antenna aperture are at least of the same order of the wavelength of the EM wave. Therefore, a baseband signal has to modulate a sinusoidal signal at radio frequency (RF), called the carrier, before transmitting it by antennas of reasonable size.

The typical modulation procedure for digital signal transmission starts with a baseband pulse code modulation (PCM) which converts the analog signal into a binary data stream. Depending on the subsequent RF modulation schemes, this binary data stream is fed into different devices. For example, fixed-length blocks of the binary data are used to determine the instantaneous frequencies of the carrier for a frequency-shift-keyed (FSK) system. On the other hand, the binary data drive a pulse generator whose output modulates the subcarrier (if used) and the RF carrier in a typical phase-shift-keyed (PSK) system. There are several different binary data formats (5). For example, non-return-to-zero (NRZ) and bi-phase (also known as the Manchester code) are the ones more commonly used. In some applications, pulse shapes other than the square pulse are selected to represent a binary digit in the data stream. The choice of a particular data format and pulse shape is determined by several factors, for example, considerations of bandwidth efficiency, inherent synchronization features, and the noise immunity of each data format.

PSK modulation has been widely adopted for deep space communications mainly because it is a very efficient type of

modulation in bit-error performance and the resulting signal has a constant envelope which allows the power amplifier to achieve the maximum efficiency by operating at the saturation point. In general, a communication system using the PSK modulation is designed to have a multiple phase-shift-keyed (MPSK) signal of which the number of phasor states, denoted by M , is any power of 2. The increased BER with M for MPSK signals, however, prevents its use when M is large. Furthermore, because of the ease of being decomposed into two orthogonal channels for further signal processing, the MPSK modulated signals commonly used for radio-telemetry are limited to the cases where $M = 4$ or lower, namely, the binary phase-shift-keying (BPSK), the quadrature phase-shift-keying (QPSK) or its variations, such as offset QPSK, unbalanced QPSK, and pulse-shaped QPSK, etc. For some applications, the quadrature amplitude-shift-keying (QAM) modulation is used to allow multilevel signals transmitted on either one or both of the mutually orthogonal channels.

It is possible to include subcarrier(s) for BPSK modulation. The purpose of using subcarrier(s) is to separate data sidebands of different signals and the carrier so that they do not interfere with each other. Whether or not to use subcarrier(s) depends on the mission design. For example, a subcarrier is preferred when a residual carrier component is preserved for spacecraft navigation or certain radio science experiments. On the contrary, it may not be a good idea to keep a subcarrier when transmission power is weak or the bandwidth efficiency becomes a major concern. Two types of subcarriers, the sine wave and the square wave, are commonly used. The sine-wave subcarrier along with the phase-modulated carrier produces fast-decaying data sidebands and, therefore, is recommended for near-Earth space missions where the interference caused by the power spillover to the adjacent frequency bands is the major concern (6). The same recommendation also suggests using the square-wave subcarrier for deep space missions because of the ease of generating a square wave onboard spacecraft and the relatively less stringent bandwidth allocation for this type of mission.

A phase-modulated (PM) telemetry signal is represented mathematically by

$$S_T(t) = \sqrt{2P_T} \sin \left(\omega_c t + \sum_{i=1}^N m_i S_i(t) \right) \quad (5)$$

and raised cosine pulse waveform, are selected where P_T is the total power of the received signal, ω_c is the carrier frequency, m_i is the modulation index associated with the i th data source, and

$$S_i(t) = \begin{cases} \sum_{k=-\infty}^{\infty} d_{k,i} P_i(t - kT), & \text{for PCM/PM (without subcarrier)} \\ \left[\sum_{k=-\infty}^{\infty} d_{k,i} P_i(t - kT) \right] \sin(\omega_{sc_i} t), & \text{for PCM/PSK/PM with a sine-wave subcarrier} \\ \left[\sum_{k=-\infty}^{\infty} d_{k,i} P_i(t - kT) \right] \text{sgn}[\sin(\omega_{sc_i} t)], & \text{for PCM/PSK/PM with a square-wave subcarrier} \end{cases} \quad (6)$$

represents either a normalized baseband waveform (with $d_{k,i} = \pm 1$) or a normalized BPSK modulated subcarrier waveform (at the frequency ω_{sc_i}). $P_i(t)$ is the pulse function of unit power and T is the reciprocal of the symbol rate. Note that, when more than one data sources are included (that is, $N > 1$ in Eq. (5)), only one data source output is allowed to directly modulate the carrier, denoted by its processing order as PCM/PM, and the rest must be put on subcarrier(s) before modulating the carrier, similarly denoted as PCM/PSK/PM in Eq. (6). An exception of this single PCM/PM rule is the code division multiplexing, in which multiple data sources are put directly on the carrier and remain distinguishable because of the unique code sequence (for example, pseudo-random codes or Walsh-Hadamard codes) used by each of them.

Typically, a deep space downlink signal (i.e., a signal sent from spacecraft to Earth) consists of two or more data sources of which, besides the telemetry signal, a ranging signal is included for spacecraft navigational purpose (7). For example, a mathematical expression of a downlink signal consisting of a sinusoidal ranging signal at the frequency ω_1 and a binary telemetry signal of NRZ format, $d(t)$, which is modulated onto a square-wave subcarrier, is given by

$$\begin{aligned} S_T(t) &= \sqrt{2P_T} \sin \{ \omega_c t + m_1 \sin(\omega_1 t) \\ &\quad + m_2 d(t) \text{sgn}[\sin(\omega_{sc} t + \theta_{sc})] + \theta_c \} \\ &= \sqrt{2P_T} \left\{ \begin{aligned} &J_0(m_1) \cos(m_2) \sin(\omega_c t + \theta_c) \\ &+ d(t) J_0(m_1) \sin(m_2) \text{sgn}[\sin(\omega_{sc} t + \theta_{sc})] \cos(\omega_c t + \theta_c) \\ &+ \cos(m_2) \left[\sum_{n=1}^{\infty} 2J_{2n}(m_1) \cos(2n\omega_1 t) \right] \sin(\omega_c t + \theta_c) \\ &+ \cos(m_2) \left\{ \sum_{n=0}^{\infty} 2J_{2n+1}(m_1) \sin[(2n+1)\omega_1 t] \right\} \cos(\omega_c t + \theta_c) \\ &+ d(t) \sin(m_2) \left[\sum_{n=1}^{\infty} 2J_{2n}(m_1) \cos(2n\omega_1 t) \right] \\ &\quad \text{sgn}[\sin(\omega_{sc} t + \theta_{sc})] \cos(\omega_c t + \theta_c) \\ &- d(t) \sin(m_2) \left\{ \sum_{n=0}^{\infty} 2J_{2n+1}(m_1) \sin[(2n+1)\omega_1 t] \right\} \\ &\quad \text{sgn}[\sin(\omega_{sc} t + \theta_{sc})] \sin(\omega_c t + \theta_c) \end{aligned} \right\} \end{aligned} \quad (7)$$

where $J_n(\cdot)$ is the n th order Bessel function and θ_c and θ_{sc} are random carrier and subcarrier phases, respectively, each uniformly distributed over 0 to 2π . The first term in this expression is the residual carrier component which is fully suppressed if the data modulation index m_2 equals $\pi/2$. The second term is the desired data-bearing component containing the telemetry information which needs to be demodulated. The third and the fourth terms contain the ranging information to be extracted separately, and the rest are from intermodulation of telemetry and ranging signals. Typically, on one hand, the ranging modulation index m_1 chosen is small (around 1/2 or smaller) so that the power consumption by this ranging signal is relatively small. On the other hand, the data modulation index selected is large (close to its upper limit $\pi/2$) to ensure that only sufficient

power goes to the residual carrier component and the rest of power is allocated solely for telemetry data transmission. The strategy of optimizing modulation indices for efficient power allocation is discussed in (8).

A QPSK phase-modulated telemetry signal is usually treated as a combination of two orthogonal BPSK signals. Mathematically, a general QPSK (or, more specifically, an unbalanced QPSK) signal takes the form

$$S_T(t) = \sqrt{\alpha P_T} \sin \left(\omega_c t + \sum_{i=1}^M m_{c,i} S_{c,i}(t) \right) + \sqrt{(1-\alpha) P_T} \cos \left(\omega_c t + \sum_{i=1}^N m_{s,i} S_{s,i}(t) \right) \quad (8)$$

where α is the percentage of transmitted power in one of the two channels. When only one binary signal of NRZ format is transmitted on each channel, that is, $M = N = 1$, with the modulation indices $m_{s,1} = m_{c,1} = \pi/2$, the QPSK signal is rewritten as

$$S_T(t) = \sqrt{\alpha P_T} S_{c,i}(t) \cos(\omega_c t) - \sqrt{(1-\alpha) P_T} S_{s,i}(t) \sin(\omega_c t) \quad (9)$$

a combination of two BPSK signals on two orthogonal basis functions. As long as the orthogonality is maintained, these two BPSK signals do not interfere with each other and the bandwidth efficiency, measured as how many bits of information transmitted over a unit bandwidth, of a QPSK signal is twice that of a BPSK signal (5). Several variants of QPSK modulation, including the offset QPSK (OQPSK) and the minimum-shift-keying (MSK), are also commonly used in near-Earth space missions. A detailed description of these modulation schemes is not covered here.

Demodulation

Demodulation is the process of transforming received waveforms back into their original state by reversing the modulation procedure. After traveling through various types of media or channels, the received waveform is corrupted in many ways. For example, it is corrupted by receiver's internally generated noise which is typically modeled as an additive white Gaussian noise (AWGN), or externally introduced interference, such as multipath, fading. Hence, for demodulation, it is important to correctly estimate the vital parameters in the transmitted signal from the corrupted waveform and apply the locally generated reference signals to remove the modulation. The following simple example illustrates how an AWGN corrupted BPSK signal is demodulated. The received signal is given by

$$r(t) = \sqrt{2P_T} \sin \left(\omega_c(t + \tau) + \left(\frac{\pi}{2} \right) \sum_{k=1}^{\infty} d_k P(t + \tau - kT) + \theta \right) + n(t) \quad (10)$$

where τ is the random propagation delay, θ is a uniformly distributed (over 0 to 2π) carrier phase, and $n(t)$ is a noise modeled as an AWGN with a two-sided power spectral density level at $N_0/2$ W/Hz. For the signal of NRZ format, the phase-modulated

signal is equivalent to the product of the carrier and the base-band binary data waveforms, which is rewritten as

$$r(t) = \sqrt{2P_T} \left[\sum_{k=1}^{\infty} d_k P(t + \tau - kT) \right] \cos(\omega_c t + \theta_c) + n(t) \quad (11)$$

where $\theta_c = (\theta + \omega_c \tau)_{\text{mod } 2\pi}$ is the total carrier phase. To demodulate the carrier, the receiver needs to generate a local reference, say $\sqrt{2} \cos(\omega_c t + \phi_c)$, where $\text{thgr}^{\text{circ}}_c$ is an estimate of θ_c . A low-pass filtered version of the product of the local carrier reference and the received signal becomes

$$r'(t) = \sqrt{P_T} \left[\sum_{k=1}^{\infty} d_k P(t + \tau - kT) \right] \cos(\phi_c) + n'(t) \quad (12)$$

where $\phi_c = \theta_c - \text{thgr}^{\text{circ}}_c$ is the phase error between the actual and the estimated carrier phases. For constant or at least slowly varying ϕ_c , the factor $\cos(\phi_c)$ represents a signal amplitude attenuation, which is inevitably translated into degradation in bit-error performance. To make a decision on each of the transmitted bits, say, the i th bit d_i , the resulting signal $r'(t)$ is sent to a matched filter whose operation is mainly to form a product of the input signal and a local replica of the pulse function followed by an integrate-and-dump (I&D) operation. The accurate timing estimate $\text{thgr}^{\text{circ}}$ is very important in this matched filter operation because an error renders integrating across two bits, which reduces the detected symbol (for coded system) or bit (for uncoded system) energy when adjacent bits are of opposite polarities and results in a higher probability of decision error.

Additional signal power degradation due to imperfect sub-carrier synchronization, similar to the carrier case given here, is expected when a subcarrier is used. The power degradation resulting from each of the carrier, subcarrier, and symbol tracking operations is discussed later.

SYNCHRONIZATION

The process of estimating the phase and timing parameters from the incoming noise-corrupted signal and using this information to keep the locally generated reference signal aligned with these estimates and, therefore, with the incoming signal is called synchronization.

As indicated previously, coherent reception and demodulation require phase information about the carrier and subcarrier (if used) and also symbol-timing information. This information must be provided and updated for coherent receivers all the time because they are usually time-varying with the changing characteristics of the channel. Therefore, individual tracking loops which continuously update their estimates of specific parameters are required to track and provide the needed information for a coherent receiver.

Although the tracking of carrier, subcarrier, and symbol timing are individually discussed in the following, one should keep in mind that, strictly speaking, all these loops are effectively coupled together in the sense that no one achieves lock without help from others, except for residual carrier tracking in which a carrier tone is separately tracked. However, in practice, each loop's performance is usually analyzed independently to keep the problem manageable.

It is also important to know that all the tracking loops discussed later are motivated by the maximum *a posteriori* (MAP) estimation which suggests only the open-loop structure of a one-shot estimator. The closed-loop structure derived by differentiating the likelihood function and equating the resulting loop feedback signal (also known as the error signal) to zero is only motivated by the MAP estimation (8).

Carrier Tracking

BPSK. The most commonly used device to track the phase of a sinusoidal signal, for example, the residual carrier component in Eq. (7), is the phase-locked loop (PLL). The PLL is composed of a phase detector, a loop filter, and a voltage-controlled oscillator (VCO) or, in a digital PLL design, a numerically controlled oscillator (NCO). The low-pass component of the phase detector output is a periodic function (of period 2π) of the phase error ϕ_c , which is called the S-curve of the loop. A stable lock point exists at $\phi_c = 0$, one of the zero-crossing points where the S-curve has a positive slope. The phase error for the first-order PLL, which has its loop filter implemented as a constant gain, is a Tikhonov distributed random variable and its probability density function (pdf) is given by (5)

$$p(\phi_c) = \frac{\exp[\rho_{\phi_c} \cos(\phi_c)]}{2\pi I_0(\rho_{\phi_c})} \quad |\phi_c| \leq \pi \quad (13)$$

where $I_k(\cdot)$ denotes the modified Bessel function of order k and $\rho_{\phi_c} = (\sigma_{\phi_c}^2)^{-1}$ is the loop SNR defined as the reciprocal of the phase error variance in radian². The loop SNR for the first-order PLL is expressed by

$$\rho_{\phi_c} = \frac{P_c}{N_0 B_L} \quad (14)$$

where B_L is the loop bandwidth. The detailed description of a PLL is discussed in another article in this encyclopedia and is not to be repeated here.

For the suppressed carrier in which no discrete carrier component appears in its spectrum, the carrier phase, embedded in the data-bearing component as the second term of Eq. (7), must be tracked by the Costas loop. The Costas loop is a phase-tracking loop whose functionality is similar to that of a PLL. Except for the same feedback path comprised of a loop filter and an NCO, a Costas loop has a double-arm loop structure, denoted, respectively, as the in-phase (I) and quadrature (Q) arms, with a phase detector and a low-pass arm filter in each. The incoming signal is first mixed with each of the two locally generated reference signals 90° apart, that is, $\sqrt{2} \sin(\omega_c t + \text{thgr}^{\text{circ}}_c)$ and $\sqrt{2} \cos(\omega_c t + \text{thgr}^{\text{circ}}_c)$, at the corresponding phase detector and then passed through the arm filter. Although the low-pass arm filter is either a passive RC-type filter or an active filter, it turns out that a matched filter (that is, an active filter) is the optimal design. The output of the two arm filters are multiplied, which effectively removes the data modulation, to form the loop feedback signal before it is fed into the loop filter. Because of this multiplication, a Costas loop is actually tracking twice the error phase. Accordingly, the Costas loop has two equally stable lock points at $\phi_c = 0$ and $\phi_c = \pi$, each corresponding to a zero-crossing point in the S-curve (of period π) where the slope is positive. These dual lock point inevitably introduce phase ambiguity such that the demodulated data has

inverted polarity if the loop locks at $\phi_c = \pi$. This 180° phase ambiguity is resolved in several ways. For example, a known sequence pattern is inserted in the transmitted symbol stream from time to time so that the inverted polarity is detected by examining the received sequence pattern. However, the most efficient method is employing a differential encoding scheme in the transmitted data so that the information is kept in the relative phase between adjacent symbols instead of in the absolute phase of each symbol (5). On the receiver side, a corresponding differential decoding scheme is applied to extract the relative phase (or the transmitted information) after the symbol decision. A small penalty in terms of error performance exists for this differential encoding/decoding scheme because one incorrect symbol decision creates two consecutive errors in the relative phase.

The phase error for the first-order Costas loop with I&D arm fitters is similarly found as a Tikhonov distributed random variable and its pdf is given by

$$p(\phi_c) = \frac{\exp\left[\frac{\rho_{\phi_c}}{4} \cos(2\phi_c)\right]}{\pi I_0\left(\frac{\rho_{\phi_c}}{4}\right)} \quad |\phi_c| \leq \frac{\pi}{2} \quad (15)$$

and the associated loop SNR is given by

$$\rho_{\phi_c} = \frac{P_d}{N_0 B_L} \left(1 + \frac{1}{2E_s/N_0}\right)^{-1} \quad (16)$$

where $E_s/N_0 = P_d T/N_0$ is the symbol SNR. Note that the term in the parentheses is usually called the squaring loss, which results from the signal-noise product in the loop feedback signal. At low symbol SNR, the squaring loss is significant.

As discussed previously, transmitted power is allocated to the residual carrier component and data-bearing component by the choice of the modulation index for the telemetry data. It has been proved that a fully suppressed carrier is the best way to maximize data throughput (9). However, if a residual carrier component is desired for purposes other than communications, it is always a dilemma to set this modulation index because, on one hand, sufficient power must be given to the residual carrier so that it is successfully tracked by a PLL, and, on the other hand, the power allocated for data transmission should be kept as high as possible to maximize data throughput. Because the residual carrier and data sidebands are coherently related, a hybrid loop (10) which consists of the phase-locked loop and Costas loop structures is used to exploit this coherence and thereby improve carrier phase tracking in this scenario. This technique is also known as sideband aiding because it utilizes the power in the data-bearing component as the second term of Eq. (7) to help residual carrier tracking.

In the hybrid loop, both error signals from the single-arm PLL structure and the double-arm Costas loop structure are weighted and added together to form an effective loop feedback signal. As a result, there are usually dual lock points for the hybrid loop, that is, $\phi_c = 0$ and $\phi_c = \pi$ and similar to those of a Costas loop. Yet, these two lock points generally are not equiprobable. It can be shown that (11) the lock point at $\phi_c = \pi$ vanishes when the modulation index is smaller than a threshold as a function of the symbol SNR.

With a given modulation index, an optimal relative weight between the PLL and Costas loop portion is derived to minimize

the hybrid-loop tracking jitter. Because the relative tracking performance between a PLL and Costas loop is determined by the relative power allocation and the additional squaring loss incurred in the Costas loop, it is not surprising to find that the optimal weight is a function of the modulation index and the symbol SNR.

QPSK. The carrier tracking of a QPSK signal is usually done by a generalized Costas loop motivated by MAP estimation theory. There are basically two variants of this generalized Costas loop: the polarity-type known as the crossover Costas loop for high SNR scenarios and the squaring-type for low SNR scenarios (7). In the crossover loop, two products are formed by multiplying the hard-limited version of one arm-filter output with the other arm-filter output before they are combined as the loop error feedback. The phase error for the first-order crossover Costas loop is a Tikhonov distributed random variable with pdf given by

$$p(\phi_c) = \frac{2 \exp\left[\frac{\rho_{\phi_c}}{16} \cos(4\phi_c)\right]}{\pi I_0\left(\frac{\rho_{\phi_c}}{16}\right)} \quad |\phi_c| \leq \frac{\pi}{4} \quad (17)$$

The associated loop SNR for this first-order cross-over loop is

$$\rho_{\phi_c} = \frac{P_d}{N_0 B_L} \times \left\{ \frac{\left[\operatorname{erf}\left(\sqrt{\frac{E_s}{2N_0}}\right) - \sqrt{\frac{2}{\pi}} \left(\frac{E_s}{N_0}\right) \exp\left(-\frac{E_s}{2N_0}\right) \right]^2}{1 + \frac{E_s}{N_0} - \left[\sqrt{\frac{2}{\pi}} \exp\left(-\frac{E_s}{2N_0}\right) + \sqrt{\frac{E_s}{N_0}} \operatorname{erf}\left(\sqrt{\frac{E_s}{2N_0}}\right) \right]^2} \right\} \quad (18)$$

where the error function $\operatorname{erf}(\cdot)$ and its complimentary $\operatorname{erfc}(\cdot)$ are defined as

$$\operatorname{erf}(x) = 1 - \operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\alpha^2} d\alpha$$

Note that, similar to Eq. (16) of the Costas loop, the term in the braces in Eq. (18) is the squaring loss of the crossover Costas loop. The squaring-type loop has a different squaring loss which is smaller than that of its polarity-type sibling in low SNR region. More details about the crossover Costas loop can be found in (13). Another alternative in tracking a QPSK signal is to use the demod-remod quadriphase tracking loop which can be viewed as a fourth-power loop with a multiplication done at the IF level (7).

Subcarrier Tracking

Subcarrier tracking is almost identical to suppressed carrier tracking for BPSK signals because there is no residual tone left for the binary-phase-shift-keyed subcarrier. The Costas loop is used here to remove the data modulation and a squaring loss associated with this process is inevitable. However, depending on the use of a sine-wave or square-wave subcarrier, tracking performance is quite different. For the sine-wave subcarrier, there

is no difference between its tracking and that of a suppressed carrier. On the contrary, additional improvement in the square-wave subcarrier tracking is realized by using a time-domain windowing function on the quadrature arm (13). In this case, the windowing function around the midphase transition of the Q-arm reference signal is treated as an approximation of the time-domain derivative of its I-arm counterpart. According to the derivation of the MAP estimation, which implies the existence of an optimal open-loop structure when one of the I-arm and Q-arm reference signals is the derivative of the other, the resulting loop SNR is greatly improved by shrinking the window size. The first-order loop SNR is given by

$$\rho_{\phi_{sc}} = \left(\frac{2}{\pi}\right)^2 \frac{P_d}{N_0 B_L} \left(\frac{1}{W_{sc}}\right) \left(1 + \frac{1}{2E_s/N_0}\right)^{-1} \quad (19)$$

where W_{sc} is the quadrature window size (between 0 and 1) relative to a subcarrier cycle. It is clear that the loop SNR is inversely proportional to the window size. However, using a small window inevitably reduces the loop's pull-in range and raises the issue of loop stability. A reasonable window size of one-quarter or one-eighth is usually used to provide a 6 to 9 dB improvement in loop SNR.

No such improvement from quadrature windowing is realized for a sine-wave Costas loop of which the I-arm and Q-arm reference signals are two sine functions separated by 90° and, therefore, have the derivative relationship between them as suggested by the MAP estimation. Applying a quadrature window in this case actually destroys the derivative relationship and renders inferior loop performance.

Symbol-Timing Tracking

Symbol synchronization has a direct impact on the data detection process because inaccurate symbol timing reduces the probability of making a correct decision. Although a separate channel may be used to send timing signals for synchronization, to extract the synchronization information directly from the data-bearing signal has the advantage that no additional power and frequency spectrum are required. Of course, to successfully extract symbol timing information from the transmitted symbol stream relies on the presence of adequate symbol transitions (zero-crossings).

The data transition tracking loop (DTTL) is widely used for symbol synchronization. Similar to the Costas loop, DTTL has a double-arm loop structure with a hard decision followed by a transition detector in its in-phase arm and a delay in its quadrature arm to keep signals on both arms properly aligned. It is important to note that the term "in-phase" refers to an operation synchronous with the timing of the received symbols and, therefore, the I-arm phase detector becomes a matched filter integrating from one symbol epoch to the next. The Q-arm phase detector performs another integration within a window, which is of a size W_{sym} (between 0 and 1) relative to the symbol interval and centered at the symbol epoch, causing the midpoint of the Q-arm integration interval offset by a half-symbol from its I-arm counterpart. Similar to square-wave subcarrier tracking, the time-domain windowing function on the quadrature arm improves the tracking performance but inevitably raises the issue of loop stability at the same time (5).

The DTTL has a single stable lock point at $\phi_{sym} = 0$. The phase error for the first-order DTTL is a Tikhonov distributed

random variable and its pdf is given by

$$p(\phi_{\text{sym}}) = \frac{\exp(\rho_{\phi_{\text{sym}}} \cos(\phi_{\text{sym}}))}{2\pi I_0(\rho_{\phi_{\text{sym}}})} \quad |\phi_{\text{sym}}| \leq \pi \quad (20)$$

with the corresponding loop SNR given as

$$\rho_{\phi_{\text{sym}}} = \frac{1}{(2\pi)^2} \frac{P_d}{N_0 B_L} \left(\frac{1}{W_{\text{sym}}} \right) \times \left\{ \frac{2 \left[\text{erf} \left(\sqrt{\frac{E_s}{N_0}} \right) - \frac{W_{\text{sym}}}{2} \sqrt{\frac{1}{\pi}} \left(\frac{E_s}{N_0} \right) \exp \left(-\frac{E_s}{N_0} \right) \right]^2}{1 + \frac{W_{\text{sym}}}{2} \left(\frac{E_s}{N_0} \right) - \frac{W_{\text{sym}}}{2} \left[\sqrt{\frac{1}{\pi}} \exp \left(-\frac{E_s}{N_0} \right) + \sqrt{\frac{E_s}{N_0}} \text{erf} \left(\sqrt{\frac{E_s}{N_0}} \right) \right]^2} \right\} \quad (21)$$

SYMBOL SNR DEGRADATION

Symbol SNR degradation is the direct cause of poor bit-error performance and is translated into the telemetry system loss as seen in the next section when the receiver performs a hard decision on each demodulated symbol. When no hard decision is performed by the receiver, symbol SNR degradation directly affects decoder performance because the demodulated symbols, called soft symbols, are fed to the decoder without going through a hard-decision device.

Because of the difficulty of analyzing the coupled-carrier, subcarrier, and symbol-tracking loops, SNR degradation of the demodulated symbol (for a coded system) or bit (for an uncoded system) caused by imperfectly synchronized references is usually approximated as a product of degradation factors of the individual loops, each factor being derived on the basis of assuming perfect tracking in other loops. The overall degradation, conditioned on the corresponding phase errors, for the telemetry signal given in Eq. (7) is expressed by

$$\begin{aligned} D_{\text{SNR}}(\phi_c, \phi_{\text{sc}}, \phi_{\text{sym}}) &= D_c(\phi_c) D_{\text{sc}}(\phi_{\text{sc}}) D_{\text{sym}}(\phi_{\text{sym}}) \\ &= [\cos(\phi_c)]^2 \left[1 - \frac{2}{\pi} |\phi_{\text{sc}}| \right]^2 \left[1 - \frac{|\phi_{\text{sym}}|}{\pi} + \frac{\phi_{\text{sym}}^2}{2\pi^2} \right] \end{aligned} \quad (22)$$

where $D_c(\phi_c)$, $D_{\text{sc}}(\phi_{\text{sc}})$ and $D_{\text{sym}}(\phi_{\text{sym}})$ are the degradation factors for the imperfect carrier, subcarrier, and symbol (or bit) synchronization, respectively. Hence, the averaged symbol (or bit) SNR degradation due to imperfect synchronization becomes

$$(\overline{D_{\text{SNR}}})_{\text{dB}} = (\overline{D_c})_{\text{dB}} + (\overline{D_{\text{sc}}})_{\text{dB}} + (\overline{D_{\text{sym}}})_{\text{dB}} \quad (23)$$

where $\overline{D_c}$, $\overline{D_{\text{sc}}}$ and $\overline{D_{\text{sym}}}$ are the averaged power degradation factors obtained by averaging over the corresponding Tikhonov distributed phase errors.

In addition to the degradation caused by imperfect synchronization, it is also important to know that there are other

sources of SNR degradation, for example, the intermodulation terms and possible interference from the ranging signal found in Eq. (7), and the subcarrier and symbol waveform distortion introduced by the bandlimited channel.

BIT-ERROR PERFORMANCE (UNCODED SYSTEM) AND TELEMETRY SYSTEM LOSS

Telemetry information is extracted from the demodulated data stream by a symbol decision process. For binary signals, it is typically a hard-limiting decision on an AWGN corrupted, antipodal, random variable. For an uncoded system, the bit- (or symbol) error probability of a BPSK signal is well known as

$$P_b = \int_{-\pi}^{\pi} \frac{1}{2} \text{erfc} \left[\sqrt{\frac{E_b}{N_0}} \cos(\phi_c) \right] p(\phi_c) d\phi_c \quad (24)$$

where $p(\phi_c)$ is the pdf of the carrier phase error given in Eq. (13), when the carrier is tracked by a PLL so that no phase ambiguity exists. However, with a fixed-loop SNR, an irreducible error probability exists no matter how large the bit SNR. This irreducible error probability is characterized solely by the carrier tracking loop SNR and, for a given loop bandwidth, is reduced only by allocating more power to the residual carrier component which serves no purpose in transmitting telemetry except being tracked by PLL.

For suppressed carrier tracking of a BPSK signal by the Costas loop, the phase ambiguity exists and must be resolved. The bit-error probability for a special case of perfect phase ambiguity resolution (say, by other means, such as a periodically inserted known sync pattern) is given by

$$P_b = \int_{-\pi/2}^{\pi/2} \frac{1}{2} \text{erfc} \left[\sqrt{\frac{E_b}{N_0}} \cos(\phi_c) \right] p(\phi_c) d\phi_c \quad (25)$$

where $p(\phi_c)$ is the pdf in Eq. (15).

If a differential coding scheme is utilized to resolve the phase ambiguity, the bit-error probability becomes (7)

$$\begin{aligned} P_b &= \int_{-\pi/2}^{\pi/2} \text{erfc} \left[\sqrt{\frac{E_b}{N_0}} \cos(\phi_c) \right] \\ &\times \left\{ 1 - \frac{1}{2} \text{erfc} \left[\sqrt{\frac{E_b}{N_0}} \cos(\phi_c) \right] \right\} p(\phi_c) d\phi_c \end{aligned} \quad (26)$$

where $p(\phi_c)$ is the pdf in Eq. (15). No irreducible error probability exists in the suppressed carrier tracking because the tracking loop SNR for a fixed loop bandwidth and bit duration product increases with the bit SNR.

So far, only the impact of bit-error probability from carrier tracking has been discussed and one can find the SNR degradation from imperfect carrier tracking, that is, $\cos^2(\phi_c)$ appears repeatedly in Eqs. (24)–(26). When the overall impact of bit-error performance from all levels of imperfect tracking, including carrier, subcarrier, and symbol, is considered, the bit-error probability becomes a threefold integration involving the

overall symbol SNR degradation given in Eq. (22). For example,

$$P_b = \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \operatorname{erfc} \left[\sqrt{\frac{E_b}{N_0} D_{\text{SNR}}(\phi_c, \phi_{sc}, \phi_{\text{sym}})} \right] p(\phi_c) p(\phi_{sc}) p(\phi_{\text{sym}}) d\phi_c d\phi_{sc} d\phi_{\text{sym}} \quad (27)$$

where $p(\phi_{sc})$ and $p(\phi_{\text{sym}})$ are Tikhonov distributed pdfs of subcarrier and symbol phase errors, respectively. In fact, $p(\phi_{sc})$ takes the form of Eq. (15) of the Costas loop and $p(\phi_{\text{sym}})$ is given by Eq. (20). Note that the product of pdfs of individual phase errors is used in lieu of the hard-to-establish joint pdf from the coupled loops.

For QPSK signals, the bit-error probability is given by (7)

$$P_b = \int_{-\pi/4}^{\pi/4} \left(\frac{1}{4} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} [\cos(\phi_c) - \sin(\phi_c)] \right\} + \frac{1}{4} \operatorname{erfc} \left\{ \sqrt{\frac{E_b}{N_0}} [\cos(\phi_c) + \sin(\phi_c)] \right\} \right) p(\phi_c) d\phi_c \quad (28)$$

where $p(\phi_c)$ is the pdf in Eq. (17).

The telemetry system loss is defined as a loss factor $L \geq 0$ dB, which represents the amount of additional bit SNR required for a lossy system to meet the same bit-error performance of a perfectly synchronized system. For example, the bit-error probability for a perfectly synchronized BPSK or QPSK system is given by

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \quad (29)$$

and, therefore, the required bit SNR for this ideal system at a given bit-error probability P_b^* is expressed by

$$\left(\frac{E_b}{N_0} \right)^* = [\operatorname{erfc}^{-1}(2P_b^*)]^2 \quad (30)$$

The lossy system with a system loss L needs L times as much bit energy, namely, $E_b/N_0 = L(E_b/N_0^*)$, to achieve the same bit-error probability or, in other words, to compensate for the loss incurred within.

ADVANCED TOPICS

Antenna Arraying

With recent space missions moving toward high data rate and low transmitting power operations, combining signals from several antennas to improve the effective SNR becomes the only viable option when the existing technologies of building larger single-aperture antenna and lowering the system noise temperature are pushed to their limits.

Three arraying techniques (14) are briefly discussed here: symbol-stream combining, baseband combining and full spectrum combining, each combining signals at a different stage of signal processing.

For symbol-stream combining, each participating antenna performs carrier, subcarrier, and symbol synchronization individually. Then the symbols at each receiver output are combined, with the appropriate weights, to form the final symbols for detection or decoding. This scheme has the advantage of a small combining loss. It is also suitable for real-time combining from intercontinental antenna sites because combining is performed at a relatively low processing rate, that is, the symbol rate. The disadvantage is that each antenna needs a full set of receiver hardware and must lock on the signal individually.

In baseband combining, each antenna needs to lock on and remove the (residual) carrier by itself. Then the resulting baseband signals, including the data-modulated subcarrier, are combined for further synchronization and demodulation. The advantage of this scheme is that less hardware is required because only a single set of subcarrier and symbol tracking devices is needed to process the combined signal. The disadvantage is that each antenna still must lock on, at least, the carrier, individually.

In full spectrum combining, signals are combined at an intermediate frequency (IF). Before they are combined, the relative time delay and phase difference must be properly estimated and compensated for so that signals are coherently combined. Then the resulting IF signal is directed to a single receiver for further synchronization and demodulation. The advantage of this scheme is that only one of the participating antennas must lock on the combined IF signal, which allows including smaller antennas in this arraying scheme even though they cannot lock on the signal. The disadvantage is the very large transmission or recording bandwidth required to carry the IF signals through the networked antenna sites for combination.

Besides the above-discussed arraying techniques, a scheme called the carrier arraying, which employs coupled carrier tracking devices from participating antennas should be mentioned here. This scheme by itself does not combine the signals and, thus, must be operated with symbol-stream or baseband combining to array the telemetry. In a carrier array scenario, a large master antenna generally locks on the signal by itself and then helps other smaller antennas to track by estimating and removing the signal dynamics in their input.

Buffered Telemetry Processing

The Deep Space Communications Complex (DSCC) Galileo Telemetry (DGT) is developed and implemented by the Jet Propulsion Laboratory to support the Galileo S-Band Mission. Many advanced technologies have been developed for this mission to cope with the failure to fully deploy the high-gain antenna of the Galileo spacecraft, making itself a showcase of future signal processing technologies in the radiotelemetry field. In the following, selected key features of the DGT and the technologies behind these features are briefly described to illustrate the concept of buffered telemetry processing in which telemetry is recorded, processed, and re-processed to minimize data loss in space missions operated with low link margins.

DGT is composed of four major subsystems, the full spectrum recorder (FSR), the full spectrum combiner (FSC), the buffered telemetry demodulator (BTD), the feedback concatenated decoder (FCD), and other control functions to coordinate the operations of these subsystems. Except for the FSR, the rest of the DGT is implemented in software and can be run on general-purpose workstations, which allows greater

flexibility of signal processing without expensive custom-made hardware.

The FSR downconverts the RF signal to IF for digitization and then further open-loop downconverts each data sideband to the baseband, individually and coherently, before it is sampled and recorded. This significantly reduces the required bandwidth for transmission through the intercontinental antenna network because the processing rate is linked to the symbol rate, instead of the much higher subcarrier frequency.

The recorded signals (residual carrier and data sidebands centered at the first four harmonics of the square-wave subcarrier are kept for the Galileo S-Band Mission) from arrayed antennas are combined by the FSC, which estimates and adjusts the time delay and phase for each recorded sideband coherently to a reference point chosen as the center of the Earth, and then combines the time- and phase-aligned signals from arrayed antennas to form an enhanced signal. The combined telemetry is archived and transferred to the BTM upon request for synchronization and demodulation.

The BTM, known as the software receiver, is the signal processing core of DGT, which provides acquisition, synchronization, demodulation, and miscellaneous monitoring functions through its carrier, subcarrier, symbol-tracking loops and associated lock indicators (15). In the BTM, the individually combined data sidebands are processed coherently and then are synthesized to form an equivalent signal as if it were a single signal processed by a regular receiver. The end product of the BTM is a demodulated soft symbol stream which is written to a file and transferred to the FCD upon request for decoding and decompression.

Because the FSR/FSC combined data are recorded on tape, the BTM is actually designed to be able to work on any segment of data off-line in either direction, forward or backward in time. In fact, with the availability of multiple CPU workstations, simultaneous BTM sessions are initiated on different segments of data. For example, one session is dedicated to process real-time samples forward (in time) while the others reprocess other recorded segments as needed. Then the soft symbol streams from these simultaneous sessions are merged into a single stream because each of them is properly time tagged. By taking advantage of the flexibility in software implementation, many noncausal signal processing techniques can be performed to process or reprocess the data to further enhance the quality of the telemetry. One important feature of the BTM is the so-called gap-closure processing (16) which greatly reduces possible data loss due to receiver acquisition, resynchronization, and loss of lock.

The need to reprocess a segment of sampled data arises from the failure of the BTM to maintain the in-lock status in any of its tracking loops or the failure of the FCD to properly decode the soft symbols. A segment of sampled data on which the telemetry cannot be extracted reliably is called a gap, and the processing of a gap to extract any valid information not available when that segment of samples was first processed is called gap-closure processing. Gaps caused by acquisition are found in the beginning of each pass or at instants when the receiver drops out of lock, whereas gaps generated by cycle slips in one of the loops occur randomly in a pass. Along with its demodulation efforts, the BTM tracks its internal states, including the lock indicators, the symbol SNR estimates, and the state variables inside the loop filter and the NCO for all three loops. These state variables are recorded at fixed intervals as

checkpoints and, with them, a software receiver is easily restored to its state at a checkpoint immediately before or after a gap. By estimating the parameters of a phase process in a region near a restored checkpoint where the phase tracking was successfully carried out, gap closure processing can start from this checkpoint and move into the gap. Two configurations, one for closed-loop and the other for open-loop, are used here. The closed-loop configuration needs to initiate the loop filter with phase parameter estimates in a particular way, so that, when the loop is closed and starts to track at the checkpoint, the loop virtually starts immediately with steady state tracking. For a relatively stable phase process and a gap of small size, an open-loop configuration is applied by using an estimated phase profile as the reference without resorting to a loop operation. Both configurations are applied to gap-closure processing in either direction, forward or backward in time, because the buffered data can be processed in either order. This is especially useful when a gap occurs at the beginning of a track so that all of the available checkpoint information is from the region behind this gap.

Another useful feature of BTM is its capability of seamless tracking through symbol rate changes. The reason for changing the symbol rate during a pass is to take advantage of the changing G/T figure as the elevation angle of an antenna changes in a pass. With a higher elevation angle, an antenna has a higher G/T figure and supports a higher symbol rate when the symbol SNR is fixed. The software implementation of BTM handles symbol rate changes without dropping a lock on symbol timing, as long as the rate changes follow a set of specific rules and their schedule is roughly known in advance.

The FCD is a subsystem that performs error-correction decoding and data decompression in the DGT. Implemented in software on a multiprocessor workstation, it employs a feedback mechanism that passes intermediate decoding information from the outer code of the concatenated code to the inner code to facilitate multipass decoding which achieves a final bit error rate of 10^{-7} at a 0.65 dB bit SNR. The architecture and the detailed operations of the FCD are described in the next section.

Advanced Source and Channel Coding for Space Applications

In this section, we use the Galileo S-Band Mission again as an example to illustrate the application of advanced source and channel coding schemes to enhance telemetry return (17). First, using the integer cosine transform (ICT) for lossy image compression is briefly explained. Then, an advanced error-correction coding scheme used to protect the heavily edited and compressed data is discussed, followed by a discussion of the interaction between data compression and error-control (containment/detection/correction) processes.

Galileo's Image-Compression Scheme. Galileo image compression is a block-based lossy image-compression algorithm that uses an 8×8 ICT. The ICT was first proposed in (18), and was streamlined and generalized in (19,20). It is an integral approximation of the popular discrete cosine transform (DCT), which is regarded as one of the best transform techniques in image coding. Its independence from the source data and the availability of fast transform algorithms make the DCT an attractive candidate for many practical image processing applications. In fact, the ISO/CCITT standards for image

processing in both still-image and video transmissions include the two-dimensional DCT as a standard processing component in many applications.

The elements in an ICT matrix are small integers with sign and magnitude patterns resembling those of the DCT matrix. Besides, the rows of the ICT matrix are orthogonal. The integral property eliminates real multiplication and real addition operations, thus greatly reducing computational complexity. The orthogonality property ensures that the inverse ICT has the same transform structure as the ICT. Notice that the ICT matrix is only required to be orthogonal, but not orthonormal. However, any orthogonal matrix may be made orthonormal by multiplying it by an appropriate diagonal matrix. This operation is incorporated in the quantization (dequantization) stage of the compression (decompression), thus sparing the ICT (inverse ICT) from floating-point operations and, at the same time, preserving the same transform structure as in the floating-point DCT (inverse DCT). The relationship between the ICT and DCT guarantees efficient energy packing and allows the use of fast DCT techniques for the ICT. The ICT matrix used in the Galileo mission is given as follows:

$$\begin{bmatrix} 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 5 & -3 & -2 & -1 & -1 & -2 & -2 & -5 \\ 3 & -1 & -1 & -3 & -3 & -1 & -1 & -3 \\ 3 & -1 & -5 & -2 & -2 & -5 & -1 & -3 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 2 & -5 & -1 & -3 & -3 & -1 & -5 & -2 \\ 1 & -3 & -3 & -1 & -1 & -3 & -3 & -1 \\ 1 & -2 & -3 & -5 & -5 & -3 & -2 & -1 \end{bmatrix}$$

Figure 1 shows the rate-distortion performance of the ICT scheme compared with the JPEG scheme. Simulation results indicate that the difference in performance between the use of floating-point DCT and the ICT is insignificant.

Galileo's Error-Correction Coding Scheme. The Galileo error-correction coding scheme uses a (255,k) variable redundancy

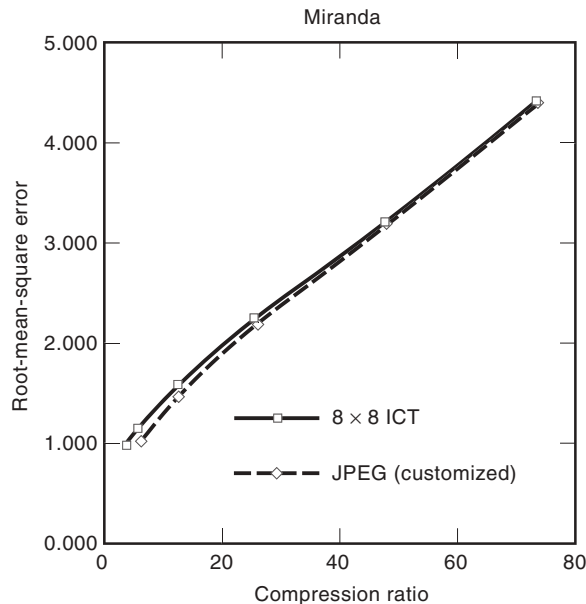


Figure 1. Rate-distortion performance of ICT.

RS code as the outer code and a (14,1/4) long constraint-length convolutional code as the inner code. The RS codewords are interleaved to depth 8 in a frame. The redundancy profile of the Reed–Solomon codes is (94, 10, 30, 10, 60, 10, 30, 10). The staggered redundancy profile was designed to facilitate the novel feedback concatenated decoding strategy (21,22). This strategy allows multiple passes of channel symbols through the decoder. During each pass, the decoder uses the decoding information from the RS outer code to facilitate the Viterbi decoding of the inner code in a progressively refined manner. The FCD is implemented in software on a multiprocessor workstation. The code is expected to operate at a bit signal-to-noise ratio of 0.65 dB at a bit error rate of 10^{-7} . Figure 2 shows a schematic of the FCD architecture. In this article, only the implementation and operational aspects of the FCD task are discussed. The FCD novel node/frame synchronization scheme is discussed in (23) and its code selection and performance analysis are discussed in detail in (24).

The (255,k) Variable-Redundancy Reed–Solomon Code. All RS codes for the Galileo mission use the same representation of the finite field $GF(256)$. Precisely, $GF(256)$ is the set of elements

$$GF(256) = \{0, a^0, a^1, a^2, \dots, a^{254}\} \quad (31)$$

where a , by definition, is a root of the primitive polynomial

$$p(x) = x^8 + x^7 + x^2 + x + 1 \quad (32)$$

[i.e., $p(a) = 0$].

In the encoding/decoding process, each power of a is represented as a distinct nonzero 8-bit pattern. The zero byte is the zero element in $GF(256)$. The basis for $GF(256)$ is descending powers of a . Note that this is the conventional representation, not Berlekamp's dual basis (25). The RS generator polynomial is defined as

$$g(x) = \prod_{i=0}^{n-k-1} (x - a^{\beta(i+L)}) = \sum_{i=0}^{n-k} g_i x^i \quad (33)$$

where n denotes the codeword length in bytes, k denotes the number of information bytes, and a^b is a primitive element of $GF(256)$. The parameter b is chosen in some applications to minimize the bit-serial encoding complexity. Because the Galileo RS encoders are implemented in software, there is little advantage in preferring a particular value of b . The parameter L is chosen so that the coefficients of $g(x)$ are symmetrical. This reduces the number of Galois field multiplications in encoding by nearly a factor of 2.

The Galileo mission utilizes four distinct RS codes. We define $RS(n,k)$ as an RS code which accepts k data bytes as input and produces n bytes as a code word, where $n > k$. An $RS(n,k)$ code corrects t errors and s erasures if $2t + s \leq n - k$. These codes are referred to as $RS(255,161)$, $RS(255,195)$, $RS(255,225)$, $RS(255,245)$. Specifically, the parameters b and L of these four codes are as follows:

- $RS(255,161)$ $b = 1, L = 81$
- $RS(255,195)$ $b = 1, L = 98$
- $RS(255,225)$ $b = 1, L = 113$
- $RS(255,245)$ $b = 1, L = 123$

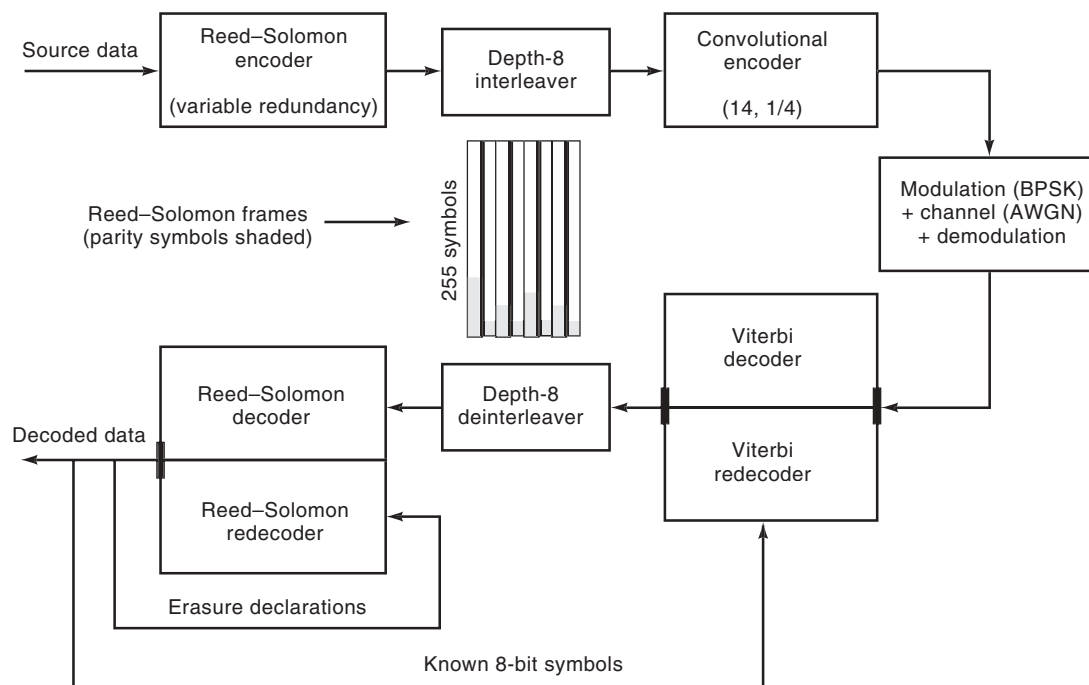


Figure 2. Schematic of the FCD.

These RS codes, being interleaved to depth 8, are arranged in a transfer frame as shown in Fig. 2. The RS decoders use a time-domain Euclid algorithm to correct errors and erasures. The details of the decoding algorithm are discussed in (26).

The (14,1/4) Convolutional Code and Its Parallel Viterbi Decoder. The (14,1/4) convolutional code used for the Galileo mission is the concatenation of a software (11,1/2) code and an existing hardware (7,1/2) code, which is hardwired in the Galileo Telemetry Modulation Unit (TMU), and by the processing speed of the ground FCD. The generator polynomials of the (11,1/2) code and the (7,1/2) code in octal are (3403, 2423) and (133, 171), respectively. The generator polynomials of the equivalent (14,1/4) code are (26042, 36575, 25715, 16723).

The Viterbi decoder for the (14,1/4) code is implemented in software in a multiprocessor workstation with shared memory architecture. The use of a software decoder is possible because of the slow downlink data rate of the Galileo S-Band Mission. The advantages of a software-based decoder are that the development cost is low and it allows the flexibility to perform feedback concatenated decoding. We examined two different approaches to parallelize the Viterbi algorithm: (1) state-parallel decomposition in which each processor is equally loaded to compute the add-compare-select operations per bit and (2) round-robin frame decoding that exploits the multiple processors by running several complete but independent decoders for several frames in parallel. Our early prototypes indicate that the first approach requires a substantial amount of interprocessor synchronization and communication and this greatly reduces the decoding speed. The second approach requires much less synchronization and communication because each processor is now an entity independent of the others. The performance scaling is nearly perfect. We chose the round-robin approach for the FCD Viterbi decoder. The details of the FCD software Viterbi decoder implementation are described in (27).

Redecoding. Redecoding, as shown in Fig. 2, uses information fed back from code words successfully decoded by the RS decoder to improve the performance of Viterbi decoding. A correctly decoded RS bit forces the add-compare-select operation at each state to select the path that corresponds to the correct bit. Thus the Viterbi decoder is constrained to follow only paths consistent with the correct path. The Viterbi decoder is much less likely to choose a long erroneous path because any path under consideration is pinned to coincide with the correct path at the locations of the known symbols. Each RS frame is decoded with four feedback passes. In the first pass, only the first code word RS(255,161) is decoded. In the second pass, the fifth code word RS(255,195) is decoded. In the third pass, the third and seventh code words RS(255,225) are decoded, and finally, in the fourth pass, the second, fourth, sixth, and eighth code words RS(255,245) are decoded. During each pass, the decoder uses the decoding information from the Reed-Solomon outer code to facilitate Viterbi decoding of the inner code in a progressively refined manner. The details of the FCD redecoding analysis are given in (24).

Interaction Between Data Compression and Error Control Processes. Packet loss and other uncorrectable errors in a compressed data stream cause error propagation, and the way the error propagates depends on the compression scheme. To maximize the scientific objectives with the limited transmission power of the low-gain antenna used in the Galileo S-Band Mission, most of the data (image and nonimage) are expected to be heavily edited and compressed. These valuable compressed data must be safeguarded against catastrophic error propagation caused by packet loss and other unforeseeable errors.

The ICT scheme for solid-state imaging (SSI) data is equipped with a simple but effective error-containment strategy. The idea is to insert synchronization markers and

counters at regular intervals to delimit uncompressed data into independent blocks so that, in case of packet loss and other anomalies, the decompressor searches for the next available synchronization marker and continues to decompress the rest of the data. In this case, the interval chosen is eight lines of uncompressed data. The error-containment strategy guarantees that error propagation does not go beyond the compressed code block where errors reside. Other options to prevent error propagation are also considered, but these options usually result in great onboard implementation complexity or excessive downlink overhead. For example, a self-synchronizing feature in Huffman code may be used to contain errors, but it is difficult to implement. A packetizing scheme with varying packet sizes may also be used to contain errors (by matching packet boundaries and the compressed data block boundary), but the packet headers introduce excessive downlink overhead in SSI data.

The SSI ICT error-containment scheme works as follows. On the compression side, every eight lines of data are compressed into a variable-length, compressed data block. The dc (steady-state bias) value is reset to zero at the start of each compressed data block, thus making every block independent of the others. A 25-bit synchronization marker and a 7-bit modulo counter are inserted at the beginning of every compressed data block. The sync marker is chosen to minimize the probability of false acquisition in a bursty channel environment. The 25-bit synchronization marker pattern is 024AAAB in hex. Simulation results indicate that this synchronization marker gives a probability of false acquisition of less than 10^{-8} . The decompression scheme consists of two program modules, the SSI ICT decompression module and the error detection/sync module. The SSI ICT decompression module reconstructs the data from the compressed data stream, and the error detection/sync module checks the prefix condition of the Huffman codes to detect any anomaly. When an anomaly is detected, a synchronization marker search is initiated to find the next available one. Decompression resumes from there on, and the reconstructed blocks are realigned by using the modulo counter. The corrupted portion of the data is flagged and reported.

The downlink overhead of the SSI ICT error-containment scheme is a function of compression ratio (CR) and image width W. It is measured by the percentage of sync data (sync marker and counter) compared to the compressed data and is given by the following equation:

$$\frac{4 \times \text{CR}}{8 \times W}$$

For example, an 800×800 SSI image has the following overhead as a function of the compression ratio:

Compressed Ratio	Overhead
2	0.00125
4	0.00250
8	0.00500
16	0.01000

Multiple Spacecraft Support

Traditionally, every spacecraft is supported by one of the ground antennas for its uplink and downlink. This dedication requires efficient scheduling of the resources on the ground,

including hardware, software, and personnel. With more and more concurrent missions, the need for multiple spacecraft support by a single ground antenna to alleviate the scheduling problem becomes evident. For example, several proposed future missions to Mars by various joint efforts of international space agencies will place more than a dozen spacecraft, including orbiters, landers, and rovers, on or around Mars in the next 10 years. For these missions, it is highly possible that more than one spacecraft will come within the same beam width of a single ground antenna, and it constitutes the opportunity to communicate with them by using this single antenna with a considerable amount of operational cost saving over multiple antennas. In multiple spacecraft support, a telecommand uplink from a single ground antenna will be shared by the supported spacecraft, and multiple telemetry downlinks originated from these spacecraft will also have to be established by a single ground antenna.

Several options have been studied to support this multiple spacecraft scenario (28). The most straightforward (and the most inefficient) option is to carefully assign different sub-carrier frequencies to the supported spacecraft, allowing sufficient guard band to accommodate Doppler effects and tolerating some degree of spectrum overlapping in data sidebands in exchange for more simultaneous support. This method requires very tedious planning and is extremely inflexible when facing a dynamic scenario.

Another option is to redesign the spacecraft transponder so that the coherent turnaround ratio (TAR), which specifies the uplink to downlink carrier frequency ratio, is programmable. Each supported spacecraft receives its unique TAR from the uplink commands. As a result, different spacecraft will be instructed to use different downlink carrier frequencies because their TARs are distinct. Currently, a new digital transponder, known as the Small Transponder Modem developed by the Jet Propulsion Laboratory, has such a built-in feature.

The third option is to use code division multiple access (CDMA) techniques similar to those used in commercial mobile cellular systems. CDMA offers far more simultaneous support than that of the previous two options. However, to support various types of spacecraft, the power dissimilarity problem between weak rover and strong orbiter signals has to be properly solved to avoid severe performance degradation for weaker signals. This option may be the best choice when more and more multiple spacecraft support scenarios emerge in the future.

BIBLIOGRAPHY

1. B. Sklar, *Digital Communications: Fundamentals and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1988.
2. J. Ziv and A. Lempel, A universal algorithm for sequential data compression, *IEEE Trans. Inf. Theory*, **IT-23**: 337–343, 1977.
3. J. Ziv and A. Lempel, Compression of individual sequences by variable rate coding, *IEEE Trans. Inf. Theory*, **IT-24**: 530–536, 1978.
4. C. E. Shannon, A mathematical theory of communication, *M. D. Comput.*, **14** (4): 306–317, 1997.
5. W. C. Lindsey and M. K. Simon, *Telecommunication System Engineering*, Englewood Cliffs, NJ: Prentice-Hall, 1973.
6. International Consultative Committee for Space Data Systems, *CCSDS, Recommendations for Space Data System Standards, Radio Frequency and Modulation Systems, part I, Earth Stations Spacecraft*, CCSDS 401.0-B, Blue Book, 1994.

7. J. H. Yuen, *Deep Space Telecommunications Systems Engineering*, New York: Plenum, 1983.
8. T. M. Nguyen, Technique to select the optimum modulation indices for suppression of undesired signals for simultaneous range and data operations, *IEEE Trans. Electromagn. Compat.*, **32**: 7–19, 1990.
9. M. K. Simon and S. Million, *Residual versus suppressed carrier coherent communications*, TDA Progress Rep. 42-127: July–September 1996, Jet Propulsion Laboratory, Pasadena, November 15, 1996.
10. H. Tsou, M. K. Simon, and S. M. Hinedi, Closed loop carrier phase synchronization techniques motivated by likelihood functions, *1994 IEEE Int. Conf. Commun. Conf. Rec.*, **2**: 1994, pp. 934–939.
11. M. K. Simon et al., The performance of a coherent residual carrier PSK system using hybrid carrier phase synchronization, *1996 IEEE Int. Conf. Commun. Conf. Rec.*, **2**: 1996, pp. 1275–1280.
12. J. K. Holmes, *Coherent Spread Spectrum Systems*, New York: Wiley, 1982.
13. W. J. Hurd and S. Aquirre, A method to dramatically improve subcarrier tracking, *IEEE Trans. Commun.*, **COM-36**: 238–243, 1988.
14. A. Mileant and S. M. Hinedi, *Overview of Arraying Techniques in the Deep Space Network*, TDA Progress Rep. 42-104: October–December 1990, Jet Propulsion Laboratory, Pasadena, 1991.
15. H. Tsou et al., A functional description of the buffered telemetry demodulator for the Galileo mission to Jupiter, *1994 IEEE Int. Conf. Commun. Conf. Rec.*, **2**: 1994, pp. 923–928.
16. H. Tsou et al., The recovery of buffered telemetry data for future low cost space missions, *1995 IEEE Int. Conf. Commun. Conf. Rec.*, **2**: 1995, pp. 919–923.
17. K. Cheung et al., Enhancing the Galileo data return using advanced source and channel coding, *NASA Technol. 2004 Conf.*, Washington, D.C., September 1994.
18. W. Cham, Development of integer cosine transform by the principle of dyadic symmetry, *IEE Proc.*, part I **136**: 276, 282, 1989.
19. K. Cheung, F. Pollara, and M. Shahshahani, *Integer cosine transform for image compression*, TDA Progress Rep. 42-105: January–March 1991, Jet Propulsion Laboratory, Pasadena, 1991.
20. K. Cheung and K. Tong, Proposed data compression schemes for the Galileo S-band contingency mission, *Proc. NASA Space Earth Sci. Data Compression Workshop*, Snowbird, UT, 1993.
21. E. Paaske, Improved decoding for a concatenated coding system recommended by CCSDS, *IEEE Trans. Commun.*, **COM-38**: 1138–1144, 1990.
22. O. Collins and M. Hizlan, Determinate-state convolutional codes, TDA Progress Rep. 42-107: July–September 1991, Jet Propulsion Laboratory, Pasadena, 1991.
23. J. Statman et al., Decoder synchronization for deep space missions, TDA Progress Rep. 42-116: October–December 1993, Jet Propulsion Laboratory, Pasadena, 1994.
24. S. Dolinar and M. Belongie, Enhanced decoding for the Galileo low-gain antenna mission, *Proc. 1994 IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, 1994.
25. E. Berlekamp, Bit-serial Reed–Solomon encoder, *IEEE Trans. Info. Theory*, **28**, 1982.
26. R. McEliece, The decoding of Reed–Solomon codes, TDA Progress Rep. 42-95: July–September 1988, Jet Propulsion Laboratory, Pasadena, 1988.
27. T. Chauvin and K. Cheung, A parallel Viterbi decoder for shared memory architecture, *SIAM Conf. Parallel Signal/Image Process. Multiprocess Syst.*, Seattle, WA, August 1993.
28. H. Tsou et al., Description of communication system options for single-aperture multiple-link (SAML) mission support, TDA

Progress Rep. 42-127: July–September 1996, Jet Propulsion Laboratory, Pasadena, 1996.

HAIPING TSOU
KAR-MING CHEUNG
California Institute of Technology

DATA RECORDING

Data are items of information. We walk through the information superhighway. Data are all around us—weather data, sports data, stock data, sonar data from underwater objects, seismic data from an earthquake, satellite data from the Mars; one-dimensional speech data, two-dimensional image data, three-dimensional video data; analog data, and digital data; some data are natural, some are synthetic.

Data are recorded for later retrieval and use. Archeological findings are recorded to trace back the past. Current scientific findings are recorded to strive for the future. Musical data are recorded to listen at a later time on demand. So, what is data recording? Data recording refers to a system that can store data and restore it as faithfully as possible. A data recording device, therefore, involves both the storage and retrieval mechanisms. A data recording system may be viewed as shown in Fig. 1.

The source generates the raw data. Data acquisition is the process of measuring and gathering data from the source in some usable format. Acquired data are usually compressed to reduce the burden on the storage media. Each of these processes will be discussed next.

DATA ACQUISITION

As mentioned earlier, data acquisition collects data in a usable format from the data source. This usually comprises a sensor, calibrator, and/or data converter. Figure 2 shows typical components of a data acquisition system. For example, for the recording of temperature of a system, a thermocouple may be used to sense the temperature that will yield a sequence of electric potential data values. This may be followed by proper calibration and finally may need to be converted to a discrete sequence of data using an analog-to-digital converter.

Data Telemetry

When the data source is at a distance of actual storage media, then data need to be transmitted through this distance. This technique of distant measurement of data is called data telemetry. The problem here is to get the data from one point to the other without having to go and physically retrieve it. With the progress of satellite communication and Internet distribution, this is becoming a wider application area day by day. The Global Positioning System (GPS) is the most active application

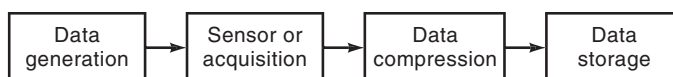


Figure 1. A data recording system flow.

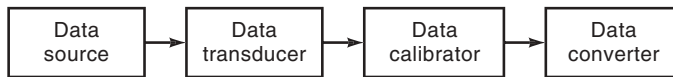


Figure 2. A data acquisition system flow.

area of data telemetry. The Automated Vehicle Locator (AVL), interactive monitoring and control of remote equipment and processes, stolen vehicle recovery systems, and interactive vehicle messaging systems are some of the application areas of GPS.

DATA COMPRESSION

Consider a sequence of National Television Standards Committee (NTSC) standard satellite video data. Each frame when represented in 24-bit true color takes up approximately 1 MB of storage space. With 30 frames/s, 1 min recording of the satellite data thus requires 1800 MB (1.8 GB), which is sufficient to fill the entire hard disk space of a modern day personal computer. Is there a way to reduce the size of these image data? That is what data compression is all about. Based on the requirements of reconstruction, there are two types of compression: lossy and lossless compression. Lossless compression is generally used for text, computer-generated data, and some types of image and video information. A particular application is the compression of bank account and credit information. Lossy compression, on the other hand, can tolerate some loss of information. For example, a reconstructed video image sequence from a compressed sequence may be well within the tolerance of viewer.

Lossless Compression

Lossless compression, in general, employs some coding techniques to reduce the size of data.

Coding Techniques

Huffman Coding. Consider a coding system in which the symbols that occur more frequently will have shorter-length codeword than the symbols, which are less likely to occur. The result is a variable-length coding having a smaller average code size. In the context of a 256-level gray image, each pixel is coded as an 8-bit code. Now assume that analyzing the probabilities of the gray levels, it is possible to have an average code size of 5 bits per pixel. That gives a compression ratio of 1.6.

Arithmetic Coding. This is used for shorter-codeword sources and instances in which the codewords have highly skewed probabilities.

Dictionary Coding. A popular realization of this is the LZW approach, which is a variation of the seminal work by Ziv and Lempel in 1977. A dictionary codeword is created from the patterns of the source. The codewords consist of the index, length, and symbol of a frequently used pattern. The LZW technique is used in industry standard applications as in the UNIX `compress` command, graphical image format (GIF) image coding, and the V.42 bis modem data compression.

Differential Encoding. Instead of directly encoding a source, the next data element is predicted from its neighboring elements and then the difference between the actual and

predicted model is encoded. This reduces the length of code size by reducing the dynamic range of possible values. The basic differential encoding is called differential pulse coded modulation (DPCM). DPCM attempts to minimize the variance of the predicted model and the actual values. The popular JPEG (joint photographic expert group) technique uses differential encoding for compressing image data.

Lossy Compression

Lossy compression reduces the data size with an acceptable loss of resolution of data. The technique of quantization is used in general for lossy compression, which is then followed by coding. Quantization maps a larger dynamic range of the original source into a smaller range of quantized values. The compression ratio now is a function of the levels of quantization desired, which obviously is much higher than that of lossless compression. The set of quantizer output values is called the codebook.

Coding Techniques

Differential Encoding. As in the case of lossless compression, this technique may also be used for lossy compression. Here the differentials of input values go through the process of quantization. Different variants of differential encoding exist in literature.

Subband Coding. Here, a data sequence is first decomposed into different bands or resolutions. A bit allocation technique is then used for an optimal combination of these bands. This may be used in both lossless and lossy compression. Wavelet decomposition is an example of subband coding. Different types of filters used for subband decomposition result in different types of information.

Transform Coding. The data sequence is first converted to another sequence using some transforms. Use of quantization and other compression techniques is then used in the transform domain instead of the original data domain. One of the goals of transformation is to have a smaller number of data samples that are representative of the original data sequence. This process is called decorrelating the input sequence, which means that the sample-to-sample correlation of the original data sequence is zero. The transform samples that do not contain much information are discarded. The reverse transformation is then applied on this compressed sequence for reconstruction. There are different types of transforms available, all of which are, to some extent, data dependent. The Karhunen–Loeve transform is an optimal transform in the sense that it minimizes the geometric mean of the variances of transform coefficients, thus resulting in an optimal decorrelation providing the largest coding gain of any transform coding method. The discrete cosine transform (DCT), discrete sine transform, and discrete Walsh–Hadamard transform are some of the other useful transforms.

Transform coding operate in three steps: First the transformation is applied. The second step involves quantization of the transformed sequence into the desired number of bits. Finally, some types of coding techniques such as fixed-size code, Huffman, or arithmetic coding are used. The JPEG standard is one of the most widely used techniques for lossy compression of images. It uses DCT as the transform and goes through the following steps.

1. The input image is first level-shifted. This is obtained by subtracting 2^{M-1} from the image having pixel values in the range $\{0, 2^M - 1\}$. Here, M is the number of bits used to represent each input pixel.
2. Block transformation with a block size of 8×8 using DCT is then applied on the level-shifted image.
3. Quantization on the transform coefficients is then done using a variable step size. The idea is to use a small step size for low-frequency components and a large step size for high-frequency components.
4. Huffman coding method is then applied on the quantized coefficients. Note that a large number of 0 values are obtained after the quantization process and these values do not need to be coded.

As mentioned earlier, the compression ratio really depends on the fidelity of the reconstructed image. In general, a ratio of 25:1 results in an acceptable reconstruction for typical image applications.

The MPEG (motion pictures expert group) standard, which is used for the compression of motion pictures, also uses the DCT transform. Here the next frame is predicted from the current frame. DCT is taken for the difference of a frame and its predicted frame.

DATA STORAGE

The ever-increasing demand for data storage coupled with multimedia requirements for text, image, video, and audio storage requirements is growing at an exponential rate. In practice, we encounter three main types of storage: primary or on-line, secondary or near-line, and archival or off-line memory. The range of on-line, near-line, and off-line data storage systems may consist of a variety of technologies such as magnetic tape drives, magnetic disk drives, optical tape, electron trapping optical memory (ETOM), holographic optical storage, and optical disk drives such as the compact disk read-only memory (CD-ROM), CD recordable (CD-R), digital versatile disk (DVD), and write-one read-many (WORM).

Storage Modes

Data storage may be analog or digital. The audio tape is an example of analog recording, whereas a CD is an example of digital recording. An analog signal is a continuously varying smooth signal having no discontinuity in it. A digital signal, on the other hand, consists of a sequence of discontinuous pulses. The amplitude of the digital signal is fixed, whereas the amplitude of an analog signal is a changing parameter that imposes a requirement on the dynamic range of an analog storage device. Figure 3(a) shows an analog signal along with discretization in 15 samples. Figure 3(b) shows how each discretized value is represented by using 4 bits.

Storage Media

Depending upon the media and technology used, storage devices may be classified as semiconductor storage, magnetic storage, optical storage, and magneto-optical storage.

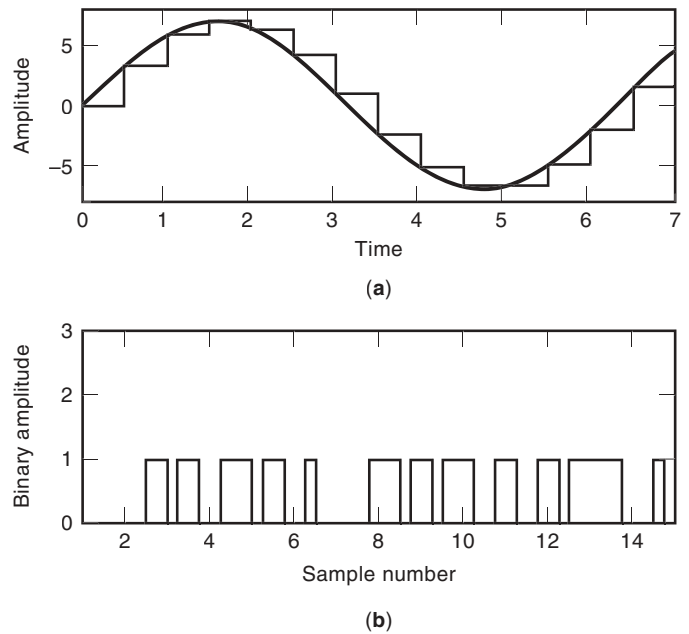


Figure 3. (a) Analog signal and the digitization process. (b) Coding as a digital signal.

Semiconductor Storage. In a computer system, semiconductor storage is used as a faster access media, which is called primary storage. This is primarily divided into read-write memory (RWM) and read-only memory (ROM). RWM historically was called random access memory (RAM), although ROM is also random access. Random access, in contrast with sequential access of a tape, means any piece of data item can be accessed randomly without accessing other data items. RAM is volatile in the sense that its content is destroyed when the power is turned off. Depending on the technology used, RAM can be static or dynamic. Static RAM (SRAM) does not need any refreshing and is usually faster than dynamic RAM (DRAM). DRAM, on the other hand, can be made denser and is thus less costly. One variant of ROM, called programmable ROM (PROM), can be programmed only once, sometimes also referred to as one-time programmable (OTP). Erasable PROM (EPROM) can be programmed and reprogrammed many times.

The basic building block of a semiconductor storage system is called a flip-flop. Figure 4 shows the storage of 1 bit of information using a complementary metal oxide semiconductor (CMOS) SRAM cell. Note that many variations of this configuration are available, resulting in different storage

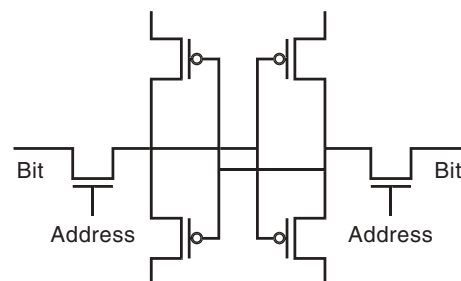


Figure 4. CMOS realization of 1-bit storage of static RAM.

capability. With the tremendous success of very-large-scale integration (VLSI) technology, semiconductor memory has become the storage device with the highest density and fastest speed. Following are some of the off-the-shelf semiconductor storage devices.

PC Cards. PC cards can be either storage or input/output (I/O) cards, which is defined by the internationally accepted Personal Computer Memory Card International Association (PCMCIA) standard. PC cards use little power and are compact, highly reliable, and lightweight, making them ideal for battery-powered notebook and palmtop computers, handheld personal digital assistants (PDAs), and personal communicator devices. Their low power, small size, and portability facilitate the use of memory cards in a wide variety of new applications, such as electronic cameras, voice recorders, portable phones, and in the area of data telemetry in general.

ROM and OTP Cards. Because the information stored on a ROM card is inseparable from the hardware, a ROM card usually is device-dependent and will not work in a computer system other than that for which it is specifically manufactured. By contrast, OTP cards are built *blank*. After manufacture, these chips can be written to with a special device. While they have the advantage of providing a form of nonvolatile storage, the biggest disadvantage is that once information is stored on the card, it cannot be altered in any way. In addition, the custom manufacturing involved in producing ROM chips makes ROM cards expensive unless produced in large quantities.

SRAM Cards. SRAM cards have an advantage over ROM and OTP cards: users can write, modify, or delete the information stored on them. However, SRAM chips are volatile. Affordable SRAM cards are available in capacities of up to 4 MB. Beyond that, the cost of SRAM is prohibitive. Therefore, SRAM cards are used primarily for low volumes of user storage.

Flash Cards. One of the latest innovations in memory cards is the use of flash memory technology. Flash cards combine the best features of both ROM/OTP and SRAM cards. Like ROM and OTP chips, flash components are nonvolatile—they retain data even without a power source. And, like SRAM, users can write, modify, or delete the information stored on them. However, use of flash memory currently is prohibitive due to cost and device capacity.

Magnetic Storage. Magnetic storage uses electromagnetism. When electric current flows through a conductor, it generates a magnetic field around that conductor. This magnetic field can influence the magnetic material in the field. When the direction of current reverses, the polarity of the magnetic field is also reversed. So, particles of the magnetic substance can be polarized magnetically in one of the two directions with an electromagnet. Thus, magnetic polarization can be used to distinguish 0 and 1. This two-way operation of electromagnetism makes it possible to record data on a disk or tape and read the data later. Magnetic disks and tapes are the two variants of magnetic storage devices.

Magnetic Disks. Magnetic disks, which were preceded by magnetic drums, are of two major types: hard disks and floppy disks. Hard disks cannot be bent or flexed—hence the term hard disk. Also, the platters of the hard disk are not removable, and for that reason it is also called a fixed disk.

Construction and Geometry. Disks consist of circular platters made of glass, plastic, or metal and coated with a magnetic

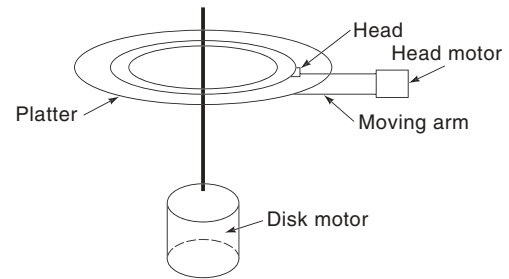


Figure 5. Magnetic disk drive components.

substance. A driver motor rotates the disk platter about its central axis. The disk drive has also head motors, which cause head rotation. Heads move radially in and out across the disk surface. Heads are connected to head motors through a moving arm. Usually there is one head per surface. Figure 5 shows the components of a typical magnetic disk drive.

Figure 6 shows the logical geometry of platters of a disk. Data are stored in concentric tracks on the surfaces of each platter. The number of tracks is an indicator of the storage capacity of the disk. Track density, which is the number of tracks per unit of radial length, is expressed in tracks per millimeter or tracks per inch. A typical disk drive can have more than 2000 tracks per inch (TPI) on its recording surface. A pair of tracks on each side of the disk with the same track number is called a cylinder. In the case of disk packs several tracks may constitute a cylinder. Each track is divided into sectors, which can store a data block. The process of organizing the disk surface into tracks and sectors is called formatting. In almost all systems, including PCs and Macintoshes, sectors typically contain 512 bytes of user data plus addressing information used by the drive electronics.

In earlier hard drive designs, the number of sectors per track was fixed, and, because the outer tracks on a platter have a

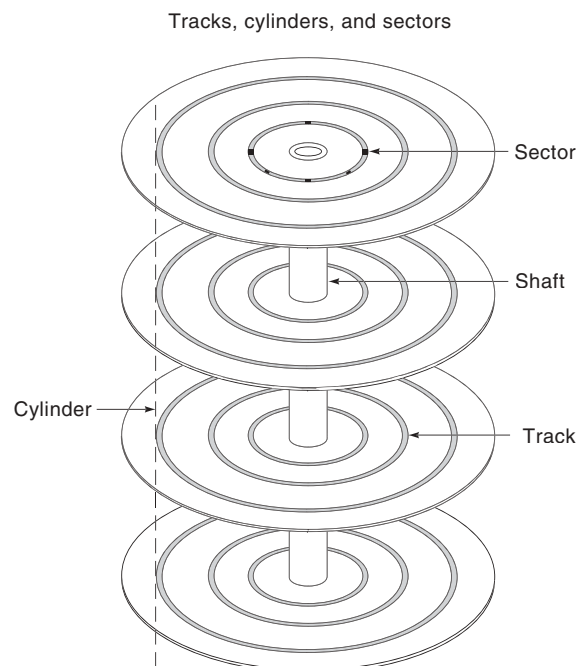


Figure 6. Logical components of a disk system.

larger circumference than the inner tracks, space on the outer tracks was wasted. The number of sectors that would fit on the innermost track constrained the number of sectors per track for the entire platter. However, many of today's advanced drives use a formatting technique called multiple zone recording that allows the number of sectors per track to be adjusted so more sectors are stored on the larger, outer tracks. By dividing the outer tracks into more sectors, data can be packed uniformly throughout the surface of a platter, disk surface is used more efficiently, and higher capacities can be achieved with fewer platters.

Read/write heads are the single most important component of a hard disk drive, and their characteristics have a great impact on drive design and performance. A head is a piece of magnetic material, formed almost in the shape of a C with a small opening or gap. A coil of wire is wound around this core to construct an electromagnet. In writing to the disk, current flowing through the coil creates a magnetic field across the gap that magnetizes the disk-coating layer under the head. In reading from the disk, the read/write head senses an electronic current pulse through the coil when the gap passes over a flux reversal on the disk.

Several different types of read/write heads exist. Among the earliest were monolithic ferrite heads or those made of a single block of ferrite, a magnetic ceramic material. An improvement on all-ferrite heads were composite heads, consisting primarily of nonmagnetic material with a small ferrite structure added. Next came metal-in-gap or (MIG) heads with very thin metal layers added inside the gap to improve magnetic performance. Currently, many drives use thin-film heads, the name of which reflects the fact that their structural elements are deposited on a substrate similar to way that microchips are manufactured. Thin-film technology allows head vendors to achieve much smaller physical dimensions and to better control the fabrication process, both of which result in higher-performance products.

Storing and Retrieving Data. The mechanism of storing and retrieving data is almost same for both the hard disk and floppy disk. The only difference is that the floppy disk has one disk platter and for this, the number of surfaces is two: the top and bottom surface of the single platter. Most hard disks contain several disk platters all mounted on the same axis.

When data are retrieved from a hard disk drive, a command is issued to open an existing file. The application program that is running prompts the user to enter the name of the file to open. It then passes the file name to the operating system, which determines where the file is located on the disk drive—the head number, cylinder, and sector identification. The operating system transfers this information to the disk controller, which drives an actuator motor connected to the actuator arm to position the heads over the correct track. As the disk rotates, the appropriate head reads the address of each sector on the track. When the desired sector appears under the read/write head, the entire contents of the sector containing the necessary data are read into the computer's main memory.

Storing data on a hard drive is a similar process to retrieving data, only reversed. The host computer operating system is responsible for remembering the addresses for each file on the disk and which sectors are available for new data. The operating system instructs the controller where to begin writing information to the disk. The controller moves the read/write

heads to the appropriate track and writing begins. When the first track is full, the heads write to the same track on successive platter surfaces. If still more track capacity is required to store all the data, the head moves to the next available track with sufficient contiguous space and writes the data there. Now let us see the encoding process involved in the read/write.

Encoding. Traditionally, drives have used analog peak-detection read channels. During a write operation, an analog peak-detection read channel converts binary, digital data into an analog signal, which the drive's read/write head uses to cause magnetic flux changes on the platter surface [Fig. 7(a)]. During a read operation, the read/write head detects the magnetic flux changes from the data and generates an analog readback signal, in the form of a wave, that it transmits to the read channel. The read channel analyzes the incoming signal to determine the high/positive peaks and low/negative peaks. Finally, it decodes each of these peaks. The limitation of analog peak-detection read channels is that, as data densities increase, the analog signal peaks start to overlap, which causes data degradation. To counter this effect, hard drives employ a data encoding scheme during write operations that separate the analog signal peaks. One of these techniques is called the partial response maximum likelihood (PRML).

With the use of sophisticated digital coding and filtering techniques, PRML read channels sample the analog signal wave at a number of points, as opposed to just at the high and low peaks, as shown in Fig. 7(b). By taking these samples, PRML read channels can determine the "shape" of the readback signal and thus can interpret the high and low peaks that represent data bits very accurately. Although using the

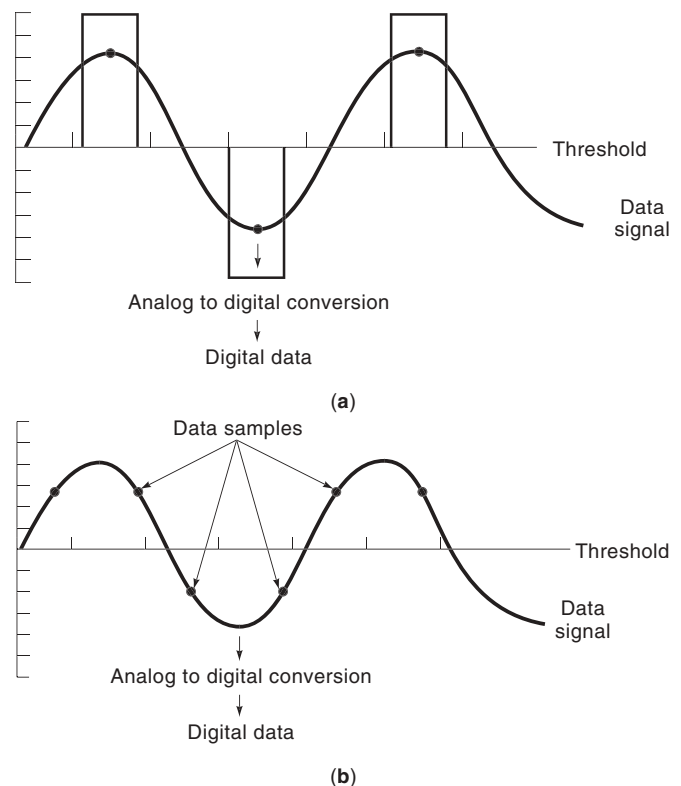


Figure 7. (a) Analog peak detection encoding. (b) PRML encoding.

Table 1. Comparison of Codes in Magnetic Recording

Code	Rate	Capacity	d, k	Packing Density	Detection Window	Dc ^a Free
NRZ-L	1	1	0	1	T	No
NRZ-M	1	1	0	1	T	No
E-NRZ	7/8	1	0, 13	0.875	T	No
FM	1/2	0.6942	0, 1	0.5	T/2	Yes
MFM	1/2	0.5515	1, 3	1	T/2	No
Mil'r Sq.	1/2	0.6509	1, 5	1	T/2	Yes
ZM	1/2	0.5515	1, 3	1	T/2	Yes
RLL 2,7	1/2	0.5174	2, 7	1.5	T/2	No
GCR 4/5	4/5	0.8792	0, 2	0.8	0.8T	No

^a Dc Free means average value is zero.

same read/write heads and media, the use of PRML technology provides a much more space-efficient error detection scheme, which yields a 25% improvement in data bit density while achieving the same, low bit error rate as analog peak detection.

Recording Codes. Numerous codes for recording are available. Table 1 shows a comparison of them. The code rate is the ratio of the number of information symbols (m) and the number of code symbols (n), $m/n \leq 1$.

The detection or sample window relates to the interval during which the recorded bit may be sampled on playback to determine its value. It is normally measured in terms of bit-cell duration T . The capacity $C = \log x$, where x is the largest root of the equation

$$F(x) = x^{k+1} - x^k - d - x + 1 = 0 \quad (1)$$

where d is the minimum number of 0s between adjacent 1s and k is the maximum number of 0s between adjacent 1s.

The density ratio (DR) is a measure of the packing density of the code: $DR = m/n(d + 1)$.

Disk Performance Specifications

Data Access Time. Access time is a measure of how long it takes to position a read/write head over a particular track and find the sectors of interest within the track for reading or writing. Thus, data access time is a combination of seek time, head switch time, and rotational latency and is measured in milliseconds.

Seek Time. The amount of time it takes the actuator arm to move the read/write head between tracks is called seek time. For a platter having N tracks on each side, moving from the current track to the next requested bit of data might entail moving just one track or up to $N - 1$ tracks. Seek time is measured in milliseconds (ms). Seek times between adjacent tracks can be as short as 2 ms, while a full-stroke seek (movement between the outer and inner track) consumes about 20 ms. The average seek time is defined as the time it takes to position the drive's read/write heads for a randomly located request and usually ranges from 8 to 14 ms.

Head Switch Time. The actuator arm moves all of the read/write heads over the platters simultaneously. However, only one of the read/write heads can be reading or writing data at a time. Head switch time measures the average time the drive takes to switch between two of the read/write heads when reading or writing data.

Rotational Latency. Once the read/write head is positioned over the proper track, it must wait for the drive to spin the platter to the correct sector. This waiting period called rotational latency, is measured in milliseconds, and is dependent on the drive's revolutions per minute (rpm). On average, the disk needs to spin half-way around before the next sector to be read or written to. Increasing the disk's rotational speed decreases this latency. However, the faster the disk rotation, the hotter the drive runs and the more wear on the drive's moving parts. Still, manufacturers are overcoming these problems, and the once universal 3600 rpm (with an average latency of 8.3 ms) is being supplanted by faster speeds like 7200 rpm (with a latency of 4.2 ms).

Data Transfer Rate. The data transfer rate depends heavily on two measures: the disk transfer rate, or how fast data are passed from the disk to the hard drive's controller, and the host transfer rate, or how fast the controller passes data to the computer's central processing unit (CPU). The data transfer rate is measured in megabytes per second (MB/s). To speed up the host transfer rate and minimize mechanical delays (from seeking and rotational latency), manufacturers have added cache buffers to the hard drive. A cache buffer built into the drive electronics provides a temporary storage space that eliminates or reduces any bottleneck in the flow of data between the disk and CPU. A cache buffer coupled with an effective caching algorithm can improve the effective data transfer between a drive and the CPU by a significant factor—often 50% to 100%.

Data Throughput Rate. Data throughput rate is a measurement reflecting both data access time and data transfer rate. It represents the total amount of data that the CPU can access in a unit of time. As such, it is a fairly comprehensive measure that factors in most of the major drive performance measurements.

Storage Capacity. Capacities are stated in two different ways: unformatted and formatted. Formatted capacity represents the real capacity, the amount of data that a user can store, while unformatted capacity represents a higher number usually used in advertisements. The unformatted capacity is the maximum number of bytes that can be recorded assuming that each track is continuously and completely recorded with 0s and 1s. The unformatted capacity of a disk is, therefore, the product of the maximum number of bytes per track times the number of tracks times the number of recordable sides. For modern modes of recording such as modified frequency modulation (MFM), one bit is represented by one flux transition. The

number of flux transitions for a track equivalent to one complete revolution, that is, 2π radian, is therefore calculated as follows:

$$\text{Flux transitions/track} = \text{flux transitions/rad} \times 2\pi = \text{bits/track}$$

The unformatted capacity C_u , of the 2-MB 3.5-in. diskette (15,916 flux transitions/rad, 80 tracks, 2 sides) is $C_u = (2\pi \times 15,916 \times 80 \times 2)/8 = 2,000,063 \text{ bytes} \approx 2 \text{ MB}$.

For a certain standardized format, the real formatted capacity is the product of the number of data bytes per sector times the number of sectors per track times the number of tracks per side times the number of sides. For the same diskette using the standardized MS-DOS format, the formatted capacity, C_r is

$$C_r = 512 \times 18 \times 80 \times 2 = 1,474,560 \text{ bytes} \approx 1.47 \text{ MB}$$

For this type of diskette the formatted capacity is about 73.7% of the unformatted capacity.

The performance of magnetic storage devices has been enhanced greatly by the emerging technologies. Two of these are mentioned below. One relates to capacity and the other to speed.

- Material improvements such as thin-film and magnetoresistive (MR) heads, have been made. Unlike current head technologies—all of which are basically tiny inductive electromagnets—MR technology uses a different material whose electrical resistance changes in the presence of a magnetic field. A small stripe of MR material is deposited on the head structure, and, as it passes over the magnetic patterns on the disk, it senses the strength of the magnetic field and creates electrical pulses corresponding to the flux reversals. This mechanism cannot be used for writing, so a conventional thin-film inductive write element is deposited alongside the MR stripe. MR head technology began to appear in drive designs in 1994, with wider acceptance and inclusion in more designs by 1995. IBM recently announced their MR-head-based disk that holds 3×10^9 bits of data per square inch.
- Seek command reordering of the disk drive is also an improvement. In the absence of command reordering techniques, the disk drive would execute the commands in the queue in the order it received them. The result is that the read/write head randomly sweeps over the surface of the platter executing each command in the queue in sequence. Command reordering minimizes this random movement of the read/write head. One technique for command reordering is called *shortest seek time-first* ordering that selects the command with the shortest seek time for the next execution. The problem with the method is that it gives higher priority to the tracks located near the middle compared to the commands requiring data from the outer and inner tracks because their seek times are usually longer. *Elevator seeking* overcomes this problem by also ordering commands such that the read/write heads move back and forth across the entire platter surface. *Optimized reordering command algorithm (ORCA)* takes command reordering one step further to calculate the optimum sequence of seeks so as to minimize the combined total of seek time as well as rotational latency. Thus it ex-

amines both the track location of the seek on the disk and the sector location of the seek on the track. Therefore, it orders the seek command tasks in a more efficient sequence, increasing the data throughput of the drive by an average of 20%.

Magnetic Tape. Commonly used on early mainframe computers, one of the first computer storage technologies was the magnetic tape drive, or simply tape drive. Magnetic tape is a sequential data storage device. To read data, a tape drive winds through a spool of tape to the exact location of the desired information. To write data, the tape drive encodes it sequentially on the tape. Because tape drives cannot randomly access or write data like disk drives and are thus much slower, they have been replaced as the primary storage device in most computer applications. However, with its high storage capabilities and low cost-to-megabyte ratio, magnetic tape is still very much in use as a storage medium for archiving large amounts of data. Recent advances in tape technology, such as faster tape drives, DLT, and digital audio tape (DAT) cartridges, have also made tape the preferred technology for backing up network servers and other critical data.

Optical Storage. Currently, optical memory devices are widely used as secondary memory in computer systems. The most widely known optical storage device is the CD-ROM, which has significantly enhanced the replication, distribution, and utilization of software, games, audio, and video intended for use in computers or in entertainment instruments. For instance, a CD-ROM containing 680 MB of data can be mass-duplicated by injection molding in a few seconds costing less than 10 cents per piece. The main attractive features of an optical data storage system are removability, portability, and reliability.

Optical Disk. An optical disk is usually constructed from a plastic or glass substrate coated by one or more thin-film layers and contains a number of tracks along which information is stored. The manufacturer may record the information on the substrate surface or it may be recorded by the user on one or more of the thin-film layers along the tracks. The storage layer is generally sandwiched between two dielectric layers and the stack is covered with a metal layer for better reflectivity. The functions of the dielectric and reflective layers include reduction of thermal cross-talk during writing, optimization of absorptivity and/or reflectivity, and protection of the storage layer.

Figure 8 shows the basic building blocks of an optical disk system. The light beam emanating from the laser is collimated and directed toward a high-numerical-aperture objective lens via a beam splitter. The objective lens focuses the light beam on the surface of the disk where information is written to or read from a given track. The light reflected from the disk surface is collected by the objective lens and directed toward one or more detectors through the beam splitter. The detector extracts the data readout signal as well as the focusing and error-tracking signals.

The main features of an optical disk are shown in the right-hand side of Fig. 8. The access window is used by the read/write head to gain access to the disk. The hub provides the mechanical interface with the drive for mounting and centering the disk on the spindle. The track shown in Fig. 8 may be concentric rings of width W_t , and adjacent tracks are separated from each other

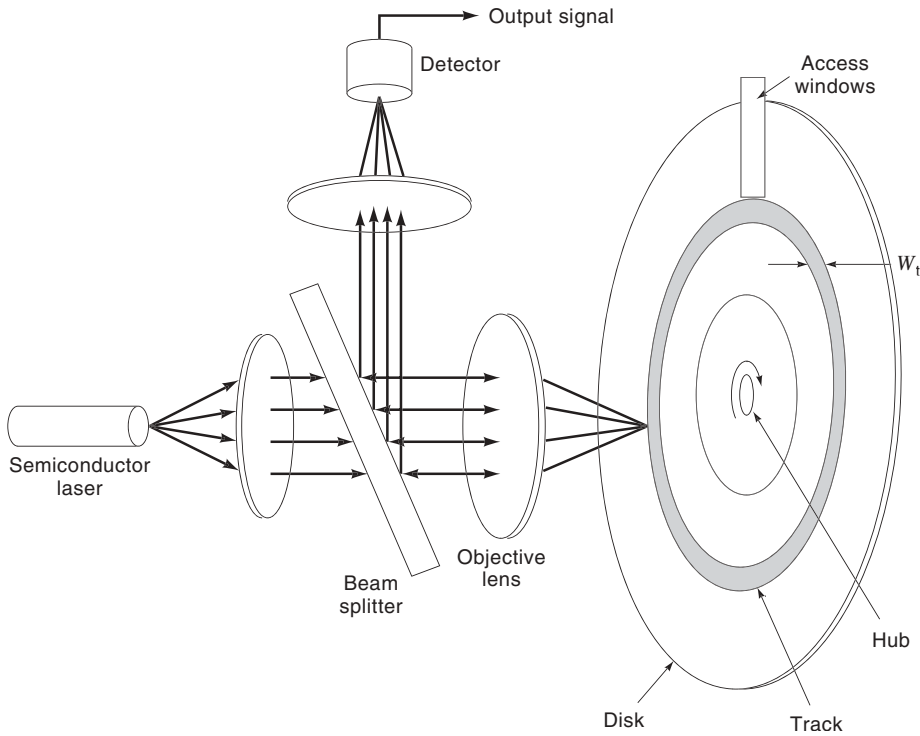


Figure 8. Basic building blocks of an optical disk system: W_t is width of track.

by a guard band of width W_b . If the binary digits (0 and 1) are stored as transparent and opaque marks of length L_m and the entire disk surface of radius r is covered with such marks, then the capacity of the optical disk may be expressed as

$$C = \frac{\pi r^2}{(W_t + W_b)L_m} \quad (2)$$

where C is expressed in bits per surface. For a typical 5.25-in. optical disk, $r = 67$ mm, $L_m = 0.5 \mu\text{m}$ (determined by the read/write laser wavelength), and $W_t + W_b = 1 \mu\text{m}$. Then the capacity of this disk, using Eq. (2), becomes approximately equal to 3.5 GB per surface, which is pretty close to reality.

In optical disk and tape storage systems, semiconductor laser diodes, having the shortest possible wavelength that can provide enough power for read, write, and erase operations for several thousand hours, are generally used. In general, the size of the focused spot varies directly with the wavelength. Therefore, shorter-wavelength lasers can be focused to smaller spots at the diffraction limit. The diameter d of the focused spot may be expressed as

$$d \approx \frac{\lambda}{\text{NA}} = \frac{\lambda}{\sin \theta} \quad (3)$$

where NA refers to the numerical aperture of the objective lens, λ is the wavelength of light, and θ is the half angle subtended by the focused cone of light at its peak. Equation (3) implies that a higher-numerical-aperture objective lens should be used for smaller spots. But the depth of focus δ of the objective lens may be expressed as

$$\delta \approx \frac{\lambda}{(\text{NA})^2}, \quad (4)$$

insinuating that the depth of focus will be smaller for higher-

numerical-aperture objective lens. The range of NA usually varies from 0.4 to 0.6.

Optical Tape. Figure 9 shows the main components of an optical tape, which consists of a substrate onto which an active layer and a dielectric film are sputtered flanked by a back and a top coating for better reliability. The active layer usually consists of thin metallic film, dye polymer, or phase change material. Information is hole-punched (called marks) into the active layer by a focused semiconductor laser beam. For data readout, the tape is scanned by a focused laser beam and the marks cause the amplitude of the reflected light to vary. These variations are detected by a detector array to generate the appropriate data readout signal.

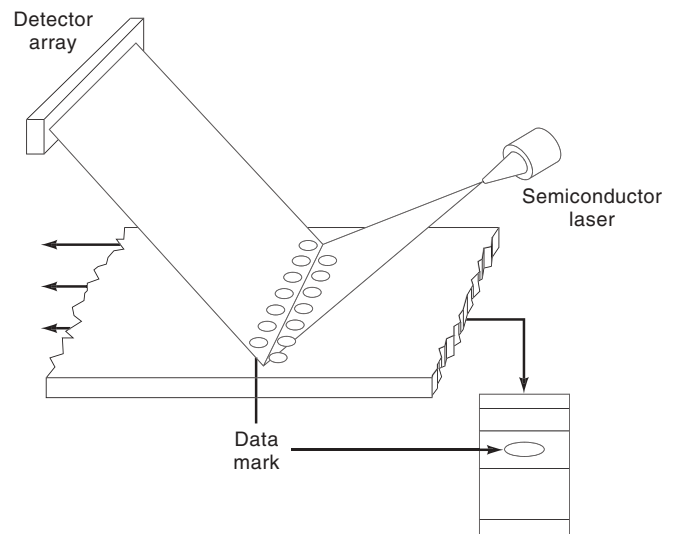


Figure 9. Basic building blocks of an optical tape system.

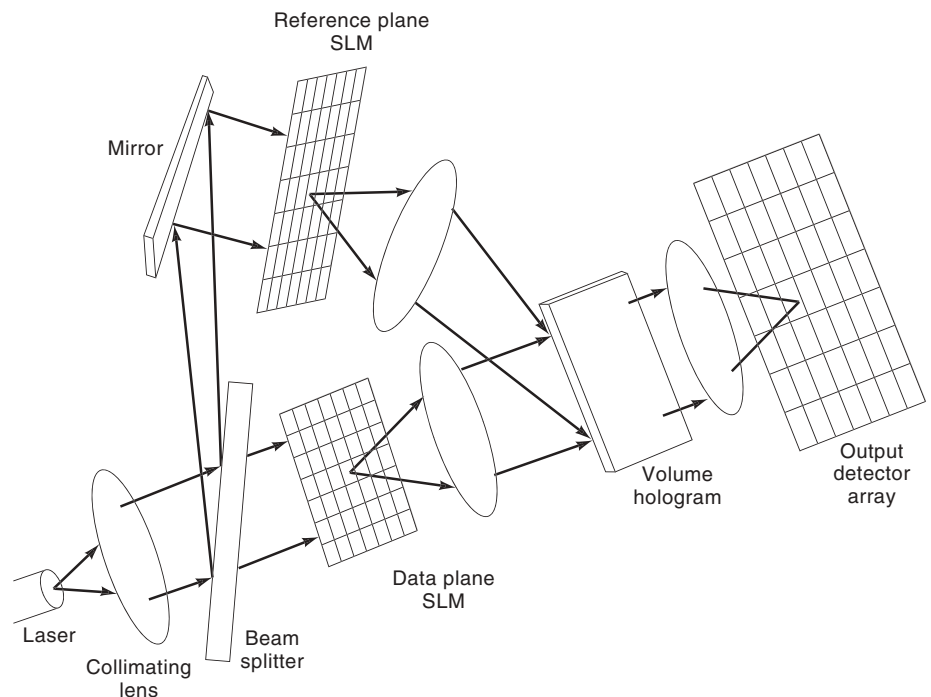


Figure 10. Basic building blocks of a volume holography storage system.

Optical tape is mainly used for high volume storage. It is so thin that staggering amounts of data can be stored in a very small volume of it. For instance, optical tape has at least 25 times the volumetric density of CDs. Although magnetic tape is economical, it is not suitable for long-term archiving because the magnetic medium is degraded by creep, track deformation, and print-through, which requires the transfer of stored information to a new magnetic tape after a given interval. Optical tape is more durable, does not suffer from print-through, and is more reliable than magnetic tape. The effective data rate for a magnetic tape is very low since the read/write process is inherently sequential. On the other hand, the effective data rate for an optical tape can be enhanced significantly by exploiting parallel read/write channels as depicted in Fig. 9.

Electron Trapping. Electron trapping optical memory (ETOM) exploits multilevel encoding through amplitude modulation, which allows it to store one of several intensity levels at a given location. For instance, if any of eight amplitude levels can be detected reliably, then 3 bits can be stored in a single location, thus tripling the storage capacity.

ETOM usually consists of an active layer of phosphorlike material coated on a substrate. To store information, the trapping material is exposed to light of one wavelength, which excites electrons into the conduction band where they are trapped. The number of electrons trapped is directly proportional to the intensity of light. To read data, the trapping material is exposed to light of another wavelength, which causes the conduction-band electrons to drop by emitting photons of a third wavelength. The number of photons emitted is proportional to the number of electrons trapped initially. Thus by counting the number of photon multiple values can be stored at a location.

The main limitations of an ETOM are that it requires a complex read/write head and three different lasers. Furthermore,

adding an extra bit for a given location requires a decibel level twice as large, thus causing the loss in signal-to-noise ratio.

Holographic Storage. Figure 10 shows the general features of a holographic data storage system. A hologram is created by recording the spatial interference pattern of two coherent beams, called the object beam and the reference beam, which are usually generated from the same laser source using a beam splitter and mirror combination. The object beam carries the information to be recorded while the reference beam is just a plane wave. The information or the data to be stored are usually a two-dimensional arrangement (called pages) of transparent and dark patterns representing binary 1s and 0s. This two-dimensional pattern is displayed in a spatial light modulator (SLM) in the data plane and modulates it onto the object beam. The object beam interacts with the reference beam in the holographic material to form a fringe pattern, which is recorded by the material, thus forming the hologram. When the hologram is illuminated with the original reference beam, it reconstructs the object beam, thus retrieving the information stored in the hologram.

As the hologram recording material becomes thicker, the reconstruction becomes particularly sensitive to the angle between the two interfering beams. Therefore, multiple holograms may be superimposed in the same volume by changing the angle of incidence of one of the beams, that is, the reference beam. This is accomplished by employing another SLM called the reference plane SLM. Each hologram can be accessed independently by using an appropriate set of reference beams.

The Bragg effect governs the diffraction from a hologram, which implies that if the reference beam is applied to the holographic medium at an angle other than that used during recording, it will not diffract the stored fringe pattern or the hologram. Therefore thousands of holograms can be recorded in a

small-volume hologram (on the order of 10^{11} bit/cm³) using a reference beam with a Bragg angle unique to each of the stored patterns. Furthermore, the page-oriented parallel access to data in a hologram offers low access times (less than 1 ms) and potentially high data rates (gigabit/s).

Magneto-Optical Storage. Magneto-optical (MO) disk systems combine the technology of traditional magnetic media, such as hard disk drives, with optical disk technology. MO technology allows users to pack hundreds of megabytes of data on a disk that looks similar to a traditional 3.5-in. floppy disk and typically comes in a 3.5- or 5.25-in. form factor. A MO disk is made out of materials that cause it to be highly resistant to magnetic fields, or coercive force, at room temperature.

A MO drive writes to the disk using a read/write head assisted by a laser. A laser heats up the disk surface to its Curie point. Then, the read/write head passes over the disk, polarizing those areas that have been heated by the laser. Because a laser can be focused on a much narrower field than a traditional, magnetic read/write head, data written on a MO disk with the assistance of a laser results in a very high data density not available with traditional hard disk drive technology. During a read operation, the MO drive uses the same laser to sense the data stored on the disk. As the laser scans the disk surface, the drive detects a reflection of light by the data bits oriented in one direction and no reflection from the data bits oriented in the opposite direction. Thus, a MO drive can distinguish between the 0 and 1 data bits stored on the disk. MO disks have many advantages:

- They provide very high data densities, achievable because of the use of a laser.
- Data can be changed at will—added, modified, or deleted.
- Data are resistant to magnetic fields. Unlike a traditional floppy or hard disk, a magnetic field alone cannot alter the data without the additional heat provided by a laser.
- Because of the use of the laser to assist in reading and writing data, the read/write head does not need to be as close to the surface of the disk as with a hard disk drive.

The disadvantage of MO technology is that, because of the high intensity of the magnetic field created with the combined use of the read/write head and laser, the drive cannot change the field polarity very quickly. Therefore, the drive must pass over the disk twice to write to it. Of course, attempts are going to compress the writing process into a single disk rotation. Thus, MO drives with one-pass writing capabilities should be available soon.

BIBLIOGRAPHY

1. C. D. Mee and E. D. Daniel, *Magnetic Recording*, New York: McGraw-Hill, 1988, Vols. I–III.
2. J. Watkinson, *An Introduction to Digital Audio*, 1st ed., Boston: Focal Press, 1994.
3. G. Held, *Data Compression: Techniques and Applications Hardware and Software Configurations*, 1st ed., New York: Wiley, 1984.

4. J. A. Storer, *Data Compression: Methods and Theory*, 1st ed., Rockville, MD: Computer Science, 1988.
5. K. Sayood, *Introduction to Data Compression*, 1st ed., San Mateo, CA: Morgan Kaufman, 1996.
6. M. H. Kryder, The special section on magnetic information storage technology, *Proc. IEEE*, **74**: 1475–1476, 1968.
7. F. Jorgensen, *The Complete Handbook of Magnetic Recording*, Blue Ridge Summit, PA: Tab Books, 1980.
8. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**: 379–423 (part 1), 623–656 (part 2), 1948.
9. N. D. Mckintosh and F. Jorgensen, An analysis of multilevel encoding, *IEEE Trans. Magn.*, **MAG-17**: 3329–3331, 1981.
10. G. V. Jacoby, Ternary 3 PM magnetic recording code and system, *IEEE Trans. Magn.*, **MAG-17**: 3326–3328, 1981.
11. M. Camras, *Magnetic Tape Recording*, New York: Van Nostrand Reinhold, 1985.
12. K. Ford (ed.), *High Density Digital Recording*, NASA, Scientific and Technical Information Branch, NASA reference publication, 1111, 1985.
13. A. S. Hoagland, *Digital Magnetic Recording*, Malabar, FL: Krieger, 1983.
14. H. Jouve, *Magnetic Bubbles*, Orlando, FL: Academic Press, 1986.
15. Y. Manolopoulos and A. F. Vakali, Seek distances in disk with two independent heads per surface, *Inf. Process. Lett.*, **37**: 37–42, 1991.
16. M. Mansuripur and G. Sincerbox, Principles and techniques of optical data storage, *Proc. IEEE*, **85**: 1780–1796, 1997.
17. G. Bouwhuis et al., *Principles of Optical Disk Systems*, Bristol, UK: Adam Hilger, 1985.
18. A. B. Marchant, *Optical Recording*, Reading, MA: Addison-Wesley, 1990.
19. T. W. McDaniel and R. H. Victoria (eds.), *Handbook of Magneto-Optical Data Recording*, Westwood, NJ: Noyes, 1997.
20. M. Guizani, ATM architectures using optical technology: An overview of switching, buffering, and multiplexing, *Int. J. Network Manage.*, **7** (4): 198–220, 1997.
21. M. A. Karim and A. A. S. Awwal, *Optical Computing—An Introduction*, New York: Wiley, 1992.
22. M. Mansuripur, *The Physical Principles of Magneto-Optical Recording*, Cambridge, UK: Cambridge Univ. Press, June 1995.
23. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991.
24. M. M. Murshed and M. Kaykobad, Seek distances in two-headed disk systems, *Inf. Proc. Lett.*, **57**: 205–209, 1996.

FARID AHMED
Penn State Erie, The Behrend
College

MOHAMMAD S. ALAM
Purdue University

M. KAYKOBAD
Bangladesh University of
Engineering and Technology

RECORDERS

Recorders are used in instrumentation and measurement to record the variations in time of some physical quantity for later analysis. This is in distinction to many audio and video recordings for which the purpose is to play back the recorded material. Recording is essential to the instrumentation, measurement,

and analysis processes, because many phenomena that must be measured occur much too rapidly to be analyzed in real time.

Modern high speed recorders share much in common with digital storage oscilloscopes. While these are often used as recorders, the emphasis in design is different. The primary design focus of the digital storage oscilloscope is convenience of capturing and displaying data to be instantly analyzed with the resolution of the eye. The primary focus in recorder design is to capture data for later analysis at potentially much higher resolution.

Important elements of any recording system are the transducers, to convert the physical quantities to be recorded into electrical signals, and the signal conditioners and instrumentation amplifiers to convert the signal to a form acceptable for recording. These elements are covered in other articles. Also covered in other articles are video recording, the medium that records the time history of two dimensional images, and data recording, the medium that records digital data. This article will be confined to a discussion of recording multiple channels of analog data that have already been converted to a voltage signal.

ANALOG-TO-DIGITAL CONVERSION IN MODERN RECORDERS

Modern recorders rely very heavily on analog to digital conversion. The signals to be recorded are converted to digital form by an analog-to-digital converter (ADC); the digital words are then recorded on media compatible with a personal computer (PC) such as magnetic disk or tape or magneto-optical or optical disk. This is generally the most economical approach, since digital data recording has become so inexpensive because of the large PC market. This approach is also very convenient, because a PC is usually the tool that will be used to analyze the recorded data. Often the recording will be done on a PC while the signal conditioning and analog to digital conversion will be on an add-in board to the PC.

IMPORTANT RECORDER CHARACTERISTICS

Sampling Rate and Bandwidth

The individual numbers stored by the recorder, which represent the amplitude of the recorded signal at a specific time, are called *samples*. The rate at which these samples are collected is called the *sampling rate* and is expressed in samples per second (Sa/s). Sampling rates of commercially available recorders vary from 1 Sa/s to 1 TSa/s (10^{12} Sa/s). Choosing an appropriate sampling rate for any particular application is important, because using too low a rate will result in loss of data while using too high a rate can add appreciably to the cost and complexity of the recorder.

The dynamic response of a recorder is usually modeled as a low pass filter followed by a perfect recorder. The frequency at which the response of this filter is 3 dB below its value at zero frequency is called the *bandwidth* of the recorder. Generally, bandwidth limitation is caused by nonidealness of all parts of the recorder between its input and (including) the ADC. Bandwidths of recorders commercially available range from 1 Hz to 10 GHz (10^{10} Hz).

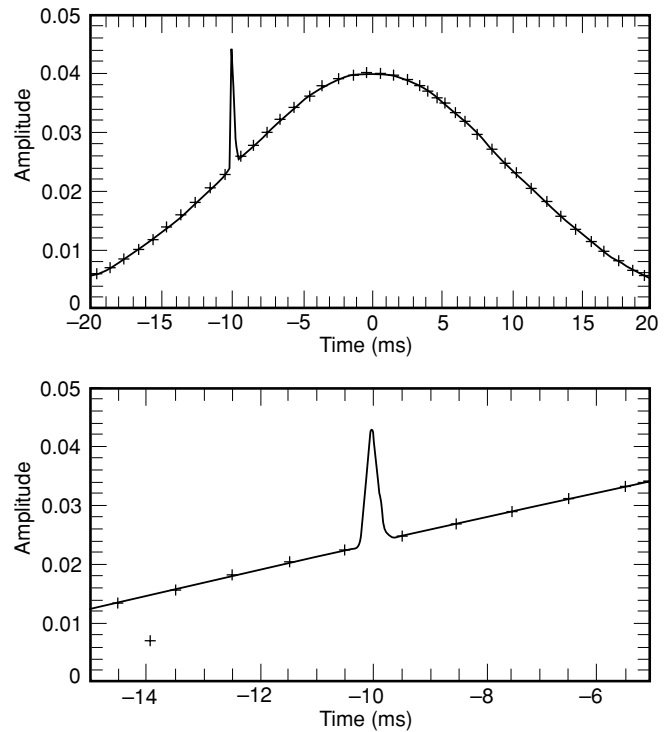


Figure 1. The signal of Eq. (1) sampled at 1 kSa/s (sample points indicated by + signs.) The short pulse at $t = -10$ ms is completely missed by the sampling. However, the Fourier transform of the signal (shown in Fig. 18) and the usual “rules-of-thumb” indicate that 1 kSa/s is an adequate sampling rate.

There is a relationship between the bandwidth required to record a signal and the sampling rate required. This comes from the sampling theorem (1,2) which states (roughly) that a signal can be reconstructed from samples taken at a rate of twice the signal's bandwidth but cannot be reconstructed from samples taken at a lower rate. In practice, it is often desirable to use a higher sampling rate. If the sampling rate is ten times the bandwidth, the signal can be reconstructed quite accurately by straight line segments between the sampled values. As the sampling rate is reduced to the theoretical limit of twice the bandwidth, much more complicated approaches to interpolation must be used to reconstruct the original signal. These are covered in Refs. (2,3).

Figures 1 and 2 show the different effects of inadequate sampling rate and inadequate bandwidth. Figure 1 is the signal

$$v(t) = \frac{1}{\sqrt{2\pi\tau_1^2}} \exp[-t^2/2\tau_1^2] + .005 \frac{1}{\sqrt{2\pi\tau_2^2}} \exp[-(t+10)^2/2\tau_2^2] \quad (1)$$

where $\tau_1 = 10$ ms and $\tau_2 = 0.1$ ms. The cross marks are every 1 ms and show where the signal might be sampled with a sampling rate of 1 kSa/s. The spike at $t = -10$ ms is almost completely missed, but if the samples were offset by 0.5 ms there would be one point at the peak of the spike. Figure 2 shows the same signal after having passed through a 500 Hz bandwidth filter. The spike has become much smaller and much wider.

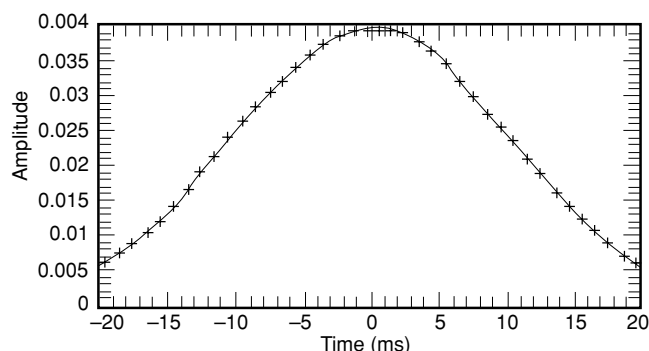


Figure 2. The signal of Fig. 1 filtered with a 500 Hz bandwidth low-pass filter. Only a minuscule hint of the pulse at $t = -10$ ms remains, but the 1 kSa/s sampling adequately represents the resulting signal.

Resolution and Dynamic Range

The resolution of a recorder is a measure of the precision of the digital words that are used to represent the sample values, though this terminology is not universal. Typically, the ADCs in a recorder convert the amplitudes to binary words, and the resolution is expressed as the number of bits in one of these words. If the resolution is N bits, the number of different values representable is 2^N . The resolutions of commercially available recorders range from 6 bits to 24 bits. For some lower speed recorders the ADC converts directly to a decimal number. In this case the resolution will be expressed as the number of digits in the decimal number. If the resolution is N digits, then the number of possible values is 10^N . Frequently ADCs produce 4×10^N different values; these are said to have N -and-one-half digits of resolution.

The dynamic range of a recorder is the ratio of the largest to the smallest signal magnitude that can be measured in a single record. This term has significant uncertainty in its meaning due to the lack of precision of the phrase "smallest signal magnitude that can be measured"; this will be covered in more detail later in the article. In most cases an upper bound on the dynamic range of a recorder with a resolution of N bits is 2^N . The dynamic range is frequently lower than this because of noise signals generated in the recorder that interfere with the measurement of small signals. Dynamic range is usually expressed in decibels rather than as a pure ratio.

Amplitude Accuracy

The accuracy of a recorder is a very complex subject that will be dealt with in more detail later in the article. It can depend very strongly on the signal being measured, and it can't be adequately quantified with a single number or even several numbers. However, for many recorders the accuracy is significantly poorer than the resolution and there are a few specifications that are often given which help quantify it. A fairly overall measure of the accuracy is the *effective number of bits*, N_e . This is specified for a specific frequency (or several of them) and means that the errors in recording a sinusoidal signal of that frequency are of the same size as would be made by an ideal ADC with N_e bits of resolution. Sometimes an accuracy specification is given as a percentage of full scale. This usually refers to the measurement of a direct current (dc) or very slowly varying signals.

Time Base Accuracy

For a digital recorder there are two primary measures of time base accuracy—the sampling rate accuracy and the aperture error. The sampling rate accuracy is the difference between the actual sampling rate and the nominal sampling rate. For time bases controlled by a crystal oscillator, this is usually in the range of 1 Hz/MHz to 100 Hz/MHz. The aperture error is either the maximum or the root-mean-square (rms) deviation between the time at which a sample is taken and the time the sample would have been taken if the sample spacing were perfectly uniform. The effect of aperture error on the accuracy of a sampled signal value is proportional to the rate of change of the signal. The error in amplitude units (e.g., volts) is the product of the aperture error in seconds and the slew rate of the signal in amplitude units per second.

Superior time base accuracy is one of the by-products of digital techniques of recording. When analog tape is used for recording the slippage of the tape in the rollers that propel it and variations in motor speed due to mechanical vibrations play a significant role in the accuracy of the time base. Similar problems occur in almost all analog recording mechanisms.

Output and Controls

A recorder should supply the recorded signals to the user in a form that is convenient for the analysis that is to be carried out. In most cases the analysis will be carried out (at least partially) on a PC, and means of transferring data from the recorder to the PC are essential. If the recorder is an attachment to the PC and records directly to the PC's hard disk, there is no problem. In other cases some additional means must be supplied. Some recorders have an integral floppy disk drive which allows the data transfer with no additional hardware or software. Often it is desirable to transfer data from the recorder to the PC without human intervention. If the recorder supports a common protocol, such as RS232 (a standard for data communications interfaces from the Electronics Industry Association) or the General Purpose Interface Bus (GPIB), this is facilitated. Sometimes it is necessary to transfer data faster than the aforementioned protocols allow, so another means, such as through a parallel port on the computer, must be available.

It is often necessary for the computer to be able to control the recorder. In this case the ability of the recorder to accept control commands through RS232 or GPIB is usually most practical. In complex instrumentation systems, recorders often interface to a standard bus, such as the Computer Automated Measurement and Control (CAMAC) bus or VXI (a high speed bus standard defined by the VXIbus Consortium), and transfer data and control commands through an intermediate controller on the bus.

Many recorders are able to output their data visually along with the outputs to a computer. This is typically through a cathode ray tube (CRT) or a liquid crystal display and/or on paper using a self-contained printing mechanism.

Other Characteristics

A few other characteristics that might be important in some recorder applications are listed here. The existence of differential inputs could be required and, if so, the common mode rejection ratio is of importance. If the recorder has multiple input channels, the maximum cross talk between channels is

of interest. If the input signals are not adequately limited, the over voltage recovery time is important. If the recorder does not record at its maximum rate continuously, then there are a few important characteristics: (1) the range of record lengths that can be recorded, (2) the *cycle time*, which is the required time delay between the end of one record and the beginning of the next, and (3) the *throughput*, which is the rate at which data can be continuously transferred out of the recorder while recording is taking place.

RECORDER CONFIGURATIONS

In this section various common recorder configurations will be covered with a description of the limitations and problems usually found in each configuration. This is not all inclusive, and there are variations on each configuration. In particular, the fact that certain problems and limitations are stated to exist for a particular configuration doesn't imply that designs haven't been successfully carried out to circumvent the particular limitations. For the most part, the more complex configurations are used to overcome sampling rate, bandwidth, or amplitude resolution limitations of the simpler configurations.

ADC—Permanent Recorder Configuration

This is the simplest configuration, in which the input signal passes through a buffer amplifier to an ADC, and the output of the ADC is recorded (usually with intervening buffer memory) on a permanent recording device. The permanent recording device could be hard magnetic disk, digital magnetic tape, magneto-optical disk, floppy disk, optical disk (CD-ROM or audio CD) or, for the slowest data rates, printed paper. These media are listed in order of decreasing speed. In this configuration there may be numerous channels, each with its own ADC, feeding the same recording medium.

Data can be written to PC compatible hard disks at rates of up to 100 Mbit/s. However, if one wishes to record at near the maximum rate, extra care must be taken. One must take special care with the software that is storing data on the disk; going through the Windows or Mac operating system is not likely to work. Also hard disks designated for audio/video use, which don't have lengthy temperature calibration cycles, must be used.

To get a perspective, the 100 Mbit/s rate can record 100 channels of audio with a 50 kSa/s sampling rate and a resolution of 20 bits. Digital audio workstations typically use this configuration with up to 50 channels.

Multiplexer Before ADC

This is the same configuration as the previous section with an analog multiplexer preceding the ADC. This is used for the economy gained by sharing one ADC among many (typically 16 to 64) channels. This configuration leads to potential cross talk problems from two sources. First, the settling time requirement for the ADC or the sample-and-hold preceding it is much more stringent than for single channel operation. Even when input signals are constant, with a multiplexer the input to the ADC will abruptly switch from the level of one channel to the level of another just before the second channel is to be converted. This can cause the reading of the second channel to be influenced by the amplitude of the first. The second cross talk

source is the bringing together on one chip of several analog channels.

ADC and Solid State Memory

When sampling rates are higher than can be recorded on a permanent medium, recorders typically store their data in solid state memory at a high sampling rate and transfer it to permanent memory later at a slower rate. This approach is necessary when sampling rates exceed 10 MSa/s. The sampling rates attainable with this approach are limited only by the ADC. Record lengths are limited by the amount of memory installed in the recorder. As memory densities increase and memory costs decrease, this becomes less of a limitation.

Frequently, for economic reasons, memories that have cycle times longer than that required by the sampling rate are used. The data from the ADC is sent thorough a demultiplexer and interleaved into several memories. For example, to interleave the data into three memories, the demultiplexer would send the first data value to memory number one, the second value to memory number two, the third to memory number three, and the fourth to memory number one, etc. Each memory in this example receives data at one-third the sampling rate.

This configuration is used at sampling rates up to 500 MSa/s. Currently at rates above 100 MSa/s technology limits the resolution attainable in this configuration to eight bits.

Interleaved ADCs

Several ADCs can be combined in one analog channel to increase the sampling rate. With n ADCs each sampling the same analog signal at a sampling rate of r_0 Sa/s, one adjusts the timing so that ADC $k + 1$ samples the signal at time $1/nr_0$ s after ADC k . When the data from the n ADCs are interleaved, one obtains the signal sampled at a rate of nr_0 Sa/s.

The technological problems involved in accomplishing this are larger than they appear on the surface. The sample and hold circuit preceding each ADC must have speed and bandwidth commensurate with the higher sampling rate of the interleaved system. The most difficult problem is caused by the requirement for very precisely matching the frequency response of the path from the recorder input to each ADC. This problem can best be illustrated with a hypothetical example. From results in later parts of this article one can determine that to maintain eight effective bits with an input signal of 500 MHz requires an rms uncertainty of less than 1 ps in the time any ADC samples the signal. A 1 ps delay error will result, at 500 MHz, from a 5×10^{-4} rad (0.18°) phase shift. This means that to meet the stated requirement (eight effective bits) the phase shift of each signal path must match to 5×10^{-4} rad.

Analog Memories

At sampling rates of 100 MSa/s and higher, current ADC technology limits resolution to eight bits. Figure 3 illustrates a configuration that has been used to obtain higher resolution at high sampling rates. The signal to be recorded is sampled repetitively by a sample-and-hold circuit, and these samples are recorded, temporarily, in an analog memory. After the entire record is recorded, the data is read from the analog memory at a much slower rate than at which it was recorded and stored in digital memory. Analog memories have substantial imperfections that must be corrected for before the data is delivered

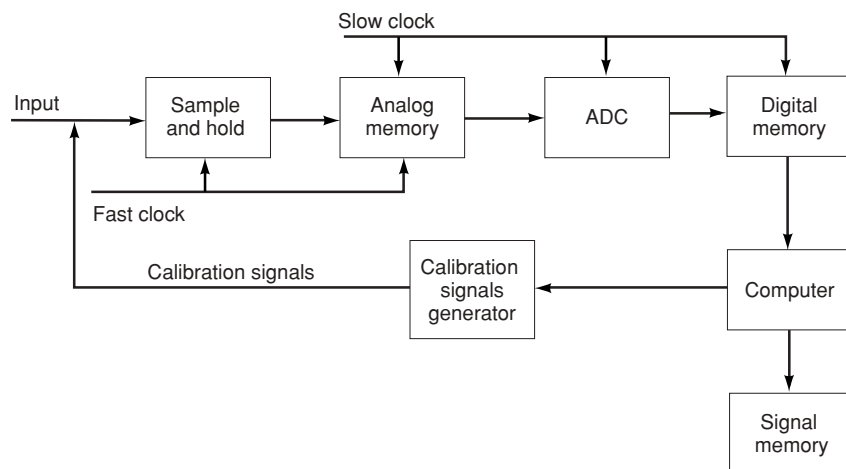


Figure 3. Representation of a recorder using analog memory. The fast clock operates at the sampling rate and stores samples into the memory. The slow clock begins after recording is completed and retrieves data from the memory at a rate at which the ADC can operate. The computer makes corrections to the data to produce a representation of the input signal and controls the calibrations which allow these corrections to be made.

to the user. The advantage of these recorders can be increased signal resolution for small signals and the resulting increase in dynamic range.

The analog memories used are of two types—sequential access and random access. In both types the signal is recorded as the amount of charge on a capacitor. Random access analog memories are constructed much like Complementary Metal Oxide Semiconductor (CMOS) dynamic RAMs, except that the write circuitry stores a variable amount of charge, and the read-out circuitry generates a voltage proportional to the amount of charge stored. Sequential access memories are charge-coupled-devices (CCDs). These contain several hundred charge storage elements coupled in a linear array. When a strobe signal is given to the device, the charge stored in each element is transferred to the next element in the array. The signal to be recorded is transferred to the first element, and the readout is taken from the last element. Analog memories that are adequate for recording applications are not off-the-shelf items; they are generally proprietary devices made by the recorder manufacturers.

This approach has difficult technical problems due to the lack of idealness of the analog memories. The relationship between the charge read in to the memory and the charge read out of the memory is nonlinear and typically different for each memory element. The charge read out may also depend on the charge stored in neighboring memory elements. Recorders in

this configuration typically require extensive internal calibration to characterize the memories and extensive calculation to correct the data. These corrections are usually far from perfect. The errors remaining after the corrections have been made are roughly proportional to the instantaneous signal amplitude, and, for signals near the full scale of the recorder are often larger than would have been obtained with an ADC operating at the full sampling rate. The advantage is potentially lower noise at small signal levels.

Sampling Oscilloscope

If the signal to be recorded can be produced repetitively, it can be recorded with a sampling oscilloscope. Sampling oscilloscopes have bandwidths of up to 50 GHz and unlimited sampling rates. They take one sample on the signal for each occurrence of the signal. Figure 4 shows the functioning of a sampling oscilloscope simplified form. A trigger signal that is synchronized with the signal must be available. To capture the leading edge of the signal, it must occur before the signal. The trigger signal starts a variable delay circuit. At the end of the programmed delay time the sample-and-hold is strobed. The voltage stored by the sample-and-hold circuit is digitized and the value is stored as the amplitude value of the signal. The corresponding time value is that which was programmed

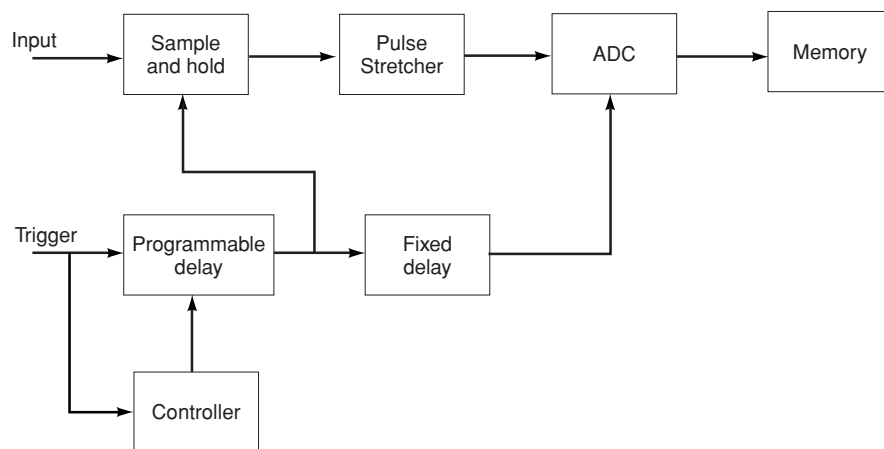


Figure 4. Representation of a sampling oscilloscope. The sample-and-hold produces a pulse whose amplitude is proportional to the signal at the time of the strobe signal. This pulse is stretched to be long enough for its amplitude to be measured by the ADC. One recorded data point is produced for each pair of synchronized input and trigger signals.

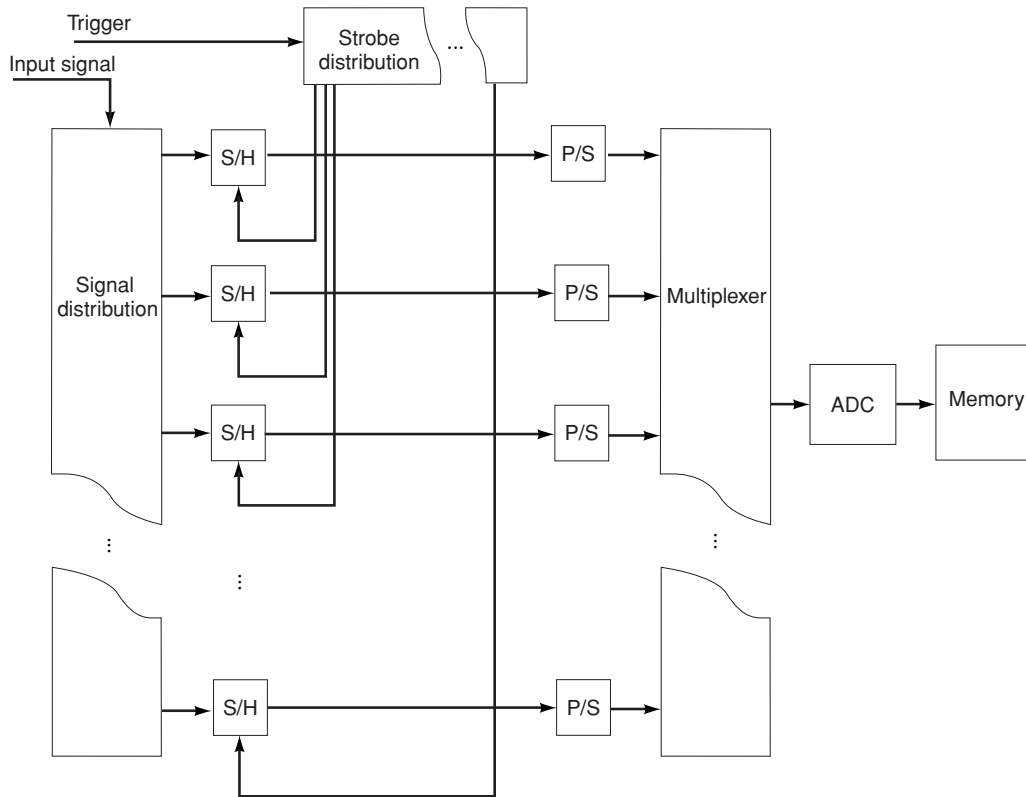


Figure 5. Representation of a recorder using many copies of sampling oscilloscope circuitry to record a single transient.

into the variable delay. The time delay is incremented by the reciprocal of the desired sampling rate, and the next sample is taken. Programming the variable delay to N different delay values gives N samples of the signal at different times. This approach is referred to as *equivalent time sampling*, because the differences in the time values associated with successive samples are not the actual time differences between when the samples were taken but represent equivalent time differences relative to the signal.

The subsystems that make up a modern sampling oscilloscope are very specialized systems. The sample-and-hold circuit must be precisely constructed out of components specifically designed for the purpose. To obtain good linearity, the sample-and-hold usually measures the difference between the applied signal and a feedback signal. The feedback signal is an estimated value of the applied signal. To obtain high accuracy with reasonably long delays, the programmable variable delay is typically an elaborate subsystem.

Multiple Samplers

One can construct recorders that use the very high bandwidth sample-and-hold circuitry of the sampling oscilloscope and record single transients. This is illustrated in Fig. 5 and in Ref. 4. This approach requires one sample-and-hold for each sample of the signal; such recorders have been made which take from 20 to 100 samples. Both the signal to be measured and the strobe signals must be distributed to each of the sample-and-hold circuits. The strobe signal must be delayed to each sample-and-hold taking into account the time, relative to the

trigger signal, that the sample is to be taken and the delay time of the signal to the particular sample-and-hold.

The construction of this type of recorder has all the problems covered earlier for interleaved ADCs, but on a larger scale. The signal must be routed to a large number of distinct circuits while maintaining the same frequency response on each path. Since this type of system is used for extremely high bandwidth, the tolerances on the matching of frequency responses (in particular phase shifts) are very tight.

Equivalent Time Sampling

The sampling oscilloscope is the simplest example of using equivalent time sampling to obtain higher sampling rates for repetitive signals. The same principal can be applied to increasing the effective sampling rate of recorders based on real time digitization. This works much the same as the sampling oscilloscope illustrated in Fig. 4, except that the trigger signal causes the recorder to sample the signal at the rate r_1 rather than to take only one sample. To increase the effective sampling rate by a factor of n , $n - 1$ additional trigger signals cause the recorder to take records delayed, with respect to the repetitive signal, by amounts of k/r_1 . This is illustrated, for $n = 4$, in Fig. 6. This approach is frequently used in digital storage oscilloscopes.

Although equivalent time sampling is applicable to any of the recording methods that have been covered in this article, it is not a trivial matter to add it to a recorder. It requires, with an exception described shortly, synchronization between the sampling times and the repetitions of the signal being measured. This is problematic if the sampling times are derived from a

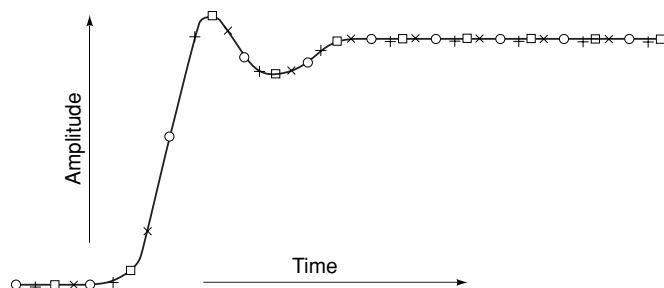


Figure 6. A step signal as recorded using interleaving. Points using the same symbol are recorded with the same ADC. This example shows four interleaved channels.

crystal oscillator. One means of solving this problem is called *random equivalent time sampling*, in which the phase relationship between the trigger signal and the sampling clock is made random, and the time is measured (rather than controlled) between the trigger signal and the first sampling time after the trigger. The time value associated with j th amplitude value of the i th record is then given by $t_{ij} = t_0 + j/r_1 - \tau_i$, where t_0 is a constant and τ_i is the measured time between the trigger and the first sampling point. This approach does not give uniform sampling.

Traveling Wave Cathode Ray Tube

The cathode ray oscilloscope has been one of the highest bandwidth recording devices for decades and remains so today. Modern recorders that use a cathode ray tube (CRT) as the recording device have bandwidths in excess of 5 GHz. An advantage of CRT recorders, besides the very high bandwidth, is their ability to withstand input signals hundreds of times their full scale signal range without damage. Historically, the cathode ray oscilloscope has been converted from a display device to a recording device by attaching a camera to record the display on the face of the CRT on film. Film recording was then replaced by an electronic camera whose output was digitized to form a digital recording of the two dimensional image of the CRT. In some modern CRT recorders the phosphorous face, which converts the electron beam to light, is omitted. Instead, the electron beam writes directly onto a CCD array. The CCD array is read out and the image digitized for analysis. The image is analyzed to produce a record of time-volts pairs.

A traveling wave deflection structure is used on the vertical axis to obtain the extremely high bandwidths that give the CRT an advantage over an ADC. In a traveling wave deflection structure the potential difference to be recorded travels as a wave along the deflection structure at the same speed and in the same direction as the electron beam to be deflected. The speed of the electron beam is typically around one-tenth the speed of light. Why use a traveling wave structure? The angle of deflection of the electron beam is proportional to the time an individual electron in the beam remains in the deflection structure. This time must be several nanoseconds to obtain adequate deflection. With a conventional deflection structure, the transition duration of a recorded step signal will be, approximately, the time it takes an electron to traverse the deflection plate. This limits the bandwidth to 100 MHz or so. With a traveling wave structure the transition duration can be reduced by a factor of 100.

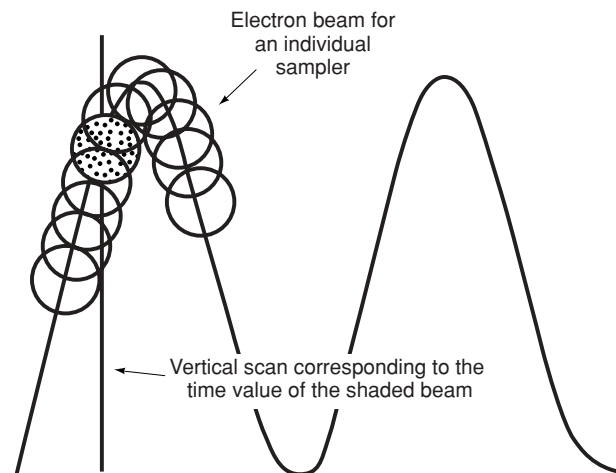


Figure 7. Each circle represents the electron beam striking the face of the CRT corresponding to one time point on the signal. The vertical line represents the scan line of the read out for the time corresponding to the shaded circle. It passes through several circles causing the y-coordinate of the output to be an average of the beam from different times. This results in an effective bandwidth reduction due to the beam size. This phenomenon is not taken into account when reporting the bandwidth of the recorder.

The errors produced by CRT recorders are generally much larger than one might predict from the published specifications. This is because of limitations inherent in the use of electron beams. Typical specifications might be: bandwidth, 5 GHz; vertical resolution, 11 bits; sweep lengths, from 1 ns to 1 ms; number of data points per record, 1000.

With a record length of 1000 samples, the sampling rate = $1000/(\text{sweep length})$. The useful sampling rate is typically an order of magnitude or more lower than this, because of the nonzero spot size of the electron beam. The useful sampling rate is obtained by replacing the number 1000 with the number of electron beam radii that fit along the time axis of the writing area of the CRT. This is because signal samples taken at a higher rate than this will interfere with each other, as illustrated in Fig. 7. The figure understates the magnitude of the problem, because the density of sample points per beam radius is likely larger than shown. Figure 8 illustrates the distortion that occurs because of the nonzero beam size. The peak of the pulse is recorded lower than the true value, because points more than one beam radius below the peak are written on by the beam on the rising and falling edges of the pulse. In the illustration it is assumed that the recorded signal will be halfway between the lower and upper edge of the trace; in actual practice the situation is usually worse than this. The trace will be more intense below the peak than above it, because the beam hits points below the peak twice—once on its way up, and once on its way down. This causes the recorded signal to be even lower than in the figure.

Another type of distortion peculiar to CRT recorders is called *wide-narrow distortion*, which is illustrated in Fig. 9. The input signal is a pure sine-wave. The recorded signal at the left of the display is wider than it should be at the positive peaks and narrower than it should be at the negative peaks. This gradually shifts to the opposite, narrow at positive peaks and wide at negative peaks, on the right side. This problem is inherent in the traveling wave structure. When the signal increases

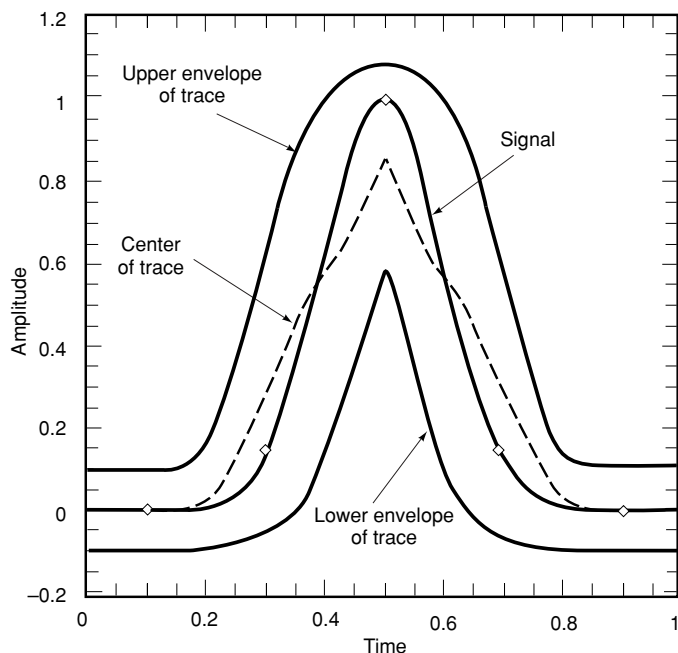


Figure 8. Illustration of the signal distortion which can be caused by the nonzero size of the electron beam in a CRT based recorder. The upper and lower envelopes are determined by having a perfectly circular beam traverse the signal. The read out is taken by scanning the beam vertically and calculating an estimate of the center of the trace. When the display has significant curvature, the center of the trace does not coincide with the signal.

rapidly with time, the voltage across the deflection structure decreases rapidly in the direction of beam travel. This decelerates the electron beam, which effects its deflection in the time direction.

The amplitude resolution of 11 or more bits typically advertised for CRT recorders is misleading. Amplitude values are typically obtained by analyzing the recorded beam intensity profile along a vertical slice (constant time) and calculating the

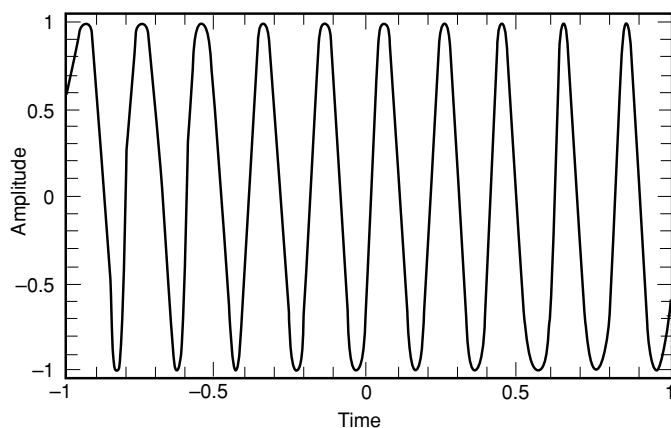


Figure 9. A display demonstrating “wide-narrow” distortion of traveling wave CRTs. The applied sine wave has wide positive peaks and narrow negative peaks on the left, wide negative peaks and narrow positive ones on the right and is relatively undistorted in the center. This phenomenon distorts the recorded width of narrow pulses.

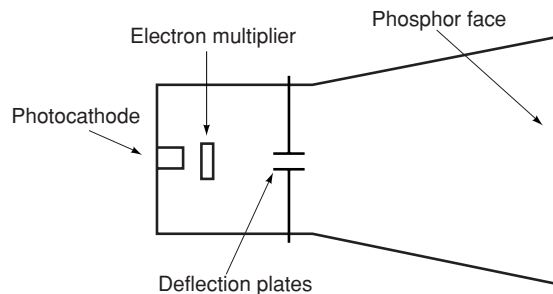


Figure 10. Top view illustration of a streak camera based recorder. Each channel is a dot of light on the photo cathode. The dots are arranged in straight line vertically. Electrons are emitted from each location on the photo cathode in a number proportional to the instantaneous light intensity. After multiplication the electron stream is deflected on the horizontal axis and recorded in the same manner as in a CRT.

centroid (or some closely related measure of the center of the trace) and relating this to a voltage value. The precision specified is usually the precision to which this calculation is rounded. The actual precision of the resulting data is close to seven bits in current commercially available recorders.

Streak Camera

A streak camera has several similarities to a CRT. Information is recorded by an electron beam striking a phosphor face. The image on the face is digitized and analyzed to determine the amplitude versus time of the recorded signal(s). The time information is obtained, as with a CRT, by sweeping the beam across the face of the *streak tube*. The amplitude information, however, is encoded in the intensity of the beam. The signals to be recorded must be converted to light whose intensity is proportional to the amplitude of the signal. The light strikes a photo-cathode that emits a number of electrons proportional to its intensity. The electron beam is amplified and swept across the face of the tube; see Fig. 10. About 20 channels can be recorded simultaneously on a streak camera. The light for each channel is placed at a different location along the vertical axis of the photo-cathode. The data for each channel is one horizontal streak across the face of the tube.

The conversion of light to electrons at the photo-cathode is nearly instantaneous, yielding potential bandwidths in excess of 20 GHz. It is often the case in making very high bandwidth recordings that the signals to be recorded are converted to light to be transmitted to the recording station by fiber-optic cable, thus eliminating the high frequency skin effect loss of coaxial cable. In such situations the streak camera is an especially suitable recorder.

RECORDER ERROR MEASURES

One of the most important features of a recorder is the accuracy with which it records. This is a complex subject, because the errors are usually signal dependent. A number of measures of recorder accuracy and test methods for determining them have been established over the years and are described in this article. Most of those described are covered in Refs. 5 and 6; others are covered in Ref. 7.

The errors made by a recorder are often classified as either *static errors* or *dynamic errors*. Static errors are those that are

evident in recording DC or very slowly varying signals; dynamic errors are those that are frequency dependent and typically disappear at low frequencies. Sometimes the classification is not clear and is made somewhat arbitrarily. Static errors will be covered first, followed by dynamic errors.

Sine Wave Tests

Sine wave tests are among the most useful tools for evaluating recorder errors. A sinusoidal signal of specified amplitude and frequency is recorded yielding a sequence (t_i, y_i) of time-amplitude pairs. One then performs a least squares fit to the data by determining values of A_0 , B_0 , C_0 , and ω_0 that minimize

$$\sum_{i=1}^M [y_i - A_0 \cos(\omega_0 t_i) - B_0 \sin(\omega_0 t_i) - C_0]^2 \quad (2)$$

where M is the number of data values used in the analysis. Details on the calculation of the unknown parameters can be found in Ref. 5. Once the coefficients are determined the *residuals* are given by

$$r_i = y_i - A_0 \cos(\omega_0 t_i) - B_0 \sin(\omega_0 t_i) - C_0 \quad (3)$$

The residuals are a good approximation to a significant portion of the errors in the recording process. They can be analyzed in either the time domain or the frequency domain (Refs. (5,6,7,8) and the following sections of this article) to give useful information. Of particular interest is the rms value of the residuals, σ_r , given by

$$\sigma_r^2 = \frac{1}{M} \sum_{i=1}^M r_i^2 \quad (4)$$

Since the amplitude, phase, frequency, and dc offset of the input signal are estimated from the data, they don't have to be accurately known to perform sine wave tests. For the same reason, the frequency response of the cabling connecting the sinusoidal signal source to the recorder, and of the recorder itself, is not an important factor in sine wave tests.

If the frequency of the test signal is chosen so that an integer number of cycles occur in one record length, that is,

$$f = \frac{J}{M} f_s \quad (5)$$

where f is the frequency of the test signal, f_s is the sampling rate, and J is an integer, then the least squares Eq. (2) can be solved by taking the discrete Fourier transform of the data. The constant C_0 is the 0 frequency term in the DFT; A_0 and B_0 can be derived from the two terms at the frequency, f , and all other terms in the DFT represent the residuals.

There are two main reasons for the prevalence of the use of sine waves in evaluating recorder errors. The first is that sine waves can be produced with very high accuracy; their purity can be measured to even higher accuracy with a spectrum analyzer, and their purity can be improved on by the use of filters. Current technology doesn't permit the economic production of signals of other shapes that are known to the accuracy we expect of many recorders. The second reason for the prevalence of sine waves in this application is they allow one to ignore the frequency response effects of the recorder and the test setup.

Quantization Error

Quantization error is an attribute of analog to digital converters. It arises from the finite number of output values that can be produced by an ADC. Figure 11 is a plot of the output of an ADC versus the input voltage. Each interval of input voltages over which the output code remains constant is called a *code bin*. The difference between the voltage at the right edge of a code bin and the voltage at the left edge is called the *code bin width*. Ideally, the code bin width is a constant, Q . The voltage value assigned to a particular output code is assumed to be the value at the center of the code bin. The difference between the input voltage and the assigned voltage is called the *quantization error*. If the input voltage is equally likely to be any value in the range of the ADC, then the quantization error is equally likely to be any value between $-Q/2$ and $+Q/2$. In this case the average quantization error is zero and the rms quantization error is given by

$$\text{rms quantization error} = \sqrt{\int_{-Q/2}^{+Q/2} \frac{x^2 dx}{Q}} = \frac{Q}{\sqrt{12}} \quad (6)$$

Although the quantization error is a deterministic nonlinearity in the transfer function, it is frequently referred to as

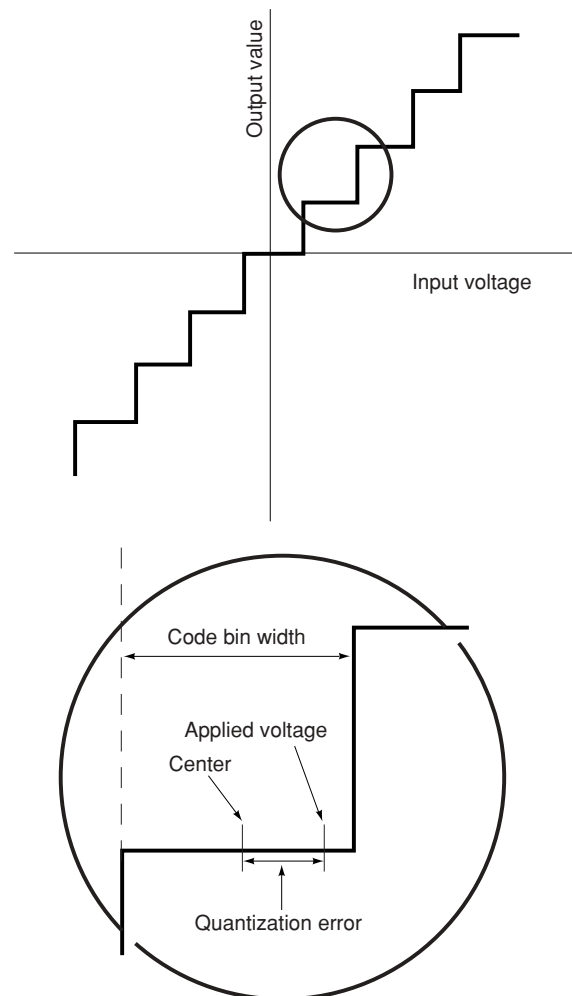


Figure 11. The stair step transfer function of an ADC.

quantization noise. In many cases the sequence of values of the quantization error is most practically analyzed as white noise with an rms value given by (6).

Gain and Offset Errors

If one thinks of a recorder as an amplifier followed by an ideal recorder, then the gain and offset errors are the deviations of the amplifier gain and offset from their nominal values. There are different means of measuring these errors that yield slightly different results. For a sequence of ordered pairs (v_i, v_r) , where v_i is an input voltage and v_r is the corresponding recorded voltage, the gain and offset are defined by the following relation,

$$v_i = Gv_r + v_0 + \epsilon \quad (7)$$

where G is the gain, v_0 is the offset error, and ϵ is the error (different for each value of v_i). The values for v_r in this equation are already corrected for the nominal gain and offset of the recorder; thus, if the recorder were ideal, we would have $G = 1$ and $v_0 = 0$. Values of G and v_0 are determined by fitting a straight line to the (v_i, v_r) pairs. Different means of fitting the line will yield different values. Values for v_i should be selected so that quantization error has negligible effect on the results. Values of v_r should be obtained by averaging so that random noise has negligible effect on the results. The *gain error* is $G - 1$ and is usually expressed as a percentage. The offset error is expressed in input units (e.g., volts).

The gain and offset errors are treated separately from other errors, because they are often larger than other errors. If necessary, the gain and offset errors can be corrected for by using the results of recording two known voltages.

Dynamic Range and Noise

The use of the terms dynamic range and noise is not totally consistent in the industry, reference (9). The most commonly used definition for dynamic range for recorders is

$$\text{dynamic range} = 20 \log \frac{\text{maximum rms signal}}{\text{rms noise}} \text{ dB} \quad (8)$$

Both the numerator and denominator in Eq. (8) are subject to multiple interpretations. The numerator is usually taken to mean the maximum rms sine wave which, for a recorder that covers the voltage range of from $-A$ to A , is $A/\sqrt{2}$. The denominator is usually taken to be the sum (in quadrature) of the noise present with no signal and the quantization noise. This quadrature combination is accurately approximated by the rms residuals from a sine wave test with a low amplitude ($5Q$ to $20Q$) input signal. For audio recorders the rms noise is often calculated with a frequency domain weighting (A weighting) based on the sensitivity of human hearing to low level sounds (11). For an ideal N bit digital recorder, one in which the only source of noise is quantization noise, the dynamic range is given by

$$\text{maximum dynamic range} = 6.02N + 1.76 \text{ dB} \quad (9)$$

Table 1 gives the dynamic range from Eq. (9) for typical values of N .

Table 1. Maximum Attainable Dynamic Range as a Function of Number of Bits

Number of Bits	Maximum Dynamic Range (dB)
6	38
8	50
10	62
12	74
14	86
16	98
18	110
20	122
22	134
24	146

Integral Nonlinearity

The concept of integral nonlinearity (INL), as it has been used for decades, is illustrated in Fig. 12. The curve represents the voltage reported by the recorder as a function of the applied voltage. The straight line represents a best fit to this curve. The INL is the maximum difference between the curve and the straight line. The difference is typically expressed as a percentage of full scale for analog instruments.

For recorders that use an ADC the situation is a little more complicated. As seen in Fig. 11, the ideal curve for an ADC is a stair step rather than a straight line. The stair step necessarily differs from the straight line by $\pm Q/2$. This difference is not considered part of the integral non linearity. For an ADC only the values of the transition levels, the values of input signal for which the output jumps from one code to the next, are considered. The INL is the maximum difference between any transition level and its ideal value. Determining the INL for an ADC based recorder requires determining the transition levels. This is a lengthy process covered in Refs. (5,6,7). If the recorder has a noise level significantly larger than the quantization noise, the INL can be estimated without determining the transition levels. Points on the curve of Fig. 12 can be measured by recording the average outputs to known arbitrary inputs. This works,

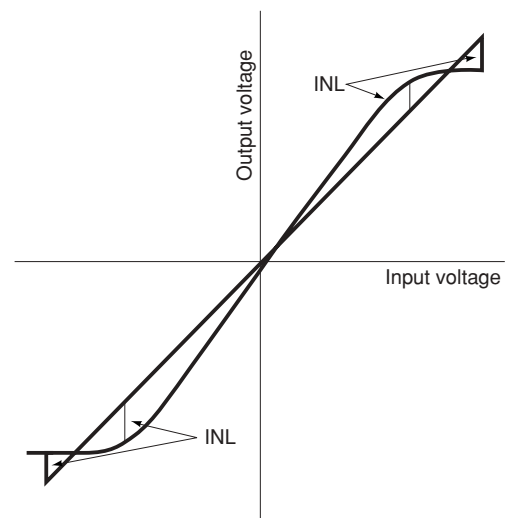


Figure 12. Illustration of INL as defined for analog instruments. The curved line is the dc transfer function of the instrument. Its maximum deviation from a straight line (occurring at four different places in this figure) is the integral nonlinearity.

Table 2. Improvement in Analog INL Due to Noise

σ/Q	INL
0	.5
.01	.48
.02	.47
.05	.38
.1	.29
.2	.14
.3	.054
.5	2.2×10^{-3}
.6	2.5×10^{-4}
.7	2.0×10^{-5}

because the noise straightens out the stair step in the average measurements. Table 2 gives the amount of apparent INL caused by the stair step as a function of the rms noise divided by Q .

Differential Nonlinearity

The concept of differential nonlinearity (DNL), like integral nonlinearity, has been used for decades in analog instrumentation. It is defined in terms of the derivative of the recorded voltage with respect to the input voltage, and is a measure of how much this derivative deviates from a constant. Letting

$$S_M = \max \left[\frac{dv_r}{dv_i} \right], \quad \text{and } S_m = \min \left[\frac{dv_r}{dv_i} \right]$$

DNL is defined as

$$\text{DNL} = 200 \frac{S_M - S_m}{S_M + S_m} \% \quad (10)$$

This is illustrated in Fig. 13, which is the derivative of the transfer function in Fig. 12.

For ADC based recorders the derivative of the stair step like ideal transfer function jumps between zero and infinity, and Eq. (10) is not appropriate. A perfectly linear ADC is one for which all code bin widths are constant. DNL for an ADC is the following measure of the maximum deviation of any code bin

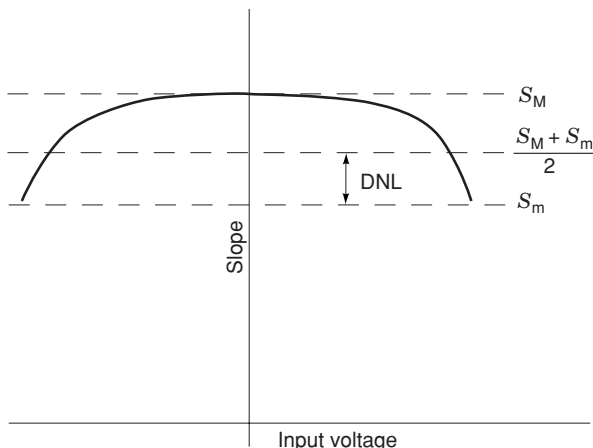


Figure 13. Illustration of DNL for an analog instrument. The curve is the derivative of that in Fig. 12.

width from a constant

$$\text{DNL} = \max \left| \frac{W(k) - Q'}{Q'} \right| \quad (11)$$

where $W(k)$ is the width of the k th code bin, and Q' is the average code bin width. The unit typically supplied with this measure is lsb (for least significant bit), while “code bin width” would be a more grammatically correct unit. For ADCs whose noise level is larger than their code bin width, DNL is not a very useful specification.

Frequency Response

The frequency response characterizes the error that is usually the dominant dynamic error of a recorder. Use of a frequency response to characterize a system is only valid if the system is linear and time invariant (2). This is a good approximation for most high quality recorders. One can imagine the recorder as a linear time invariant filter followed by blocks that produce all of the other sources of error of the recorder. For signals for which the errors are small ($<10\%$), the order of the blocks has a second order effect on the overall error. The error produced by the equivalent filter at the input of the recorder is the *frequency response error*. Keep in mind that the frequency response error may actually be caused by limitations in several parts of the recorder; placing the problem in a single filter is just a mathematical convenience.

The quantity of most interest is actually the *impulse response*, $h(t)$, of the recorder. If $x(t)$ is the input signal to be recorded, the recorded signal (accounting only for frequency response errors) is given by

$$r(t) = \int_{-\infty}^{\infty} x(t-t')h(t') dt' \quad (12)$$

The integral in Eq. (12) is called the *convolution integral*. It allows one to calculate the recorded signal for any input signal, and thus estimate the frequency response error—the difference between $r(t)$ and $x(t)$. This error depends on the signal being recorded.

The frequency response is the Fourier transform of the impulse response,

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp(-j2\pi ft) dt \quad (13)$$

The complex frequency response is usually expressed as a gain, $G(f)$, and a phase shift, $\phi(f)$, both of which are real valued and are related to $H(f)$ by

$$H(f) = G(f) \exp[j\phi(f)] \quad (14)$$

These are the gain and phase shift observed when recording a sinusoidal signal of frequency, f . In both Eqs. (13) and (14) j is the imaginary unit. For an ideal recorder (one with no frequency response error) $G(f)$ will be unity for all f , and $\phi(f)$ will be a linear function of f .

Typically the specifications for a recorder will give a lower and upper frequency between which $G(f)$ is within a certain tolerance of unity. This tolerance is usually 3 dB, but sometimes 1 dB is used. One should bear in mind that with a tolerance of 3 dB (1 dB) this specification only gives a range of frequencies

for which the gain error is less than 30% (11%), and tells nothing about phase shift errors. If one needs to record signals to better accuracy than these tolerances, more detail concerning the frequency response error is required. A convenient approach (5) is to determine the gain error and non-linear phase error as a function of frequency. The gain error is $|1 - G(f)|$, and is expressed as a percentage. The nonlinear phase error is the phase shift with a straight line subtracted, usually the line that passes through the origin and passes through the phase shift curve at a phase of $\pi/2$ rad (45°).

Measuring the Frequency Response. The most common method for measuring the frequency response of a recorder is to record the step response, numerically differentiate it (using a simple first difference) to determine the impulse response, and calculate the frequency response by taking the discrete Fourier transform of the result. The step response is recorded rather than the impulse response for a number of reasons. It gives more accurate results at low frequencies at which accuracy is most desired; it is less subject to aliasing errors than the impulse response (12), and the amplitude of the recorded step response is more easily predictable. Other types of signals are used to determine the frequency responses of systems (13), but they have not found wide acceptance in recorders.

If one only wants crude information about the frequency response, the results of sine wave tests can be used. This can determine if the gain error at certain frequencies is less some tolerance. The ratio of the fitted amplitude of the sine wave to the known input amplitude is the gain at the frequency of the test signal.

Step Response

Frequently, characteristics of the step response of a recorder are used as quality measures. These include the settling time, the amount of overshoot and the 10 to 90% transition duration (or rise time). The *settling time* is the amount of time it takes the step response to stabilize to within a certain tolerance of its final value. For example, a specification that the settling time is $5 \mu\text{s}$ to 0.1% means that for any time later than $5 \mu\text{s}$ after the onset of the step, the recorded value stays within 0.1% of its final value. Settling time is frequently the only dynamic specification that relates to precision measurements. The amount of overshoot is zero unless the value of the step response exceeds its final value at some time. Otherwise, the overshoot is the difference between the peak value of the step response and the final value, expressed as a percentage of the final value. Sometimes an overshoot specification will be contingent on the 10 to 90% transition duration of the input signal being longer than some specified value. The settling time and overshoot are illustrated in Fig. 14.

The 10 to 90% transition duration is a crude measure, as is the 3 dB bandwidth, of the response time of the recorder. Two step responses can have the same rise time and have much different distorting effects on a signal. To seriously estimate dynamic errors one should have a complete step response record and use the convolution integral to estimate distortion. Software to do the calculation is readily available.

Noise and Distortion

The sine wave tests described earlier can be used to measure the combined effect of many error sources. The error sources

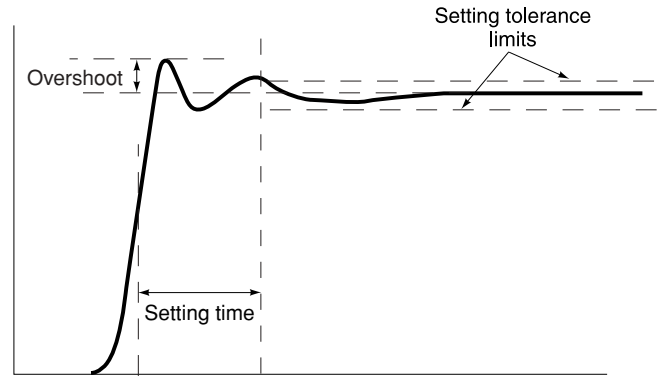


Figure 14. An pictorial example illustrating the definitions of settling time and overshoot.

left out are gain and offset errors, frequency response errors, and errors associated with uncertainties in trigger delay. Most other sources of error will show up in sine wave tests. In fact, one of the advantages of these tests is that they expose sources of error that aren't even known to the tester.

Signal to Noise Ratio, Effective Bits and Dynamic Range. Of greatest utility is the rms value of the residuals, given by Eq. (4). This is referred to as the total *noise and distortion*, and will usually depend of both the amplitude and frequency of the applied signal. The noise and distortion is often referred to as simply the noise. A related quantity is the *signal to noise ratio* (SNR), given by

$$\text{SNR} = 20 \log \left(\frac{A}{\sigma_r} \right) \text{ dB} \quad (15)$$

where A is the rms amplitude of the applied signal, and σ_r is the rms value of the residuals. The amplitude is selected so that the signal covers nearly the entire range of the recorder. Such signals are called *large amplitude* signals. The SNR looks very much like the dynamic range, given by Eq. (8). They are both the ratios (in decibels) of the rms amplitude of a large sine wave divided by the rms value of the noise. There is, however, a significant difference. For SNR, the noise used in the ratio is the noise that occurs in the presence of the large signal. For dynamic range it is the noise in the presence of a small signal.

Another commonly used related quantity is the *effective bits*, E , given by

$$E = \log_2 \left(\frac{V}{\sqrt{12}\sigma_r} \right) \text{ bits} \quad (16)$$

where V is the full scale range of the recorder (maximum recordable voltage minus minimum recordable voltage). If the only source of noise were quantization noise, then $E = N$, the number of bits. The number of effective bits is reduced by one for each doubling of the rms noise. The SNR and the effective bits are related by,

$$\text{SNR} = 6.02E + 1.76 + 20 \log \left(\frac{2A}{V} \right) \text{ dB} \quad (17)$$

For large amplitude signals (>90 % of full scale) the last term

is between -0.9 and 0 . If E_0 is the effective bits measured for a very small input signal, the dynamic range is given by (see Eq. 9)

$$\text{dynamic range} = 6.02E_0 + 1.76 \text{ dB} \quad (18)$$

Time Errors Versus Amplitude Errors

One can separate errors on the time axis from errors on the voltage axis by examining the variation in effective bits with frequency. If one assumes errors τ_i on the time axis and e_i on the voltage axis that are independent of each other and independent of the signal, then the recorded signal for sinusoidal input is

$$\begin{aligned} y(t_i) &= A \sin(2\pi f t_i + \tau_i) + e_i \\ &\cong A \sin(2\pi f t_i) + e_i + 2\pi A f \tau_i \cos(2\pi f t_i) \end{aligned} \quad (19)$$

where τ_i is the i th discrete sampling time, and the error induced on the voltage axis by that on the time axis is estimated using a first order Taylor expansion of the sin function. The second and third terms are a very good approximation to the residuals. This gives for the rms residuals

$$\sigma_r^2 = \sigma_e^2 + 2\pi^2 f^2 A^2 \sigma_\tau^2 \quad (20)$$

where σ_e and σ_τ are the rms values of e_i and t_i . The first term dominates at low frequencies, the second term at high frequencies. Figure 15 shows this converted to effective bits and plotted against frequency on a logarithmic scale. At low frequencies this is constant; at high frequencies it has a slope of -1 bit per octave. The time component of the error can be determined from the frequency f_B at which the two lines intersect by the relation

$$\sigma_\tau = \frac{\sigma_e}{\sqrt{2\pi} f_B A} \quad (21)$$

where σ_e is the rms value of the residuals at low frequency.

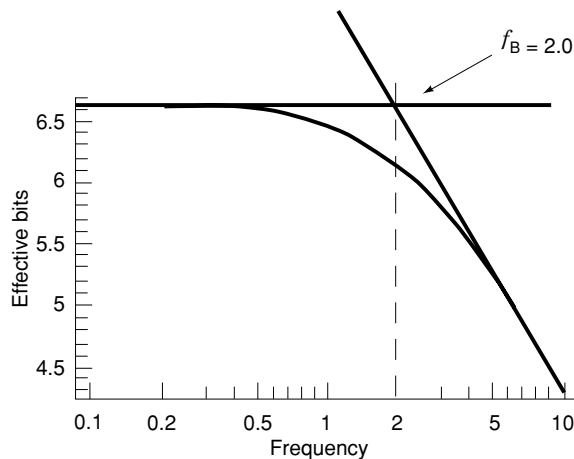


Figure 15. Effective bits versus frequency for a combination of amplitude errors and time errors. The time errors cause the 1 bit per octave roll off at high frequency. The frequency, f_B of the intersection of the two straight lines can be used in equation 21 to estimate the rms value of the time errors.

Harmonic Distortion. If the recorded signal resulting from an applied sine wave contains frequencies that are multiples of the applied frequency, the recorder is said to have harmonic distortion. Harmonic distortion is a consequence of nonlinearity. For a static nonlinearity for which the recorded signal, $v_r(t)$, is given by $v_r(t) = f(v_i(t))$, there is a direct relation between the function, f , and the harmonic distortion. If f is a polynomial of degree n , then there will be harmonic distortion up to order n . If a signal of the form $v_i(t) = A \cos(\omega t)$ is applied, and we let $f_A(x) = f(x/A)$ and expand $f_A(x)$ as

$$f_A(x) = \sum_{k=0}^n a_k T_k(x) \quad (22)$$

where the $T_k(x)$ are the Chebychev polynomials (13,14), then the recorded signal has the form

$$v_r(t) = A \sum_{k=0}^n a_k \cos(k\omega t) \quad (23)$$

Harmonic distortion is specified as a ratio of the combined rms value of the harmonics to the rms value at the signal frequency. The ratio is expressed either as a percentage or in dB. The units used when expressing it in dB are dBc, meaning decibels relative to the carrier. For the example of Eq. (23) we have

$$\text{harmonic distortion} = 10 \log \left[\frac{\sum_{k=2}^n a_k^2}{a_1^2} \right] \text{ dBc} \quad (24)$$

Harmonic distortion typically increases with amplitude and is usually measured with large amplitude signals.

The source of harmonic distortion described previously is closely related to INL, a static nonlinearity, and is, therefore, independent of the frequency of the applied signal. Harmonic distortion can result from dynamic nonlinearity and be frequency dependent. A typical example of a dynamic nonlinearity is a time delay that is proportional to the slew rate of the signal, that is, $v_r(t) = v_i(t - \tau)$, where $\tau = \epsilon(dv_i/dt)$. Estimating the time delay effect by first order expansion as for Eq. (19) gives

$$v_r(t) = v_i(t) - \epsilon(dv_i/dt)^2 \quad (25)$$

Substituting a sinusoidal signal for $v_i(t)$, one obtains second harmonic distortion which is proportional to the signal's amplitude squared and to its frequency squared.

Harmonic distortion is most conveniently measured by performing a sine wave test and calculating the DFT of the recorded signal. The record length must be an integer number of cycles of the applied signal (see Eq. 5) and a large signal should be used. The DFT values are converted to decibels and the dB value for the applied frequency is subtracted from all of the values. This gives harmonic distortion readings directly in dBc. Figure 16 shows a simulated example result. The recorder sampling rate is 2 units, and the record length for the DFT is 1024 points. The input signal was full scale in amplitude with a frequency of 0.2012. The recorded signal has harmonic distortion given in Fig. 16 and had eight bit quantization noise applied.

Interleaving Errors and Noise Spectrum. The methods used to measure harmonic distortion can be used to quantify

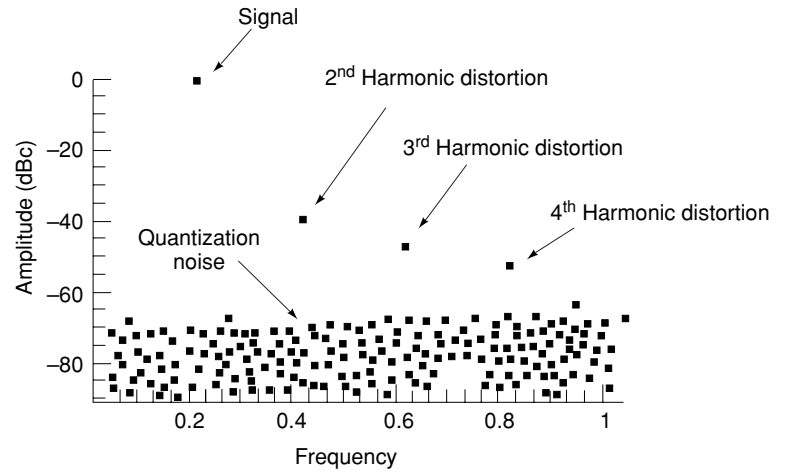


Figure 16. DFT of sine wave test results showing 1% second 0.5% third and 0.25% fourth harmonic distortion. A record length of 1024 points was used. The points below -60 dBc are the result of 8 bit quantization noise. This noise floor can be made as low as desired by using a sufficiently long record length.

interleaving errors. When a number of recording channels is interleaved to increase the sampling rate (see section titled “Interleaved ADCs”), errors result from the various channels not being perfectly matched. These are called interleaving errors. The interleaving factor, n , is the number of recorders interleaved. The sampling rate for each recorder is f_s/n , where f_s is the sampling rate of the recorder. When a sine wave of frequency f is recorded, interleaving errors produce additional signals at frequencies

$$\text{interleave error frequencies} = kf_s/n \pm f \quad (26)$$

for integer values of k . When these frequencies are larger than the Nyquist frequency, $f_s/2$, they get aliased down to frequencies in the Nyquist band. However, at most $n - 1$ distinct frequencies below the Nyquist frequency are produced.

The measurements and data reduction for examining interleaving errors are identical to those for harmonic distortion. There is, however, an additional restriction on the input frequency, f . The frequency should satisfy

$$f = \frac{Kn}{M} f_s \quad (27)$$

where K is an integer, and M is the record length. This causes all of the frequencies created by the interleaving errors to be DFT frequency values. Figure 17 shows results with $f_s = 2$, $n = 4$ and $f = 0.297$. All of parameters are the same as for Fig. 16 except that the harmonic distortion has been replaced by distortion due to the four interleaved channels having different gain errors. The gain errors for the channels are (in succession) 3%, 1%, -1% and -3% .

Spurious Signals—Spurious Free Dynamic Range. Spurious signals are any frequencies, other than harmonics of the applied signal, that occur in the recorded signal. The frequencies generated by interleaving errors are just one example. The spectral lines related to the spurious signals are called *spurs*. The test methods used to determine the magnitudes of the spurs are identical to that of the last two sections.

The negative of magnitude of the largest spur (in dBc) is called the *spurious free dynamic range* (SFDR) of the recorder. This specification is much more common for ADCs themselves

than recorders in which they are used. Note that the SFDR can be much larger than the previously defined dynamic range, which was based on the noise level in the time domain. The observed noise floor (due to random noise, quantization noise, and various other sources) is much smaller in the frequency domain than in the time domain, and it can be made as small as desired by increasing the record length. Therefore, the SFDR is unaffected by this noise. SFDR is only meaningful if the recorder is going to be used in narrow band applications, such as communications or spectrum analysis.

Intermodulation Distortion. When two sine waves of different frequencies, f_1 and f_2 , are passed through a nonlinear system, new signals are produced with frequencies $k_1 f_1 \pm k_2 f_2$, for integer values of k_1 and k_2 . This is referred to as *intermodulation distortion* (IMD). The phenomena in the recorder that are responsible for IMD are the same as those responsible for harmonic distortion. However, there are situations in which harmonic distortion measurements will lead one to believe that the recorder is more accurate than it actually is. When input frequencies are above one-half of the bandwidth of the recorder,

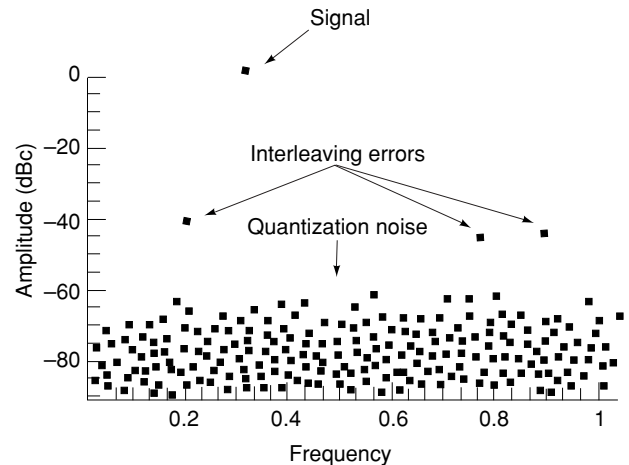


Figure 17. DFT of sine wave test results showing the results of the interleaving errors for the example given in the text. The interleave error frequencies given by Eq. (26) are 0.203, 0.703, and 0.797.

harmonics can be attenuated. This can hide some of the high frequency errors. If two high frequency signals are summed, the nonlinearities will cause a number of spectral lines that are well within the bandwidth of the recorder to be produced, making the errors easier to detect and quantify.

There is another situation in which IMD measurements have an advantage over harmonic distortion measurements. One can use test signals with harmonic distortion larger than the distortions one wishes to measure, because the harmonics will be at different frequencies than the IMD products.

Noise Spectrum for Cathode Ray Tube Based Recorders. Sine wave tests and frequency domain analysis of the residuals are also useful for recorders that use a CRT as the recording device. However, because the error phenomena are different some changes in approach have to be made. Two potentially substantial sources of error in a CRT, distortion of the image due to misalignment and time base nonlinearity, produce errors at the same frequency as the applied sinusoidal signal. The spectrum of the errors is spread out over a frequency range on the order of the reciprocal of the record length. This causes two difficulties when one attempts to use data displays such as those of Figs. 16 and 17. First, since the errors are at the same frequency as the signal, they become masked by the signal. Second, since the errors are spread out over a frequency band, their rms value isn't related to the peak value in the spectrum but to an integral. The first problem can be eliminated by performing the DFT on the residuals rather than on the recorded signal, the second by displaying the integral of the power spectrum (8).

The types of distortion shown in Figs. 8 and 9 also have particular signature in the frequency domain. The distortion at the signal peaks due to the spot size exhibits itself as odd order harmonic distortion, which is independent of amplitude and increases rapidly with frequency. The wide-narrow distortion has the same form as Eq. (25) with ϵ a linear function of t that passes through 0 near the mid point in time of the record. This produces second harmonic distortion and a zero frequency distortion proportional to $(Af)^2$, where A is the amplitude and f is the frequency of the input signal. The distortions at these two frequencies are also spread out over a frequency ranges on the order of the reciprocal of the record length.

The random noise and the quantization noise of ADC based recorders are typically white, having nearly constant power spectrum amplitude up to the Nyquist frequency. This is not true for CRT based recorders. The random noise for CRT recorders begins rolling off with increasing frequency at a cut-off frequency related to the spot size of the CRT beam. This limits the SNR enhancement that can be attained through low pass filtering.

Time Base Errors

The two types of time base errors are long term stability and aperture errors. The long term stability is a measure of the long term drift of the sampling rate from its nominal value. This was covered in the earlier *Time base accuracy* section. This error shows up in sine wave tests as a deviation between the fitted frequency and the actual frequency of the input signal. This error doesn't show up in the residuals.

Aperture errors do show up in the residuals of sine wave tests. In the *Noise and distortion* section various sources of

aperture error were related to distortions that appear in sine wave tests and whose magnitudes increase with the frequency of the input signal.

OTHER RECORDER QUALITY ATTRIBUTES

There are some other recorder attributes that frequently appear in specifications that don't quite fit into the previous categories.

Cross Talk

Cross talk is the phenomenon, in multichannel recorders, of a small part of the signal into one channel appearing in other channels. The cross talk for any pair of channels, numbered i and j , is defined as

$$c_{ij} = \frac{\text{rms signal recorded on channel } i}{\text{rms signal applied to channel } j} \quad (28)$$

This is measured with all channels terminated in their usual source impedance and with a sinusoidal input signal applied only to channel j . The cross talk typically increases with frequency. The cross talk in practice is the sum of the values from all channels.

Overvoltage Recovery Time

The situation in which the input signal to a recorder exceeds the recording range for a period of time but doesn't exceed the maximum safe input voltage of the recorder is called an overvoltage. The *overvoltage recovery time* is the length of time, after the overvoltage has been removed, for recording at the specified accuracy to resume. The recovery time may depend on the magnitude and duration of the overvoltage.

Trigger Delay and Trigger Jitter

Most high speed (above audio frequency) recorders begin recording upon the receipt of a trigger signal. The *trigger delay* is the length of time between the trigger signal and the first time in the recorded signal. This time can be negative for recorders with pretrigger capability. The trigger jitter is (with the exception given shortly) the standard deviation of the trigger delay. For recorders with a continuously running clock signal, there is an inherent jitter of $1/f_s \sqrt{12}$, where f_s is the sampling rate. Some recorders provide a measurement of the time between the trigger signal and the first clock signal. For these recorders the trigger jitter is the standard deviation of the error in this measurement.

Common Mode Rejection Ratio

Common mode rejection ratio (CMMR) applies to recorders with differential amplifiers at the input. The meaning of CMMR for recorders is the same for amplifiers (15), the ratio of the output for differential signals to that for common mode signals of the same amplitude. The CMMR is especially easy to measure for recorders, because a DFT can be performed on the recorded data to more accurately determine the magnitude of a small common mode signal that is below the noise level. The CMMR is usually frequency dependent, generally increasing with increasing frequency.

RECORDER APPLICATION COMMENTS

Determining Recorder Sampling Rate and Bandwidth Requirements

The cost of recorders can increase drastically as sampling rate and bandwidth increase. Furthermore, the amount of data that must be stored increases linearly with sampling rate for a fixed record length (measured in time units). Hence, overestimating the bandwidth and sampling rate requirement for a recording system can drastically increase its cost, while underestimating them can lead to complete loss of the information that one desires to record. Thus, proper estimation of these requirements is an important part of recorder selection. The author has frequently seen errors of a factor of ten in estimated sampling rate requirements and has found errors by a factor of one hundred not to be uncommon.

The Sampling Theorem and Its Misuses. The sampling theorem (1,2) states that a signal of bandwidth f_B Hz can be exactly reconstructed from samples taken at a rate of $2f_B$ Sa/s. There is a common folklore that says that, to be safe, one should record a signal at a rate of ten times its bandwidth. These concepts lead people to ask "What is the bandwidth of the signal to be recorded?" as the first step in determining sampling rate requirements. This is, in most practical situations, a fruitless approach. The bandwidth in the sampling theorem refers to a frequency above which the Fourier transform of the input signal is exactly zero; real signals never have this property for any frequency upper limit. So for real signals the "bandwidth" is not a well-defined quantity. Often people will use the -3 dB bandwidth of a signal as an approximation. This, coupled with the five times over sampling by sampling at ten times the bandwidth rather than two times, seems reasonable. The error in this will be illustrated with a brief example.

Consider the signal of Eq. (1) which is shown in Fig. 1. Figure 18 shows the magnitude of the Fourier transform of this signal; the -3 dB point is at 83 Hz. However, from Fig. 1 one can see that sampling the signal at 1 kHz (more than ten times the 3 dB bandwidth) can completely miss the peak that occurs early in the signal. Figure 2 shows that passing the signal through a system with a 500 Hz bandwidth removes the peak. Thus, if one wishes to record the first peak, the common rules of thumb

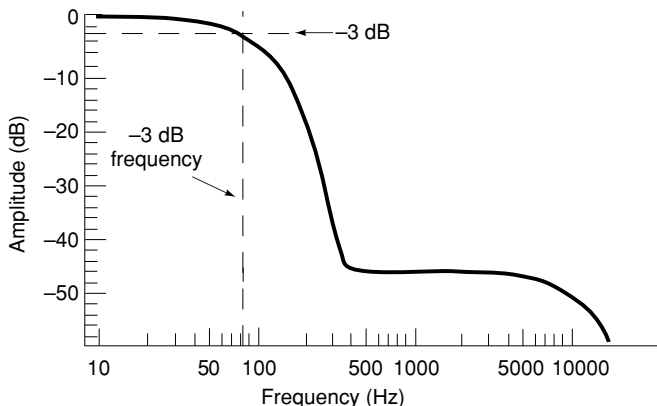


Figure 18. The magnitude of the Fourier transform for the signal in Fig. 1. The shelf at -46 dB is from the narrow pulse at -10 ms.

completely fail. However, if the first peak (and any other narrow peaks) were noise, then it would be desirable to use a lower bandwidth and sampling rate to reduce the cost of the recording system.

Use of the Convolution Integral. The best way to determine the bandwidth requirement of a recorder is through simulation. One obtains sample input signals, simulates passing them through filters of various bandwidths, examines the result at each bandwidth and determines the smallest bandwidth that produces acceptable distortion. Simulations are done by evaluating the convolution integral, Eq. (12). To do this one must have sample input signals, $x(t)$, and estimates of an impulse response, $h_B(t)$, parameterized by the bandwidth. The exact form of $h_B(t)$ is seldom critical. A particularly convenient form to use in simulations is

$$h_B(t) = (t/\tau^2) \exp(-t/\tau), \text{ with } \tau = 0.102/B \quad (29)$$

where B is the -3 dB bandwidth of the filter. With this impulse response the convolution can be rapidly carried out using an IIR filter (1,2,3). These filters are available in many software packages. Another useful impulse response model is the Gaussian,

$$g_B(t) = (1/\sqrt{2\pi}\sigma) \exp(-t^2/2\sigma^2), \text{ with } \sigma = 0.133/B \quad (30)$$

Which of these one uses makes little difference. Of course, if one is interested in the adequacy of a particular recorder for which the impulse response has been measured, the known impulse response could be used.

There are situations in which estimates can be made with very simple calculations. Often the requirement is that pulses with widths exceeding a certain value have their peak values recorded to a certain tolerance. Assuming both the pulse to be recorded and the impulse response of the recorder to be Gaussian, one can use the relation

$$e = \frac{4.9}{(BW)^2} \% \quad (31)$$

where B is the -3 dB bandwidth of the recorder, W is the width (full width at half maximum), and e is the error in the peak amplitude of the pulse. This relation is valid for $e \leq 20\%$. For the example used previously of Eq. 1 and Fig. 1, the width of the first peak is $W = .24$; to obtain an error in the peak of less than 4.9% requires that $B \geq 1/W = 4.2$ kHz.

Use of Antialiasing Filters

An antialiasing filter is a low-pass filter applied to a signal before it reaches the recorder. The cut-off frequency of the filter is at (or slightly higher than) half the sampling rate (the Nyquist frequency) to be used in recording. When a recorder's bandwidth greatly exceeds half of its sampling rate, antialiasing filters should be employed. If not, both noise and signal components at higher than the Nyquist frequency will be aliased down to lower frequencies where they can't be filtered out. This problem is greatest when recorders with variable sampling rate are used at much lower than their maximum rate, because the bandwidth of the recorder will typically be suitable for the highest sampling rate.

Table 3. Extra Bits of Precision Attainable by Oversampling

Oversampling Ratio	Extra Bits of Precision
4	1
16	2
64	3
256	4
1024	5
4096	6
16384	7
65536	8

If the noise in the signal source is white, the antialiasing filter will reduce the noise by a factor of $10 \log(B_f/B_r)$ dB, where B_f and B_r are the bandwidths of the filter and the recorder. The bandwidths in the preceding are the *equivalent noise bandwidths*, rather than the -3 dB bandwidths, but the difference in these is less than 11% for filters of second order or higher. Note that the antialiasing filter does not reduce noise that is produced within the recorder (often a major source) only the external noise.

Noise Reduction through Oversampling

Noise can be reduced and accuracy and dynamic range improved by sampling the signal at a higher sampling rate than required followed by digital filtering of the recorded signal. This process, called *oversampling*, reduces the effect of noise produced within the recorder. If B is the bandwidth required to record the signal, then recording it at a sampling rate of $nB/2$ and filtering the recorded signal to a bandwidth of B reduces the noise level by $10 \log(n)$ dB for the case of white noise. The factor, n , is called the oversampling ratio. To gain a precision of k bits requires an oversampling ratio of $n = 4^k$. The value of n required for a given value of k is shown in Table 3.

Buying a recorder with much higher sampling rate than required in order to apply this technique is usually not cost effective. A 256 MSa/s 8 bit recorder can emulate the accuracy of a 1 MSa/s 12 bit recorder with this technique. The 1 MSa/s 12 bit recorder would probably be cheaper to buy and would require much less (a factor of 170) storage capacity. The oversampling approach is, however, cost effective when the higher sampling rate recorder is already available.

The success of oversampling to reduce noise requires that the noise not be concentrated at frequencies lower than B , the bandwidth of the digital filter. This low frequency concentration of the noise can occur when the primary error source is quantization noise (when the number of effective bits is approximately equal to the number of bits) and the signal is nearly constant. In this situation performance can be significantly improved by adding noise or dither to the signal.

Increasing Dynamic Range with Multiple Recorders

In many applications the average signal level varies with time, and the precision required during a particular time interval is a fixed percentage of the average signal level in the interval. In some cases the dynamic range of expected signal levels exceeds that which can be accurately recorded with a single recorder. In these situations the dynamic range can be increased by using multiple recorders as in Fig. 19. The signal is sent to several

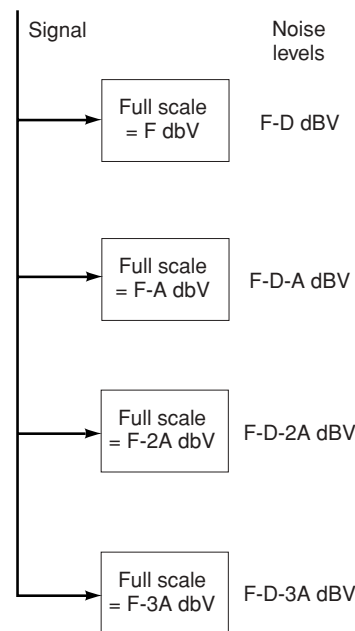


Figure 19. Four recorders used to record 1 signal. Each is A dB more sensitive than the one above it. The signals recorded by the four units are assembled together as described in the text to construct one record.

recorders, and each recorder is set for its full scale voltage to be reduced by A dB with respect to the previous recorder. The data from the three recorders is assembled into a single record after recording is completed by using, for each data point, the largest value from the various recorders that isn't off scale. One must make gain and offset corrections for each recorder to avoid discontinuities in the data.

The optimum value for A and the required number, N , of recorders can be calculated in a straightforward manner. Let $V_{\max}$ be the largest signal to be recorded, D be the dynamic range of the recorder (Eq. 8) in dB and S be the required signal to noise ratio for the signal recording in dB. To use the recorders for this application at all we must have $D > S$. If $V_{\min}$ is the smallest signal level that must be recorded with signal to noise ratio, S , we define the signal dynamic range, $R = 20 \log(V_{\max}/V_{\min})$. The number of recorders required is then $N = R/S$, and the ratio in dB between the ranges of successive recorders is $D - S$.

FUTURE DIRECTIONS

The art and technology of recording have changed drastically in the last decade or so as digital recording has virtually supplanted analog recording. There are several factors that have driven this. One is the rapid decrease in cost and increase in speed and capacity of digital recording media. Another is the advancement of analog to digital converter technology. A third factor is the use of computers within the recorder to control the generation of calibration signals and analyze the recorded results. Future progress is expected to occur along the same lines and is probably more controlled by demand than by technology.

One would expect to see an increase in the bandwidth and sampling rate of ADCs. As the bandwidth moves from its

present limit of 1 to 2 GHz to the 5 to 10 GHz range, the more limited technologies of traveling wave CRTs and streak cameras will probably disappear. One could expect over the next decade to see ADC based recorders reach the same 20 to 30 GHz bandwidth range that sampling oscilloscopes now have.

Increased precision is another area for which there is steady progress. ADCs can be roughly grouped into the 8, 12, 16, and 20 bit classes. Each class has an upper sampling rate limit that can be economically accomplished. These upper frequency limits steadily increase. The increased sampling frequency at a fixed precision of ADCs translates directly into the same improvements for recorders. However, recorders improve at an even faster rate. As the cost, size, and power consumption of ADCs go down it becomes practical to interleave more of them into a single recorder channel—increasing the sampling rate more than that of the ADCs.

An area ripe for improvement at lower sampling rates (10 MSa/s and less) is increased capacity through real time data compression. This involves analyzing and compressing the data between the time it is digitized and the time it is written to disk, so that the disk can hold much longer records. Very specialized computing hardware and algorithms are already in use to accomplish this for audio and video recording. This could be extended to more general situations.

Many users would like to be able to predict a recorder's performance in hypothetical applications based on various recorder specifications. Some work has been done in this area (5) but this is far from straightforward. More research into modeling recorders with sufficient precision to predict performance in a wide range of practical applications is desirable.

A related situation is the current lack of precisely known test signals other than sine waves. One can't test a recorder's performance in a hypothetical situation by generating the hypothetical test signal, recording it, and examining the errors, because the art of accurately generating arbitrary test signals is lagging behind the art of recording them.

BIBLIOGRAPHY

1. A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975.
2. S. S. Stearns and D. R. Hush (eds.), *Signal Processing Algorithms*, Englewood Cliffs, NJ: Prentice-Hall, 1990.
3. S. S. Stearns and R. A. David (eds.), *Digital Signal Analysis*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1988.
4. T. E. McEwan and J. D. Kilkenny, A 32 gigasample-per-second single-shot transient digitizer, *R&D Mag.*, **69**: 108–116, 1993.
5. *IEEE Standard for Digitizing Waveform Recorders (IEEE Std 1057-1994)*, New York: Institute Electrical Electronics Engineers, 1994.
6. *IEEE Instrumentation and Measurement Society Technical Committee on Waveform Measurement and Analysis*. Available URL: <http://grouper.ieee.org/groups/1057/>
7. *IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters (IEEE Std 1241-2000)*, New York: IEEE.
8. J. J. Blair, A method for characterizing waveform recorder errors using the power spectral distribution, *IEEE Trans. Instrum. Meas.*, **41**: 604–610, 1992.
9. *The new IEEE Standard Dictionary of Electrical and Electronics Terms*, 5th ed., New York: Institute Electrical Electronics Engineers, 1992.
10. F. H. Irons and D. M. Hummels, The modulo time plot—a useful data acquisition diagnostic tool. *IEEE Trans. Instrum. Meas.*, **45**: 734–738, 1996.
11. F. Rumsey and T. McCormick (eds.), *Sound and Recording: an Introduction*, Oxford: Focal Press, 1992.
12. J. J. Blair, Error estimates for frequency responses calculated from time domain measurements, *IEEE Trans. Instrum. Meas.*, **47**: 345–353, 1998.
13. R. W. Landee et al. (eds.), *Electronics Designers' Handbook*, 2nd ed., New York: McGraw-Hill, 1977.
14. W. H. Press et al. (eds.), *Numerical Recipes in C*, 2nd ed., Cambridge, UK: Cambridge Univ. Press, 1992.
15. J. C. Whitaker (ed.), *The Electronics Handbook*, Boca Raton, FL: CRC Press, 1996.

JEROME J. BLAIR
Bechtel Nevada

TESTING METHODS

EDDY CURRENT TESTING

Eddy current testing is a nondestructive evaluation method that is based on the principle of electromagnetic induction. While an alternating current passes through a coil, an alternating magnetic field is produced that is oriented perpendicularly to the direction of the current and parallel to the axis of the coil. If a conductive object is located in proximity to the coil, circular eddy currents will be induced within its surface layer normal to the magnetic field. Eddy currents in turn generate a secondary magnetic field that is in opposition to the primary coil field. The interaction between the two fields causes a partial decreasing of the primary field, hence a change in coil impedance or coil voltage. Therefore, a coil and a conductive object close to each other couple into a unified system through the interaction of alternating current in the coil and the induced eddy current in the body. A schematic representation of the system is shown in Fig. 1. Essentially, the eddy current testing means to measure the change of coil impedance (1). The magnitude of the coil impedance is related to either the coil construction or the coupling strength between the coil and tested object. The coupling strength is affected by a number of factors, such as electrical conductivity σ , magnetic permeability, μ of the object (target) material, size, shape, and the distance (or clearance) x between the coil and object. Coil construction parameters include outer diameter D , inner diameter d , thickness b , and the number of turns w , and are designed according to particular testing requirements. They are all the factors that affect the magnitude of coil impedance, but they are always kept constant after the coil, the eddy current probe, is built up.

When the object is machined of homogeneous metallic material, and the conductivity σ , permeability μ , and its shape and dimensions are all fixed, then changes in magnitude of the coil impedance vary with changes in the distance x . This phenomenon is called *lift-off effect* (2). Based on lift-off effect, an eddy current transducer provides displacement measurement; hence, we can measure vibration, motion trace, metal foil and sheet thickness, and thickness of cladding material, either for nonmetal plating on metal material or for a nonmagnetic layer on magnetic material.

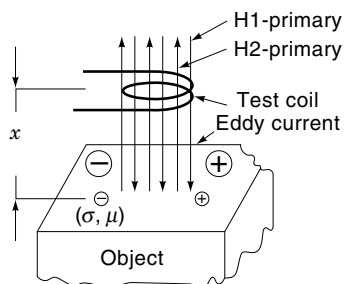


Figure 1. Eddy current principle. I_1 is the exciting ac current; x the distance between coil and tested object; σ , μ are the conductivity and magnetic permeability of the object material, respectively.

If the distance x is kept stationary, the magnitude and changes in coil impedance indicate the combined influences of conductivity σ and permeability μ of the object material, upon which flaw detections are possible. When the permeability μ is also fixed, the coil impedance becomes a function of the conductivity σ . Thus one could determine the impurity content of pure metal, heat treatment condition of an alloy, concentration of dielectric medium, etc. But otherwise, with conductivity σ fixed, impedance will vary with permeabilities, and one could inspect grain-size metallic materials, thermal related strain, hardness, and so on. Therefore eddy current testing is a multivariable detecting technology and has extensive usage. Some major applications will be stated later.

Eddy current was discovered in 1824 when Gambery (3) noted the eddy current damping phenomenon when the oscillations of a suspended bar magnet rapidly stopped whenever a copper plate was held under it. Subsequently, many scientists dedicated themselves to the study of eddy current theory and its practical use; however, progress was very slow. Förestier first investigated the influence of radial cracks in a metal bar on coil impedance in 1954 (10). Dodd and Deeds (11) in 1968 and Libby (4) in 1971 successfully put forward a theory for analytically calculating the induced eddy current within the cross section of a metal bar in an ac magnetic field, eddy current technology has developed rapidly. Now, numerous versions of eddy current test equipment as nondestructive measurement tools have been successfully developed and are commercially available (2).

Eddy current testing has many advantages, such as the following:

1. A probe coil need not contact the tested object (specimen).
2. It has high sensitivity for measuring the surface or sub-surface of conductive materials.
3. It has a fast response—can be used for either static or high-speed dynamic testing.
4. It is unnecessary to have some actuating medium between probe and specimen. Neither is there a problem even if dust, oil, or any other nonmagnetic and primarily non-conductive medium gets between them. Therefore, eddy current testing can be done in unsatisfactory conditions.
5. A variety of testing circuits, such as bridge circuits, resonant circuits, feedback circuits, and phase discriminators are available to generate corresponding outputs of voltage, current, phase, or frequency to reflect coil impedance and its changes.

Just as everything has strengths and weaknesses, however, the eddy current testing method does have inherent limitations (12), as follows:

1. It is applicable only for testing of conductive materials.
2. It detects flaws mainly for surfaces or surface layers, but cannot determine the shape and types of flaws.
3. It is difficult to inspect specimens with complex shape.

4. Eddy current testing can be used to deal with multiple variables; however, this very advantage can bring about interference signals in the results, so that special signal processing is often necessary.

PRINCIPLES OF EDDY CURRENT TESTING

The impedance of an eddy current coil can be analyzed under three distinct conditions (5).

Idling Impedance

This is the coil impedance with tested object very far from the coil, thus not influencing its impedance. Idling impedance Z_0 is

$$Z_0 = R_1 + j\omega L_1 \quad (1)$$

where R_1 , L_1 are the resistance and inductance of the coil, respectively (Fig. 2), and ω is the angular frequency ($\omega = 2\pi f$).

First-Stage Impedance

The impedance of a coil coupled with an object of known physical properties is defined as *first-stage impedance*. It indicates the characteristics of a coil impedance that varies with the distance between coil and object, such as lift-off effect. As mentioned previously, when a metal specimen is placed adjacent to a coil carrying an alternating current, secondary or eddy current will be induced within the surface layer of the specimen. Figure 2 shows the equivalent circuit. According to Kirchhoff's voltage law, a set of equations can be written:

$$(R_1 + j\omega L_1)I_1 - j\omega MI_2 = U \quad (2)$$

$$-j\omega MI_1 + (R_2 + j\omega L_2)I_2 = 0 \quad (3)$$

where R_2 , L_2 are resistance and inductance of eddy current loop in tested specimen. M is the mutual inductance between the coil and the tested specimen. From Eqs. (2) and (3), the first-stage impedance Z of the coil can be derived:

$$Z = \left[R_1 + \frac{\omega^2 M^2 R_2}{R_2^2 + (\omega L_2)^2} \right] + j\omega \left[L_1 - \frac{\omega^2 M^2 L_2}{R_2^2 + (\omega L_2)^2} \right] \quad (4)$$

The equivalent resistance R of the coil is a function of the mutual inductance M :

$$R = R_1 + \frac{\omega^2 M^2 R_2}{R_2^2 + (\omega L_2)^2} \quad (5)$$

Obviously, R increases with the decreasing distance between coil and object. Note that this variation is independent of

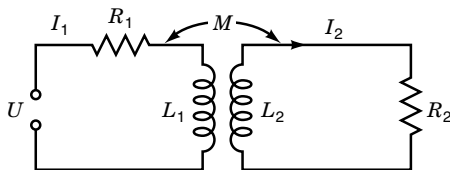


Figure 2. Equivalent circuit of the coil coupled with a conducting object.

material characteristics (magnetic or nonmagnetic). From the second term of Eq. (4), the equivalent inductance L of the coil is

$$L = L_1 - \frac{\omega^2 M^2 L_2}{R_2^2 + (\omega L_2)^2} \quad (6)$$

L is influenced by two physical effects. The first term L_1 is related to magnetostatic effects, so that it is dependent on whether the material is magnetic or nonmagnetic. The second term $\omega^2 M^2 L_2 / [R_2^2 + (\omega L_2)^2]$ is generally considered as a reflected inductance caused by eddy current effect. The result of these two effects on equivalent inductance L is opposite. When the distance between coil and specimen decreases, coil inductance L increases as a consequence of magnetostatic effect, but it decreases, because of eddy current effect. Based on the analysis above, we can relate the variation in distance between coil and specimen to the variation of impedance.

Secondary Stage Impedance

Here we define the impedance of a coil adjacent to an object of unknown physical properties as *secondary stage impedance*. It demonstrates characteristics of coil impedances as related to tested objects that have different physical properties. In non-destructive testing, it is necessary to gain information about conductivity, permeability, and various flaws, so the concept of secondary stage impedance was born at the right moment. From Eq. (5), the resistance increment ΔR of equivalent resistance of a coil is

$$\Delta R = \frac{\omega^2 M^2 R_2}{R_2^2 + (\omega L_2)^2} \quad (7)$$

And from Eqs.(4) and (6), the reactance increment $\omega \Delta L$ of equivalent reactance of the coil is

$$\omega \Delta L = \omega \frac{\omega^2 M^2 L_2}{R_2^2 + (\omega L_2)^2} \quad (8)$$

When we divide Eq. (7) by Eq. (8), we obtain

$$\frac{\Delta R}{\omega \Delta L} = \frac{R_2}{\omega L_2} \quad (9)$$

The resistance increment ΔR and inductance increment ΔL of the eddy current coil generally depend upon metal conductivity. When the mutual inductance M is selected as a fixed value and ωL_2 is assumed as a constant, Eq. (9) tells us that there are different ratios of $\Delta R / \omega \Delta L$ for different metals. This is the theoretical basis of measuring metal conductivities. If we take the increments ΔR and $\omega \Delta L$ as horizontal and vertical axes, respectively, an impedance information graph (showing the relationships of $\omega \Delta L$ and ΔR in impedance Z) can be drawn by plotting $\omega \Delta L$ against ΔR (5). The slope of the plotted straight line is $\omega L_2 / R_2$. Figure 3 shows that different conductivities of tested objects have different lines. Here we see that the particular exciting angular frequency ω is very critical. Too high a frequency is unsuitable for measuring metallic conductivity. Using Eqs. (7) and (8), we can eliminate R_2 ; then we have

$$(\Delta R)^2 + \left(\omega \Delta L - \frac{M^2 \omega}{2L_2} \right)^2 = \left(\frac{M^2 \omega}{2L_2} \right)^2 \quad (10)$$

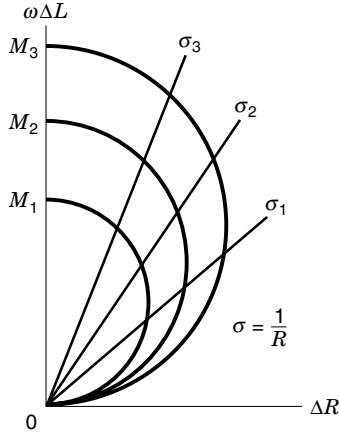


Figure 3. Relationships between metallic conductivity and information impedance.

Equation (10) shows that when the angular frequency ω and mutual inductance M are fixed at certain values, the impedance information graph follows semicircle law. The center of the semicircle sits in the vertical axis of $\omega\Delta L$, its radius is $M_2\omega/2L_2$. Different semicircles can be plotted for different M . Each semicircle passes through the origin of the coordinates and is tangential with the horizontal axis (ΔR axis). For a particular metal conductivity and mutual inductance, the intersecting point of the above straight line and the circle shows the value of the coil impedance information precisely.

The impedance diagram is the basis of eddy current testing. The relationship between coil impedance and flaws, cracks, lift-off, conductivity, permeability, and frequency, as well as the degree of filling, can be clearly known from the impedance diagram (2). Now, we let R be the horizontal axis and ωL the vertical axis, resulting in a complex impedance plane in which the terminal locus of impedance Z becomes the impedance diagram (6). It is the advantage of the complex plane diagram that the locus is rather clear and readily perceived, as shown in Fig. 4(a). We find that the shapes of loci are similar to each

other, but their magnitudes and positions are much different, being affected by different radii of coils and exciting frequencies. Therefore this kind of impedance diagram is difficult to study and utilize. In order to overcome this limitation, a normalized impedance diagram is usually used. As shown in Fig. 4(a), this normalization is achieved first by moving the vertical axis right to the idling coil resistance R_1 , eliminating the left part of the information impedance with the tested object free. Then the two coordinate axes are divided by the idling coil reactance ωL_1 . Through this modification, the normalized impedance diagrams are the same no matter what the exciting frequency and radii of coils are changed to (5), as shown in Fig. 4(b). The fractional resistance $\Delta R/\omega L_1$ and reactance $\omega L/\omega L_1$ are all dimensionless and all less than 1. Therefore normalized diagrams have identical forms and are extensively comparable.

In eddy current testing, there are many variables that cause changes in coil impedance, and so it is very complicated to analyze their influence. For simplification, Förestor proposed a conception of *effective permeability*. He assumed an ideal model in which a cylindrical tested object is placed in an infinitely long solenoid coil that carries ac current. A constant field exists at any cross section of the cylinder, but magnetic permeability changes in the section along its radial direction. However, the related magnetic flux equals that in the real cylinder. Under these assumptions, he came to a conclusion that the real alternating magnetic intensity and constant permeability could be replaced by a constant magnetic intensity and an alternating permeability. This assumed altering permeability is called *effective permeability*, symbolized as μ_{eff} , and is defined by the equation

$$\mu_{\text{eff}} = \frac{2}{\sqrt{-j}} \cdot \frac{J_1(\sqrt{-j}kr108108)}{J_0(\sqrt{-j}kr108108)} \quad (11)$$

where $K = \sqrt{\omega\mu\sigma}$, r is the radius of the cylinder, $J_0(\sqrt{-j}kr)$ is a zero-order Bessel function, and $J_1(\sqrt{-j}kr)$ is a first-order Bessel function. After introducing the concept of effective permeability, Förestor defined the frequency at which the modulus of the Bessel function argument (kr) equals 1 as the *character frequency* f_g and called it the *limiting frequency* (2):

$$f_g = \frac{1}{2\pi\mu\sigma r^2} \quad (12)$$

Obviously, for a common testing frequency f , the following equation is valid:

$$kr = \sqrt{2\pi f\mu\sigma r^2} = \sqrt{f/f_g} \quad (13)$$

Effective permeability μ_{eff} changes with variable (kr), so μ_{eff} can be calculated as long as the ratio of f/f_g is known. Conventionally, f/f_g is taken as a parameter for the analysis of coil impedance.

In practical eddy current testing, a tested cylinder with diameter d usually cannot fully fill the testing coil, which has an inner diameter D , because a gap between coil and tested object is needed for relative movement. Here we define the fill factor η (1) as

$$\eta = (d/D)^2 \quad (14)$$

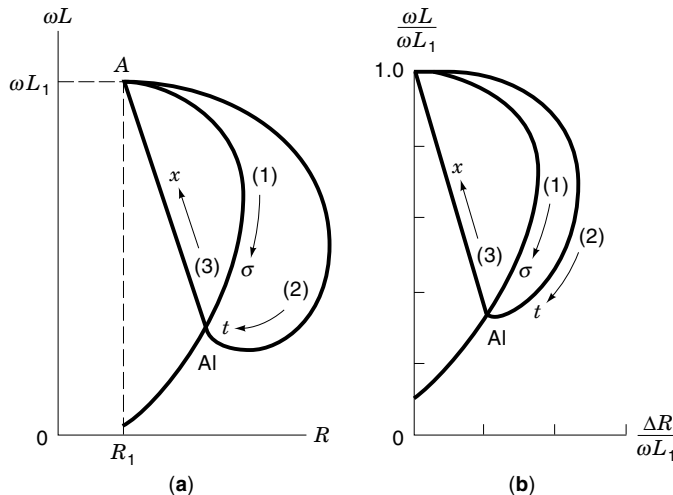


Figure 4. Impedance diagram of aluminum. (a) Impedance changes with variables. (b) Normalized impedance diagram. (1): Conductivity σ . (2): Thickness t . (3): Clearance (lift-off distance) x .

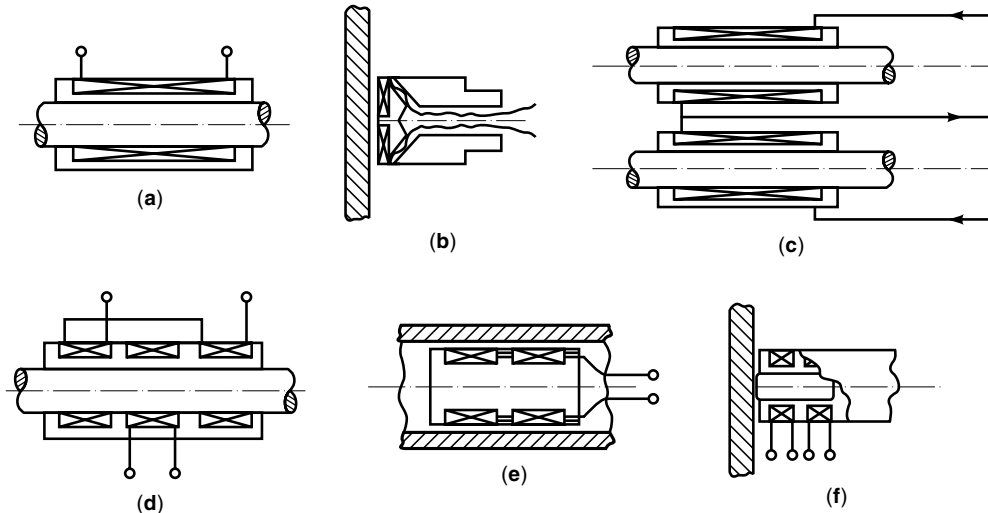


Figure 5. Several designs of eddy current probes. (a) Solenoid-type coil (or single coil) around cylindrical specimen, absolute measurement. (b) Pancake-type coil (or surface-mounted type probe) for testing planar objects. (c) Double coils for comparison measurement. (d) Multi-solenoid coils for differential measurement. (e) Bobbin probe (double or multiple coils) for use in tubular specimen. (f) Pancake-type coils with ferrite cores, the front coil used as a sensing coil, the rear for temperature compensating.

Undoubtedly, the influence of a tested object with different η on coil impedance is different.

From the analysis above, the effect on coil impedance of the opposing magnetic field produced by eddy currents in tested objects is completely determined by the fill factor η and effective magnetic permeability μ_{eff} . On the other hand, we can see from Eqs. (11) and (13) that μ_{eff} is determined by the ratio of f/f_g . Therefore the variation of coil impedance is actually determined by factors η and f/f_g . Thus, for a constant value of η , the distribution of eddy current and magnetic flux density in tested objects is a function of f/f_g . This result leads to a new conclusion: For two different tested objects, if η is kept constant and the corresponding frequency ratios f/f_g are the same, then the geometric distribution of effective permeability, eddy current density, and magnetic flux density are also the same, respectively. This is the so-called *law of similitude* of eddy current testing (5). From Eq. (13) the similitude condition can be written as

$$f_1 \mu_1 \sigma_1 r_1^2 = f_2 \mu_2 \sigma_2 r_2^2 \quad (15)$$

where the subscripts 1 and 2 describe the physical parameters and geometric sizes of tested bodies 1 and 2, correspondingly.

The law of similitude is a theoretical basis for proper simulation experiment. When problems that are solved neither by mathematical method nor settled by direct measurement are encountered in eddy current testing, we can have reasonable results according to the law by simulation experiment. For example, the law of similitude is applied to detect discontinuity flaws of materials, as long as the frequency ratios f/f_g are equal, and discontinuity flaws with geometric similarity (such as flaws having definite depths and widths that are all described as percentages of the cylinder diameter) will cause equal eddy current effect and equal variation of effective magnetic permeability. So, by means of a simulation test model with artificial flaws, the relationship between the variation $\Delta\mu_{\text{eff}}$ of

effective permeability and the depth, width, and location of the flaw can be demonstrated. According to the similitude law, we can take these well-established results as a valid basis for practical evaluation of existing flaws. Thus, in testing of metal wires and small-size tubes, the influences of cracks on the probe coil's parameters may be understood by study of a test model with a magnified cross section and artificial flaws. Fortunately, in the testing of large tubes with eccentricities, nonuniform wall thickness, as well as other flaws, simulation testing makes the evaluation much easier. For particular applications, impedance plane diagrams are usually drawn with selected f/f_g as a parameter. Experimental results indicate that frequency ratios f/f_g within the range of 5 to 150 are of high sensitivity and practical significance.

Eddy Current Probe

The eddy current probe is one of the key components in eddy current testing equipment (2). It consists of a sensing coil, a coil frame, and connecting cables. Performance of the coil directly affects testing accuracy and data reliability. Figure 5 shows several types of eddy current probes for testing metallic tubes, cylinders (wire), and planar objects.

APPLICATIONS OF EDDY CURRENT TESTING

Applications of eddy current testing are generally classified into three main kinds: nondestructive flaw detection, material examination, and displacement and vibration measurement.

Nondestructive Flaw Detection

Flaw Detection in Metallic Tubes. The main purpose of tube flaw detection is to understand flaw kinds, their geometric shapes, and their locations. However, it is difficult to calculate theoretically the sites and shapes of flaws (13). Simulation

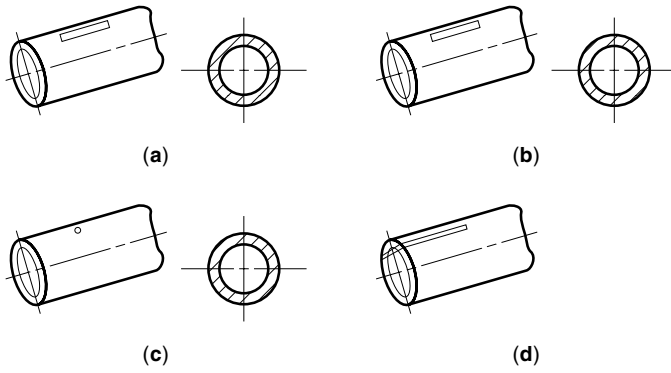


Figure 6. Several types of artificial defects. (a) Rectangle slot. (b) V-type slot. (c) Hole (or blind hole). (d) "Wire cutting" gap.

tests, as pointed out before, are available to acquire knowledge about various flaws (such as shapes, sizes, and locations) within different materials under different ac exciting frequencies. The resulting data, tables, and curves can provide reference criteria for other practical testing (14). For that reason, reference specimens with standard artificial flaws, called *standard scars*, are necessary. Figure 6 shows four types of artificial flaws.

With wide use of eddy current testing, standardization of artificial flaws in many countries is increasingly accurate (8). A number of symbols, shapes, and sizes of standard flaws has been clearly designated today, making the selection, application, and reproduction of artificial flaws much easier. For the testing of tubular materials, single-coil solenoids are often employed. With reference to wall thickness, tubular materials can be put into two categories: thin-walled tubes and thick-walled tubes. The influence of tubular material on coil impedance is determined by the conductivity σ , relative permeability μ_r , outer and inner diameters d_2 and d_1 , and wall thickness w . Of course, flaws within the outer and/or inner surface of a tube and eccentricity are also factors that affect coil impedances. Figure 7(a) shows that impedance varies with outer diameter d_2 (η is altered), whereas the ratio of d_2/d_1 is kept constant and the impedance diagram presents a set of semicircular curves in the case of $\eta = 1, 0.75, 0.5$, respectively. Each of the curves indicates coil impedance changes with conductivity σ , inner diameter d_1 , and wall thickness w ; and with frequency ratio of f/f_g under the condition of outer diameter $d_2 = \text{constant}$ (η is fixed). The crosswise-oriented curves indicate direction of the variation in coil impedance with outer diameter d_2 .

If the inner diameter d_1 of the tubular material is fixed, variation of outer diameter d_2 may cause two types of effects: One is outer diameter effect, where the crosswise curve in Fig. 7(a) shows the changes of impedance. The second is wall thickness effect, which makes f/f_g change greatly while the impedance value reaches to a new place corresponding to the new point of f/f_g . Because these two effects happen at the same time, impedance variation along with the crosswise curve caused by outer diameter effect is not very obvious, but it changes clearly along the semicircular curve symbolized as f/f_g . So the "total effect" makes impedance vary along the curve labeled, "varying d_2 and w " as shown in Fig. 7(b).

The eddy current effect of cracks in thin-wall tubes is the same as that of decreasing w of wall thickness. Therefore the

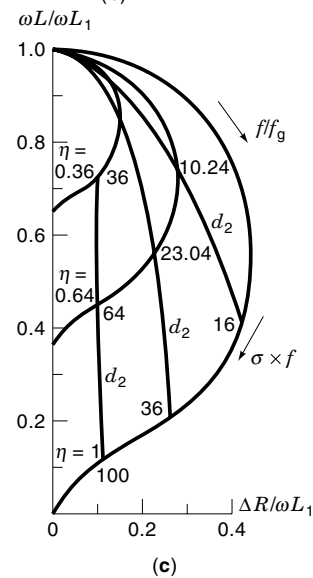
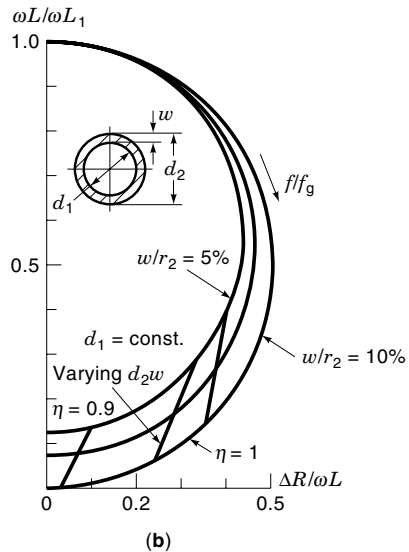
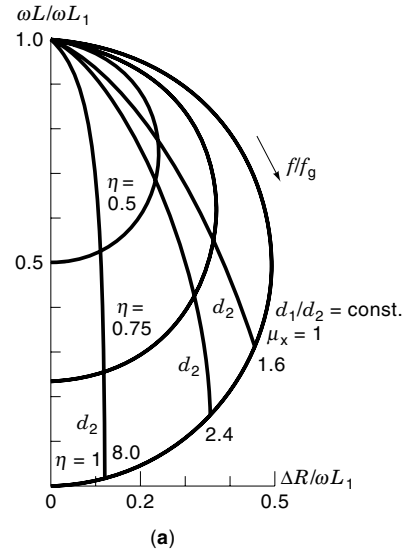


Figure 7. Impedance diagram for nonmagnetic thin wall tubes. (a) Coil impedance changes with d_2 as $d_1/d_2 = \text{const.}$ (b) Coil impedance changes with d_2 as $d_1 = \text{const.}$ (c) Coil impedance changes with d_2 as $d_1/d_2 = 80\%$.

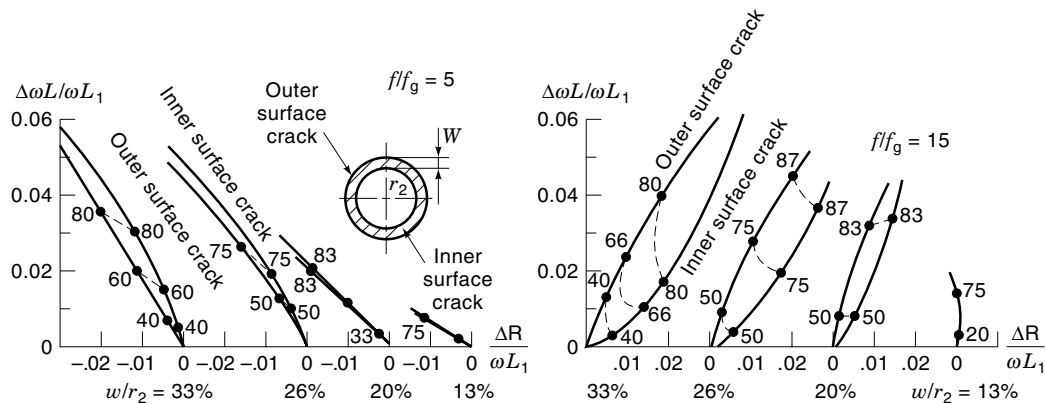


Figure 8. Coil impedance affected by cracks in nonmagnetic metal tube. Numbers on curves represent the depth of cracks as percent of wall thickness W .

effect owing to the existence of cracks in the outer surface layer is identical to that caused by altering outer diameter d_2 while inner diameter d_1 remains fixed—see Fig. 7(b). Similarly, the effect caused by cracks in the inner surface layer is identical to that caused by altering inner diameter d_1 while outer diameter d_2 is kept constant (wall thickness w varies), as shown in Fig. 7(a).

A frequency range corresponding to $f/f_g = 0.2 \sim 2.4$ is usually selected as the exciting frequency because there is greatest sensitivity at the point of $f/f_g = 1$.

Impedance curves of single coil testing of thick-walled tube are located, in the impedance diagram, between curves tested from cylinders and thin-walled tubes. If the ratio of d_1/d_2 of a thick-walled tube holds constant, whereas the outer diameter d_2 changes, we will obtain the impedance diagram shown in Fig. 7(c). Figure 7(c) also shows a set of curves for coil impedance that varies with frequency ratio f/f_g under the condition of $d_1/d_2 = 80\%$ and η (or d_2) as a parameter. As the outer diameter changes, the coil impedance changes chordwise.

For understanding of interior flaws, a great amount of model testing is necessary (5). Figure 8 shows the effects of cracks in nonmagnetic tubes, localized at different sites and depths, on coil impedance at frequency ratios $f/f_g = 5$ and 15. We can see from the diagram that there is a phase shift between impedance curves of cracks in inner and outer walls and the impedances increase with f/f_g and w/r_2 . The effect of cracks within the material on the coil impedance is slightly less than that of cracks having the same depths in the inner or outer surface layer.

In practical applications, single coil testing is generally employed for tubes with diameters smaller than 75 mm, whereas the bobbin-type coil is suitable for tubes with diameters larger than 75 mm. Usually, a single-pancake coil or surface-mounted probe is preferable for tubes with very large diameters.

Flaw Detection in Metallic Rods and Wires. Even though eddy current testing can only detect flaws in the surface or surface layers, it is widely used in the area of surface quality evaluation for some metallic rods and wires. For flaw detection of the batch process rods and wires, a similar way to detect flaws in metal tubes is available. However, eddy current penetration

is smaller in rods and wires than in tubes, and its distribution is also different. In order to raise eddy current sensitivity, a much lower test frequency should be selected than that used for tube testing (5). For these applications, single coil may often be used as the detecting coil.

As mentioned in the last section, the conductivity σ , magnetic permeability μ , dimensions and flaws of tested materials, and testing frequency are the major factors that influence coil impedances. The effect of flaws on coil impedance can be considered as the combined result of conductivity and dimensions. Flaws that are characterized by such qualities as their shape, depth, and location are very difficult to calculate theoretically. Hence flaw detections currently have to be achieved by means of model testing.

Figure 9 shows the effects of flaws on coil impedance and the curves resulting from model tests at frequency ratios of $f/f_g = 5$ and 15 for nonmagnetic cylinders (rods) with flaws of various locations, shapes, depths, and widths. The zero points of all these curves relative to flaw-free objects are located at the point that is determined by the frequency-dependent effective permeability μ_{eff} .

With $f/f_g = 15$, for example, one line segment marked Δd on Fig. 9(b) expresses the “diameter effect” relative to variation of diameter, and the numbers on it indicate percent decrease in diameter. Another line segment marked $\Delta\sigma$ expresses the “conductivity effect,” and the numbers indicate percent increase in conductivity. Other line segments marked 10, 15, 20, and 30 show the regularity of coil impedance variation in the case of a tested cylinder with narrow cracks and width-to-depth ratio of 1:100 and when depths equal 10%, 15%, 20%, and 30% of diameter, respectively. The numbers 3.3, 2, and 1 at the right-hand side express the distance of upper terminals of internal cracks to the object (cylinder) surface as being 3.3%, 2%, and 1% of diameter. The numbers 4:1, 2:1, 1:1 express width-to-depth ratios of cracks. We also can see from the diagram that, for a subsurface crack with a depth of 30% diameter, as its upper terminal gets farther from the surface, the coil impedance will vary along the curve marked with the numbers 1, 2, and 3.3. When the depth of a “V-type” crack varies, the coil impedance will vary along the curve marked with the ratios 4:1, 2:1, and 1:1. Furthermore, with the increasing width-top-depth ratio, the orientation of crack effect becomes that of diameter effect.

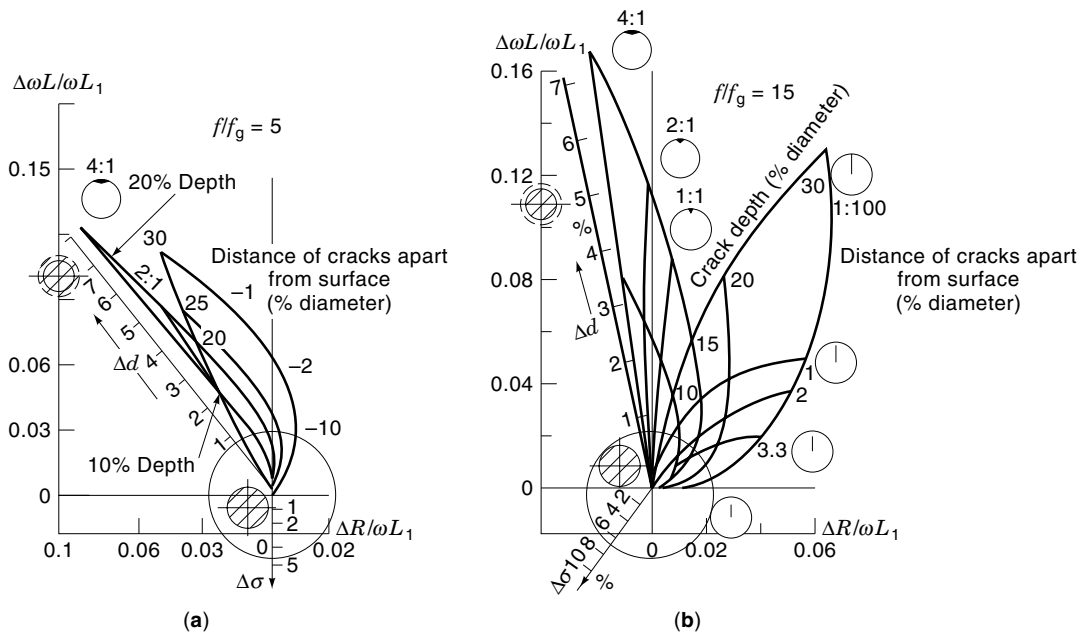


Figure 9. Cracks' effect on solenoid coil impedances. (a) $f/f_g = 5$. (b) $f/f_g = 15$.

According to the analysis above, the danger of cracks can be evaluated. For instance, the larger the angle included between the directions of crack effect and diameter effect, the deeper the crack. A harmful "sharp crack" is often the case. Otherwise, the heavy scratch mark usually has a large width-to-depth ratio, but the included angle between crack and diameter effect is very small or near zero, so it does not cause any danger.

In practical eddy current nondestructive crack detections, a frequency ratio f/f_g ranging from 5 to 150 is valuable. The optimum ratio for searching surface cracks is within 10 to 50. And a ratio of 4 to 20 is the best range for searching for subsurface cracks, whereas a ratio of 5 to 10 is useful for detecting both surface and subsurface cracks (5).

Material Examination

Eddy current testing of materials is achieved by measuring the variations of electrical conductivity and magnetic permeability of tested objects (9). Usually, material quality condition such as chemical constitution and impurity content affect conductivity and permeability.

Determining Chemical Composition and Impurity Content of Nonmagnetic Metals. The relative permeability μ_r for nonmagnetic metallic material is $\mu_r = 1$, approximately that of air, and it is generally taken as constant. Therefore, the problem is simplified because only conductivity need be measured by the eddy current method.

The purity of metallic materials is closely related to their conductivities. If a small amount of impurities is melted into the materials, their conductivities will be decreased rapidly. Figure 10 shows the variation of electrical conductivity of metallic copper with different impurities and different content amounts. We can see from the diagram that even if the impurity contents in copper are 0.1% to 0.2% Fe, Si, or P, and so on, the conductivity decreases dramatically. Furthermore, the

variation of conductivity is roughly proportional to the impurity content. Here unit IACS in diagram is the international annealed copper standard (6). Conductivity of copper according to it is defined as 100% IACS. In the copper industry, the amount of impurity is evaluated by its conductivity.

In brief, once we obtain the definite relationships between conductivities and impurity content, we can easily conclude what impurities and how much of them the material contains.

Determining Chemical Composition of Magnetic Materials.

Relative permeabilities μ_r of magnetic materials are very much larger than that of nonmagnetic ones, usually in the order of 10^2 to 10^4 . Therefore, the magnetic parameter becomes the major factor in eddy current testing for magnetic materials.

Magnetization (B - H) curves of magnetic materials are chemical composite dependent (15). For example, the curve of

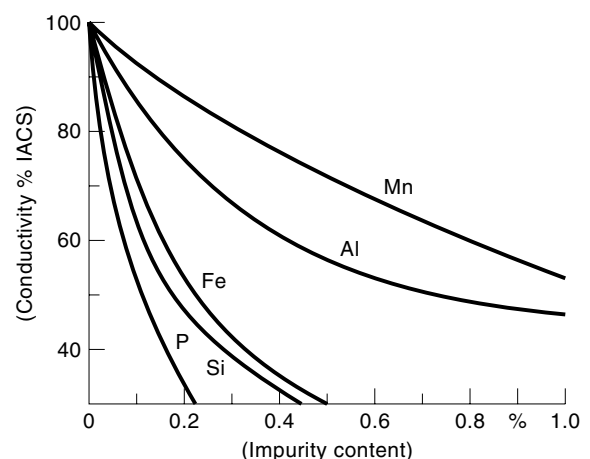


Figure 10. Effect of impurity content on copper conductivity.

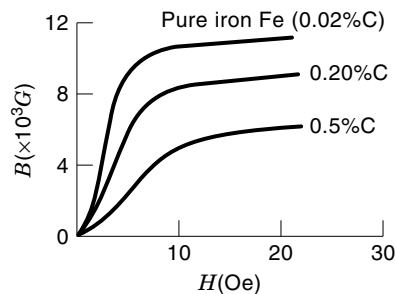


Figure 11. Effect of carbon content in carbon steel on magnetization curves (annealed condition).

carbon steel changes with different carbon content in it, as shown in Fig. 11. Generally, magnetic permeability, residual magnetization, and saturation flux density will all decrease with increasing carbon content. We can use these relationships between magnetic parameters and chemical composites and their contents to evaluate qualities of materials.

Displacement and Vibration Measurement

Displacement Measurement. Generally, the surface-mounted probe, or pancake-type coil, is acceptable for displacement measurement. Parameters of the coil such as diameters, turns, and thickness are required for test range and accuracy. For instance, in order to measure large displacement, a coil with the necessary axial uniform distribution of magnetic field is needed to form a large linear range. If high sensitivity is needed, the variation of eddy current dissipations with the relative movement of coil and testpiece along the coil axis should be large enough, thus a little bit thicker coil is suitable.

Once the conductivity σ , permeability μ , and exciting frequency f have been selected, coil impedance is only the function of distance x between coil and testpiece, as per the lift-off (or "clearance") effect stated earlier. In flaw detecting, however, lift-off effect is an interference factor and needs to be avoided. Curves of lift-off effect are expressed as impedance variation with spacing under conditions of specific materials measured and constant testing frequency, whereas the impedance of a surface-mounted coil is of great significance to lift-off effect. Figure 12 shows an impedance diagram of surface-mounted coils.

Here we should point out that the curves in Fig. 12 were obtained with testing frequency equal to 60 kHz and validated with the test coil only. Coil impedance diagrams will differ when using different coils, different exciting frequencies, and different materials. Thus, these diagrams cannot be for common use but rather for reference. Because the scope of a magnetic field generated by a coil carrying ac current is limited and only its uniform range is usable, linear range of displacement conventionally takes $\frac{1}{3}$ to $\frac{1}{5}$ of the outer diameter of the coil.

Vibration Measurement. Vibration problems are always present in operations of rotational machinery. Serious vibrations will affect normal operations, so that vibration measurement becomes an important aspect of engineering. The eddy current method serves as a noncontact detecting technique for testing various vibrations.

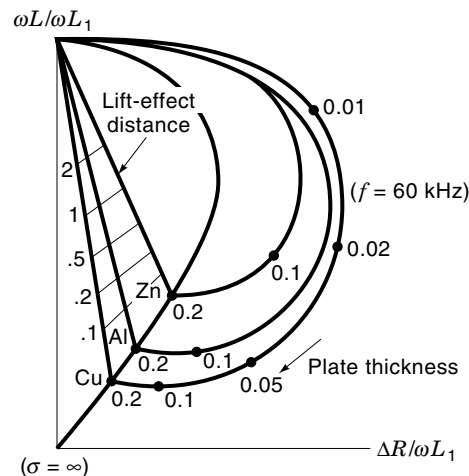


Figure 12. Normalized impedance diagram for surface-mounted coil (or pancake coil).

The testing process is very simple. As long as we place a surface-mounted coil facing the vibrating body within a certain distance from the body, while the vibrating body has reciprocating movements in repeating patterns, the distance between them will be altered periodically, and so will the magnitude of the coil impedance. Testable range of vibration magnitude is generally within several millimeters. The vibration signal can usually be input to an oscilloscope so that its waveshape can be observed directly. Figure 13 shows radial and axial vibration measurements for a turning shaft.

Eddy current methods based on the lift-off effect also have wide use in research activities and industrial processes such as thickness measurement, rotatory angle and rotatory speed measurement, and counting of products.

PROSPECTIVE DEVELOPMENT OF EDDY CURRENT TESTING

In the early 1950s, Förster put forward an "impedance analysis method"—a new way to discriminate various factors that affect eddy current signals. Since then, eddy current testing technology has achieved an essential breakthrough and has come into a new period of practical applications. After the 1970s, the rapid advances in integrating electronics, computer science, and signal processing encouraged the development of

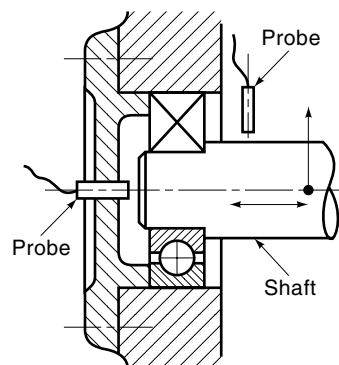


Figure 13. Vibration measurement of rotating shaft.

eddy current testing and pushed forward theoretical study, practical research, and new equipment production (2). Its application is constantly growing.

Multifrequency Eddy Current Technology

The application of multifrequency eddy current testing is relatively new. After thorough analysis of electric magnetic theory, Libby pointed out that a variety of interference signals can be significantly rejected and valuable signals more easily acquired by means of multifrequency eddy current testing methods. Now, multifrequency technology has been applied to tube flaw detecting, graphic display of tube cross sections, thickness measurement of multilayer metal film, and more (16), even though it isn't widely used yet owing to limitations of component performance, analytical methods for particular problems, complicated circuits, and so on. However, with the development of eddy current theory and computer-based signal processing, the multifrequency method will certainly play an important part in eddy current testing because of its attractive advantages.

Remote Field Eddy Current Testing Technology

The *remote field* method is a low-frequency eddy current technology (17) with which magnetic flux is able to penetrate metallic plates or tubular walls. The eddy current probe for tube flaw detecting works with an exciting coil and a relatively small searching coil that is installed apart about twice that of inner diameter from the exciting coil. The search coil can pick up the magnetic flux that comes from the exciting coil penetrating the tube wall and then returning into the tube. Thus flaws within the inner walls of tubes and decreasing thicknesses of tubular walls can be effectively detected. Because remote field technology has high sensitivity to metal tubes (especially to ferromagnetic tubes), it has become a practical way to inspect various flaws in tubes having long lengths. For that reason, remote field technology has been widely applied to tubes for petroleum, natural gas, and municipal coal gas transporting.

Evaluation of Surface Quality of Metallic Materials

As mentioned before, the eddy current effect is very sensitive to electromagnetic conditions in surface layers of metallic materials. Through the study of relationships between eddy current effects and surface layer performances, the evaluations of surface qualities of metallic materials can be fairly well achieved (9). Therefore we expect that the extensive studies of eddy current testing for residual stress, fatigue cracks, and crack extension and initiation will advance at a great pace.

Computer-Aided Eddy Current Testing

In order to raise reliability and automation of eddy current testing, digital and intelligent instruments have been developed. Numerical calculation for complicated theory problems of eddy current testing will be realized with the most advanced computer technology. Computer technology will be widely used for automatic control and data processing of eddy current testing (18). Microprocessor-based intelligent testing apparatuses

that are portable, multifunctional, and easily operated will be emerging.

Signal Image Processing and Discrimination

In eddy current testing, the eddy current information caused by flaws can be extracted by multifrequency technology, after which a variety of images relative to eddy current data can be analyzed by means of computer image processing. Once the images reflecting flaws of tested objects are displayed on screen, the flaws are shown visibly and can be quantitatively analyzed.

Probe Research

The eddy current probe, as pointed out earlier, plays an important role in eddy current testing activities. Its performance mainly comprises linearity, sensitivity, resolution, range, copper resistance, temperature stability, reliability, and probe dimensions. Most of these are connected with probe construction parameters such as coil turns, inner and outer diameters, coil thickness or length, whether air-cored or ferrite-cored, stranded wire-wound coil or metal film-type coil, and the combined use of multiple coils. Consequently, optimum design of probe parameters has become an important project (19). However, there is still a lack of theoretical analysis for probe design. The design is frequently carried out by means of experiment and experience. Optimum design is important, since, for example, a well-built stranded wire-wound coil can reject or reduce temperature drift over a wide temperature range if it is made with the proper diameter of single wire, the correct number of wire strands, and the proper choice of ac current frequency (20).

In short, with the development of science and industrial technology, the theoretical study of eddy current testing and improvement of testing equipment will certainly advance to a new stage. Eddy current testing technology has splendid applications in the fields of aircraft (21), navigation, metallurgy, machinery, electric energy production, chemistry, nuclear power generation, and more.

BIBLIOGRAPHY

1. R. Halmshaw, *Non-destructive Testing*, London: Edward Arnold Publisher, 1987.
2. E. Paul and P. E. Mix, *Introduction to Nondestructive Testing, a Training Guide*, New York: Wiley, 1987.
3. NASA SP-5113, *Nondestructive Testing—A Survey*. Washington, D.C.: National Aeronautics and Space Administration, NASA SP-5113, 1973.
4. H. L. Libby, *Introduction to Electromagnetic Nondestructive Test Methods*, New York: Wiley—Interscience, Wiley, 1971.
5. R. Ji-lin, *Electromagnetic Nondestructive Testing*, Beijing, Aviation Industry Press (China), 1989.
6. B. Hull and V. John, *Non-destructive Testing*, Houndmills, Basingstoke, Hampshire, and London: Macmillan Education Ltd, 1988.
7. R. S. Sharpe, *Research Techniques in Nondestructive Testing*, Volume V, London: Academic Press Inc., 1982.
8. H. Berger, ed., *Nondestructive Testing Standards—A Review*, American Society for Testing and Materials, Philadelphia, 1977.
9. C. O. Roud and R. E. Green, eds., *Nondestructive Methods for Material Property Determination*, New York: Plenum Press, 1984.
10. F. Förster and H. Breinfeld, *Z. Metallkunde*, **45** (4): 1954.
11. C. V. Dodd and W. E. Deeds, *J. Appl. Phys.*, **39** (6): 1968.

12. G. Van Drunen and V. S. Cecco, Recognizing limitations in eddy current testing, *NDT: International*, **17** (1): Feb. 1984.
13. M. D. Halliday and C. J. Beevers, The sizing and location of small cracks in holes using eddy current, *NDT*, **21** (3): June 1988.
14. J. Blitz, D. J. A. Williams, and J. Tilson, Calibration of eddy current test equipment, *NDT International*, **14** (3): June 1981.
15. B. M. Ma and Y. L. Liang, et al., The effect of Fe content on the temperature dependent magnetic properties of Sm (Co, Fe, Cu, Zr) and Sm sintered magnets at 450°C, *IEEE Trans. Magn.* **32** (5): Sept. 1996.
16. J. Blitz and T. S. Peat, The application of multifrequency eddy current testing ferromagnetic metals *NDT International*, **14** (1): Feb. 1981.
17. S. M. Haugland, Fundamental analysis of the remote-field eddy current effect. *IEEE Trans. Magn.* **32** (4): July 1996.
18. T. Stepinski and N. Maszi, Conjugate spectrum filters for eddy current signal processing, *Material Evaluation (USA)* **51** (7): July 1993.
19. A. Powell and T. Meydan, Optimization of magnetic speed sensors, *IEEE Trans. Magn.* **32** (5): Sept. 1996.
20. X. B. Zhuge and B. M. Ling, Analysis of temperature drift rejection of coil impedance, *IEEE Instrumentation and Measurement Technology Conference* **1**: 1996.
21. G. L. Fitzpatrick et al., Magneto-optic/eddy current imaging of aging aircraft: A New NDI Technique, *Material Evaluation (USA)* **5** (12): Dec. 1993.

XIANGBIN ZHUGE
BAOMING LING
Zhejiang University

EDDY CURRENT NONDESTRUCTIVE EVALUATION

Eddy current methods of nondestructive testing (NDT) (1,2) are one of the most commonly used methods for evaluating the integrity of materials in industry. Although there are several different eddy current methods, they all rely on the principles of electromagnetic induction to ascertain the condition of a given test specimen. The basic principle underlying such methods can be illustrated with a simple arrangement shown in Fig. 1.

Consider a coil placed over an electrically conducting, non-ferromagnetic test specimen. If the coil is excited by an alternating-current source, an alternating magnetic field is established. The alternating magnetic field causes currents to be induced in the conducting test specimen in accordance with the Maxwell-Faraday law. The induced currents are called eddy

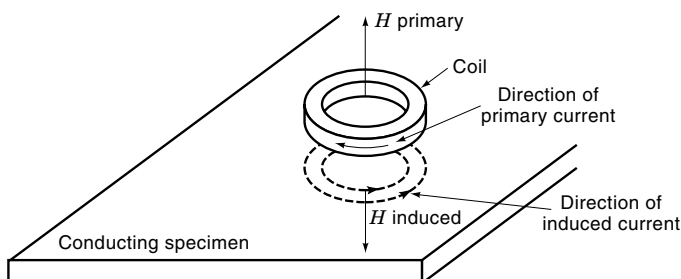


Figure 1. Eddy current probe over a conducting test specimen.

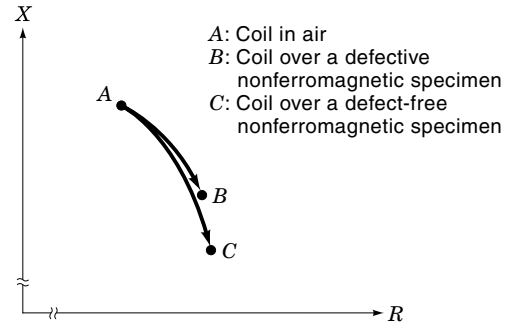


Figure 2. Exaggerated view of the impedance-plane trajectory of a coil over a conducting nonferromagnetic test specimen.

currents since they follow closed circulatory patterns that are similar to eddies found in bodies of water. The alternating eddy current, in turn, establishes a field whose direction is opposite to that of the original or primary field. Consequently, the net flux linkages associated with the coil decreases. Since the inductance of a coil is defined as the number of flux linkages per ampere, the effective inductance of the coil decreases relative to its value if it were to be suspended in air. The presence of eddy currents in the test specimen also results in a resistive power loss. The effect of this power loss manifests in the form of a small increase in the effective resistance of the coil. An exaggerated view of the changes in the terminal characteristics of the coil is shown in Fig. 2, where the variation in resistance and inductance is plotted in the impedance plane. When a flaw or inhomogeneity whose conductivity differs from that of the host specimen is present, the current distribution is altered. Consequently, the impedance of the coil changes relative to its value obtained with an unflawed specimen, as shown in Fig. 2. A system that is capable of monitoring the changes in impedance can, therefore, be used to detect flaws in a specimen that is scanned by a coil. It is not necessary to rely on impedance measurements to detect the presence of flaws. Systems that rely on the measurement of the coil currents and voltages have also been used for detecting flaws. A more detailed discussion relating to variations of the method as well as some simple transducer configurations follows after a brief introduction to the underlying theory.

MATHEMATICAL MODEL

The eddy current nondestructive evaluation phenomena, in common with all electromagnetic phenomena, are governed by Maxwell's equations (3). Three of these equations that are pertinent to eddy current testing are

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (2)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (3)$$

where $\mathbf{E}$ is the electric field intensity (V/m), $\mathbf{H}$ is the magnetic field intensity (A/m), $\mathbf{B}$ is the magnetic flux density (Wb/m²) and $\mathbf{J}$ is the current density (A/m²). The coil excitation frequencies are typically very low (<10 MHz) and consequently the

displacement current term ($\partial \mathbf{D}/\partial t$) in Eq. (2) can be neglected. Equation (2), therefore, reduces to

$$\nabla \times \mathbf{H} = \mathbf{J} \quad (4)$$

Since $\mathbf{B}$ is divergence free, it can be expressed as

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (5)$$

where $\mathbf{A}$ is called the vector magnetic potential. Substituting Eq. (5) in Eq. (1) and using Eq. (4) as well as the constitutive relationships $\mathbf{B} = \mu \mathbf{H}$ and $\mathbf{J} = \sigma \mathbf{E}$, the following result can be derived:

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{A} = \mathbf{J}_s + \sigma \frac{\partial \mathbf{A}}{\partial t} \quad (6)$$

where $\mathbf{J}_s$ is the applied or impressed current density. Equation (6) is a parabolic partial differential equation that governs the physical process underlying eddy current phenomena and is applicable for situations involving both sinusoidal as well as nonsinusoidal excitation. If the coil is excited by a sinusoidal source, then, the governing equation can be simplified, assuming steady state conditions to the elliptic equation,

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{A} \right) = \mathbf{J}_s - j\omega\sigma \mathbf{A} \quad (7)$$

where $\omega = 2\pi f$ is the angular excitation frequency (rad/s) and $j = \sqrt{-1}$. The partial differential equation represented by Eqs. (6) or (7) can be solved with appropriate boundary conditions to obtain the vector magnetic potential A in the region of interest.

Other measurement quantities of interest, such as the coil impedance Z , induced coil voltage V_i , and eddy current density $\mathbf{J}_e$ can be derived using

$$Z = \left(\frac{j\omega}{I_s} \right) \int_c \mathbf{A} \cdot d\mathbf{l} \quad (8)$$

$$V_i = (j\omega) \int_c \mathbf{A} \cdot d\mathbf{l} \quad (9)$$

$$\mathbf{J}_e = -j\omega\sigma \mathbf{A} \quad (10)$$

The solution of the underlying partial differential equation is fraught with several challenges due to the awkward boundary conditions. Analytical as well as numerical approaches to addressing the problem are discussed in a later section.

The electromagnetic field decays exponentially as a function of depth within the test specimen. Eddy current methods, as a general rule, are not very effective for detecting defects that lie deep in the material. The rate of decay is a function of the excitation frequency f , as well as the conductivity and permeability of the test specimen. A useful index that is often employed in industry is called the skin depth. The skin depth δ is defined as the depth at which the eddy current decays to $1/e$ or 36.8% of the value at the surface. If an infinite sheet of current is induced at the surface of a conducting half-plane, the current decays to $1/e$ of the value at the surface, at a depth δ

given by

$$\delta = \frac{1}{(\pi f \mu \sigma)^{1/2}} \quad (11)$$

Although the value of the skin depth as given by Eq. (11) is applicable for the theoretical case of a half-plane specimen, it is nevertheless often used as a guide for choosing the excitation frequency for a given test specimen. Lower excitation frequencies have to be employed for detecting flaws that are located deep in the specimen. Likewise, the skin depth “rule” dictates that, other factors being equal, lower excitation frequencies have to be used when the specimen conductivity is high.

Eddy Current Probes

A variety of eddy current probes are in use today, and a full discussion relating to their type and characteristics is beyond the scope of this short article. The eddy current probe described in the introductory section is an absolute probe. In practice the change in the impedance of the coil due to a defect is very small relative to its quiescent value. The challenges associated with the measurement of these small perturbations in the presence of variations contributed by changes in the environment (temperature, noise, etc.) can be formidable. Differential probes (1,4) are often used to overcome some of these problems. Figure 3 shows a simple differential eddy current probe that is used for inspecting tubes. The probe consists of two identical coils that are spaced apart by a small distance. Consider a situation in which the probe is moved past a small defect such as a crack. When the probe is positioned in a defect-free region of the tube, the impedances of the two coils are identical. Consequently the differential impedance is zero. As the leading

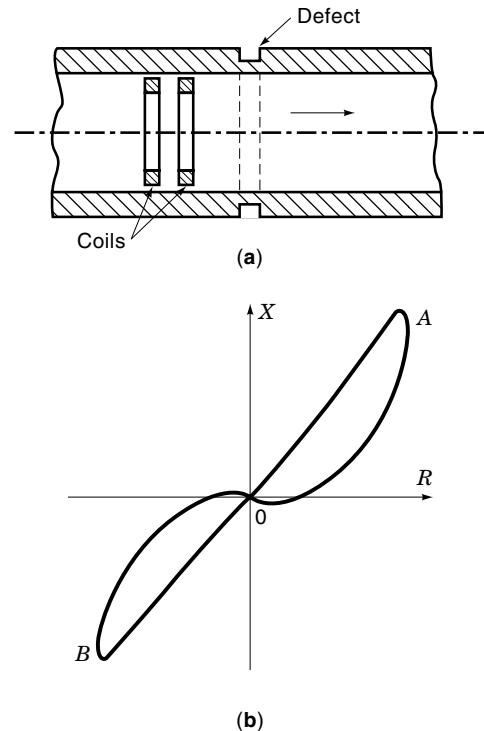


Figure 3. (a) Differential eddy current probe used for inspecting tubes; (b) Impedance-plane trajectory traced for an OD slot on the tube.

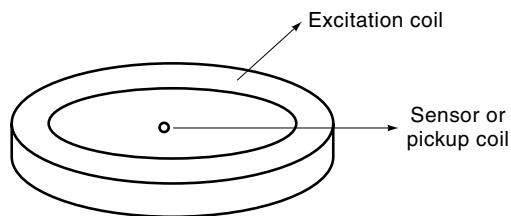


Figure 4. High-resolution eddy current probe.

coil moves past the defect, the change in its impedance causes the differential impedance to trace one of the lobes (lobe A) shown in Fig. 3. When the defect lies between the two coils, the impedances of the both coils are equal and hence the differential impedance is zero. When the trailing coil moves past the defect the second lobe (B) in the impedance plane trajectory is traced. After the trailing coil has moved past the defect, the differential impedance reduces to zero. The shape of the impedance plane trajectory is a function of the shape of the defect, excitation frequency, and probe design. Defect characterization usually involves analysis of the impedance plane trajectory. The differential nature of the probe makes it relatively insensitive to effects of temperature and variations in the spacing between the specimen and the probe (often called the lift-off). The probe is also, unfortunately, less sensitive to long axial cracks. Several variations of the basic design have been proposed in recent years to overcome this deficiency. Commercial absolute probes often use a differential arrangement in which the reference coil is located away from the specimen. The difference in the impedances of the two coils is measured.

The electromagnetic "footprint" of most eddy current probes is large. This contributes to poor resolution and the need for using deconvolution procedures. One of the most commonly used techniques is to employ a ferrite core with a small tip to improve resolution. The ferrite core concentrates the field, thereby improving the resolution. Other methods include the use of copper shielding (4) as well as active compensation methods (through the use of auxiliary coils) (5). Figure 4 shows an alternative approach involving the use of a large excitation coil combined with a very small sensor or pickup coil. The small size of the sensor coil improves the resolution. Other probe designs include those that rely on the establishment of a uniform field and the detection of small perturbations in the field. Nuclear utilities also employ probes that rotate for inspecting tubes (4).

A number of new field sensors have emerged in recent years. High-sensitivity Hall sensors and magnetodiodes as well as self-contained sensors that include amplifiers and temperature compensation circuits are increasingly being used as sensors instead of coils. Another exciting development is related to the availability of relatively low-cost superconducting quantum interference devices (SQUIDS) (6), which are capable of measuring extremely low fields. Several researchers are investigating the feasibility of using such sensors for detecting extremely small and deeply embedded defects.

EDDY CURRENT TECHNIQUES

The test setup used in the introductory section to explain the concept of eddy current testing involved the use of a single coil

excited by an alternating current source. Several variations of this basic scheme are used in industry.

A simple extension of the concept is to employ multiple excitation frequencies to exploit the fact that the depth of penetration is a function of the excitation frequency. Defects and artifacts that are close to the surface can be detected with greater levels of sensitivity using high-frequency excitation signals. Low excitation frequencies can be used to detect defects that lie in the recesses of the material. The responses obtained at the two excitation frequencies can be combined appropriately to suppress contributions selectively from extraneous artifacts that are present in the vicinity of the defect and prevent direct observation of the defect signal. The excitation signals can be applied to the eddy coil either sequentially in a time division or simultaneously in a frequency-division multiplexed manner.

As an example, multifrequency methods are used extensively in nuclear utilities for extracting defect signals that are masked by unwanted signals from artifacts such as support plates. Consider a case where a small defect exists in a tube, which is anchored and held in place by a support plate as shown in Fig. 5. If the tube is inspected using a differential eddy current probe, the signals due to the defect and the support plate would overlap, particularly if the flaw were to be close to the plate. This is evident from Fig. 6(a), which shows a composite impedance-plane trajectory. The defect signal is clearly corrupted by the larger support-plate signal. Multifrequency methods involve measurement of the eddy current signal at two different excitation frequencies. The first signal is captured using a high excitation frequency so as to obtain a response that is relatively more sensitive to the defect. The second signal is captured using a low excitation frequency to obtain a relatively high level of sensitivity to the support plate. Figures 6(a) and 6(b) represent impedance-plane trajectories obtained at high and low excitation frequencies, respectively. The two signals are then "mixed" to suppress the support-plate signal. Mixing is performed by first transforming (rotating, scaling, and translating) the second impedance-plane trajectory. The

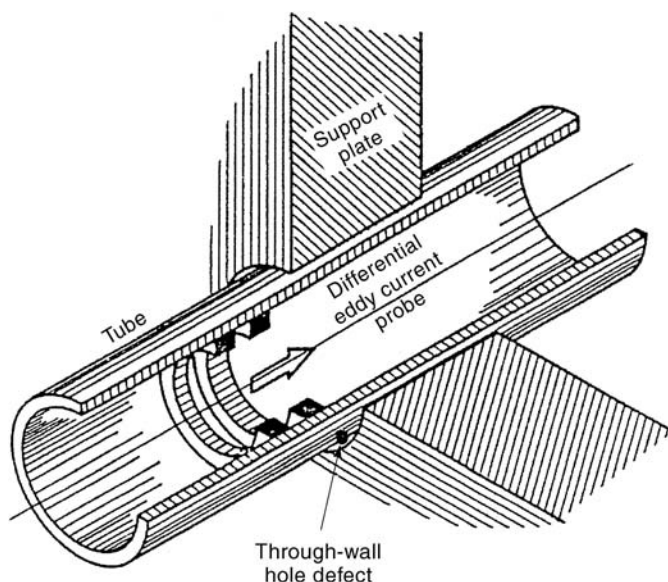
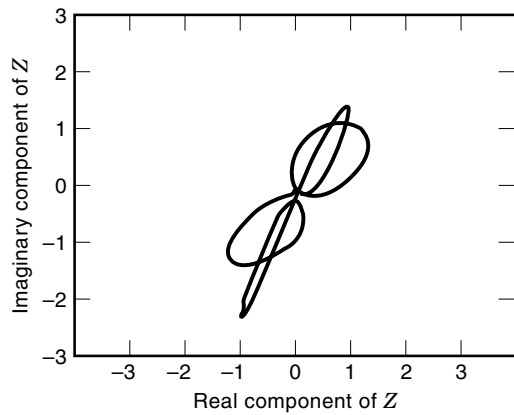
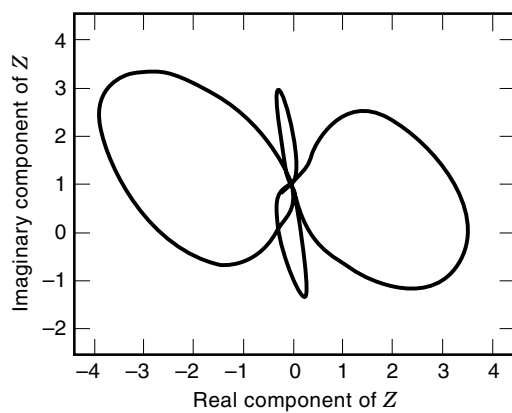


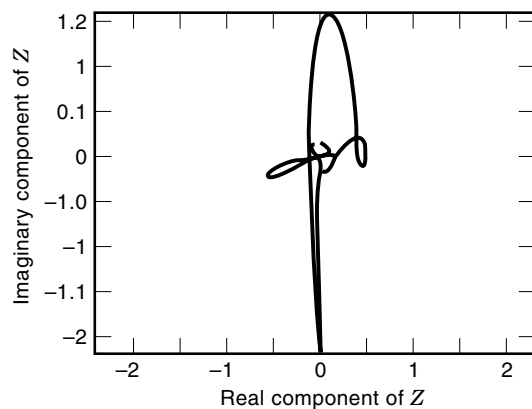
Figure 5. Differential eddy current within a tube and moved by a ferromagnetic support plate. The tube contains a defect near the support plate.



(a)



(b)



(c)

Figure 6. (a) Composite impedance-plane trajectory obtained at 400 kHz excitation frequency; (b) Composite impedance-plane trajectory obtained at 200 kHz; (c) Signal obtained after mixing signals shown in (a) and (b).

transformation parameters are chosen such that the transformed and the second impedance-plane trajectory are as similar as possible to each other. The transformed signal is subtracted from the first signal to obtain the defect signal. Figure 6(c) shows the result of mixing the signals shown in Figs. 6(a) and 6(b). The mixing can be accomplished either manually or automatically (7,8) by estimating the transformation parameters in a manner that minimizes an appropriate cost function. The cost

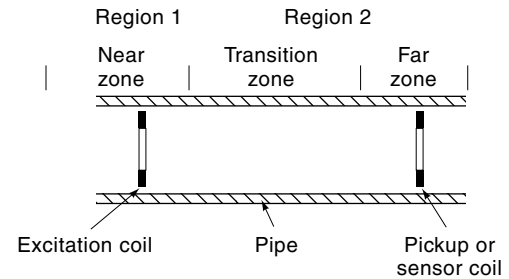


Figure 7. Typical remote-field eddy current probe arrangement.

function could, for example, be the energy in the error between the transformed and first impedance-plane trajectories. This example shows how a pair of eddy current signals obtained at two different excitation frequencies can be “mixed” to suppress indications from unwanted artifacts. The concept can be easily extended for mixing more than two signals for suppressing one or more artifact indications.

Skin-effect considerations limit the use of conventional eddy current methods to the detection of either surface breaking or shallow defects. Inner-diameter (ID) eddy current probes, for example, are more sensitive to defects that lie on the inner surface of thick-walled tubes. Special techniques that increase the level of penetration or alternatively the use of high-sensitivity sensors such as SQUIDs are required if ID probes are to be used. Recent years have witnessed the growth in popularity of a new technique that exploits the remote-field eddy current phenomenon (9,10). A major advantage lies in the sensitivity of the technique to inner- as well as outer-diameter (OD) defects. Figure 7 shows a typical remote-field eddy current probe used for the inspection of pipes. An excitation coil that is energized by a relatively low-frequency ac source establishes the eddy current field. Remote-field eddy current methods employ measurements taken at a distance from the excitation coil, unlike conventional methods that rely on field measurements in the vicinity of the excitation coil. The sensor coil is typically located two to three pipe diameters away from the excitation coil along the axis of the pipe. The eddy current field distribution can be divided into three regions. The field in the vicinity of the excitation coil (region 1) follows intuition in that the field magnitude decays exponentially with increasing radial distance from ID to OD. In the remote region (region 3), which occurs at a distance two to four pipe diameters away from the excitation coil along the axis of the pipe, the field magnitude decays from the OD to ID. This is in complete contrast to the behavior of the field in the vicinity of the excitation coil. The field distribution in the transition zone between these two regions (region 2) is characterized by very rapid changes. Energy that is directed outward, away from the excitation coil, interacts with the energy that travels close to the outer surface before traveling inward. The interaction causes the eddy current and vector magnetic potential magnitudes to drop to zero (potential valleys) at points in the pipe wall. Similarly, the phase of the eddy current undergoes rapid transitions (phase knots) in the pipe wall. Figure 8 shows the energy-flow pattern in the pipewall. The phenomenon is interesting in that it is characterized by a process wherein the field levels decay both from ID to OD (region 1) as well as from OD to ID (region 3). This renders the method sensitive to both ID as well as OD flaws in contrast to conventional eddy

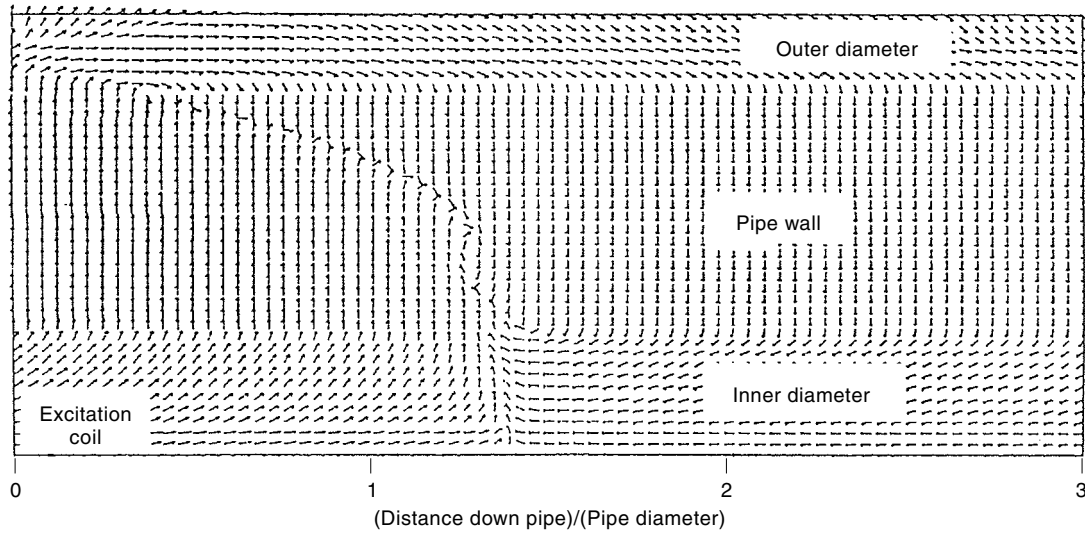


Figure 8. Energy-flow pattern in the pipewall.

current methods that are sensitive to ID flaws only (for an ID probe).

Unlike conventional eddy current methods, the remote-field methods typically involve the measurement of the phase difference between the excitation signal and the voltage induced in the sensor coil. A lock-in amplifier is typically used since the field values in the remote and transition regions are very low and measurement is difficult due to poor signal-to-noise ratios.

The remote-field eddy current methods are used largely for the inspection of ferromagnetic tubes. Extensions of the method for the inspection of nonferromagnetic pipes as well as flat sheets and plates have been proposed in recent years. The industry has been quick to exploit the phenomena and many commercial systems are now available.

Multifrequency methods use two or more excitation frequencies to exploit the skin-effect phenomenon and to heighten selectively the flaw detection sensitivity at various depths. An alternative is to use a spectrally rich signal such as a pulse to excite the probe (11). Figure 9 shows a typical setup using a pulsed eddy current to inspect spot welds. A relatively high-energy pulse energizes the excitation coil. The energy diffuses through the material and induces a signal in the sensor coil. The sensor coil signal is analyzed to ascertain the condition of the test specimen. Pulsed eddy current methods are used for characterizing thin films, spot welds, and a host of other applications.

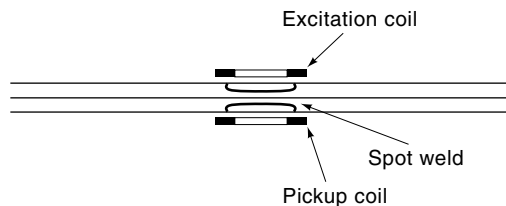


Figure 9. Typical pulsed eddy current probe setup for inspecting spot welds.

Instrumentation. The instrumentation used is simple and straightforward. Figure 10 shows a block diagram of a typical eddyscope. The excitation coil is energized by a variable frequency sinusoidal current source. If the impedance of the coil is $Z \angle \theta$ then the voltage across the coil is

$$V = I|Z| \cos(2\pi f_c t + \theta) \quad (12)$$

where f_c is the excitation frequency (Hz), $\theta = \tan^{-1}(2\pi f_c L/R)$ (rad), L is the effective inductance of the coil (H), and R is the effective resistance of the coil (Ω). The signal V is amplified using a low-noise, wideband amplifier.

The real component of the eddy current probe impedance is obtained by demodulating the amplified signal with the in-phase output ($A \cos 2\pi f_c t$) of the variable-frequency oscillator and low-pass filtering the result. The output of the demodulator is

$$\begin{aligned} V_p &= k \cos(2\pi f_c t + \theta) \cos(2\pi f_c t) \\ &= \frac{k}{2} [\cos(4\pi f_c t + \theta) + \cos(\theta)] \end{aligned} \quad (13)$$

where k is related to the magnitude of the eddy current probe impedance Z . The high-frequency component is filtered by a low-pass filter. The output of the low-pass filter is

$$V_r = C_1 |Z| \cos(\theta) \quad (14)$$

where C_1 is a constant.

Similarly, the imaginary component of the eddy current probe impedance is obtained by demodulating the amplified signal with the quadrature output ($A \sin 2\pi f_c t$) of the oscillator and low-pass filtering the output. The output of the demodulator is

$$\begin{aligned} V_q &= k \cos(2\pi f_c t + \theta) \sin(2\pi f_c t) \\ &= \frac{k}{2} [\sin(4\pi f_c t + \theta) + \sin(\theta)] \end{aligned} \quad (15)$$

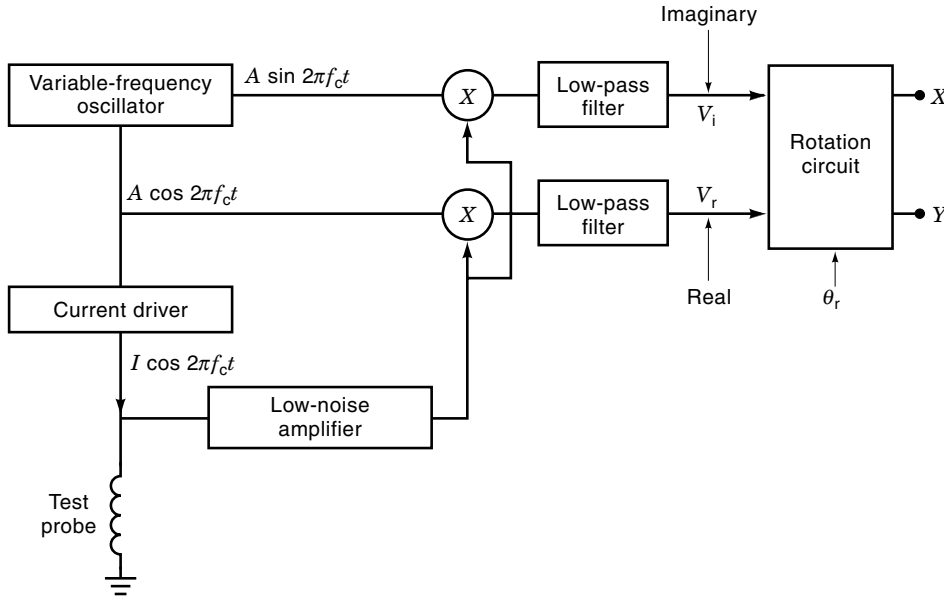


Figure 10. Block diagram of a single-frequency eddyscope.

The output of the low-pass filter is

$$V_i = C_2 |Z| \sin \theta \quad (16)$$

where C_2 is a constant. V_r and V_i are outputs that are related to the real and imaginary components of the probe impedance. Most commercial instruments allow the signals to be sampled, digitized, and stored in memory. Eddyscopes also offer a feature that permits the impedance-plane trajectories to be rotated through an arbitrary angle θ_r . The output of the rotator is given by

$$x = V_r \cos \theta_r - V_i \sin \theta_r \quad (17)$$

$$y = V_i \cos \theta_r + V_r \sin \theta_r \quad (18)$$

The rotation feature allows the user to rotate the signal and ensure that signal changes due to lift-off (distance between the probe and the test specimen) variations result in a trajectory that is either along the real or imaginary axis. After the rotation angle is adjusted, subsequent measurements are projected along the other axis to obtain relative immunity to the effects of lift-off.

Most eddyscopes are either housed in or interfaced to a personal computer. Highly sophisticated signal-processing algorithms can be implemented on the computer to analyze the signals.

Multifrequency eddy current instruments are similar to single-frequency instruments except that it is possible to apply two or more excitation signals to the eddy current coil either simultaneously or in a time-multiplexed mode. If the excitation signals are applied in a time-multiplexed mode, an appropriate time-windowing scheme is employed to separate the individual eddy current responses at each excitation frequency. If the excitation signals are applied simultaneously, classical frequency-division demultiplexing methods can be used to isolate the individual eddy current responses (12).

FORWARD PROBLEM MODELING

Computational eddy current models are extremely valuable in providing a visualization of the field distribution around the probe coil and the test sample. Such capabilities help not only in understanding the physics of the underlying process but also in optimization of probe design as well as the development of defect characterization schemes for mapping measured eddy current signals onto defect profiles. Eddy current models developed to date fall broadly into two main classes, namely, analytical and numerical.

Analytical Models

One of the earliest analytical results for the eddy current phenomenon was obtained by Dodd and Deeds (13). The governing equation in terms of the vector magnetic potential $\mathbf{A}$ for a linear isotropic, homogeneous media due to applied current density $\mathbf{J}_s$ is

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J}_s + \mu \sigma \frac{\partial \mathbf{A}}{\partial t} \quad (19)$$

In the case of an axisymmetric geometry with only the ϕ component of $\mathbf{J}$ and $\mathbf{A}$, and a sinusoidal excitation, the equation in cylindrical coordinates is

$$\frac{\partial^2 \mathbf{A}}{\partial r^2} + \frac{1}{r} \frac{\partial \mathbf{A}}{\partial r} + \frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{\mathbf{A}}{r^2} = -\mu \mathbf{J}_s + j\omega \mu \sigma \mathbf{A} \quad (20)$$

Assuming the probe to be a δ -function coil at (r_0, z_0) , the equation becomes

$$\frac{\partial^2 \mathbf{A}}{\partial r^2} + \frac{1}{r} \left(\frac{\partial \mathbf{A}}{\partial r} \right) + \frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{\mathbf{A}}{r^2} - j\omega \mu \sigma \mathbf{A} = \mu \mathbf{J}_s \delta(r - r_0) \delta(z - z_0) \quad (21)$$

This equation is solved using the separation of variables method for a coil over a planar conductor or a coil inside a cylindrical

tube as shown in Fig. 3(a). The solution for both geometries are given in terms of integrals of Bessel functions. The vector potential due to other coil configurations is then obtained using the superposition principle.

In an attempt to establish a more quantitative interpretation of the eddy current response, Auld et al. (14) formulated a mathematical flaw response model for rectangular surface-breaking defects based on the Lorentz reciprocity relation

$$\nabla \cdot (\mathbf{E}' \times \mathbf{H} - \mathbf{E} \times \mathbf{H}') = 0 \quad (22)$$

where $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{E}'$, $\mathbf{H}'$ are solutions of Maxwell's equations, at any point in the presence and absence of the flaw, respectively.

A third approach for obtaining an analytical solution proposed by Sabbagh et al. (15) uses the volume integral formulation. Based on an algorithm used for simulating the electromagnetic responses of three-dimensional bodies in layered media (16), Sabbagh et al. developed an eddy current NDT model where the inhomogeneities in a conducting specimen are replaced by an equivalent current distribution approximated by pulse basis functions. The current source representing the flaw produces a perturbation field at the probe coil that is then responsible for the eddy current signal. The change in probe impedance due to a flaw is defined as

$$\Delta Z = \frac{(\text{emf induced by the perturbed field at the coil})}{(\text{excitation current})} \quad (23)$$

The conductivity $\sigma_0(\mathbf{r})$ of the test material is expressed as $\sigma_0(\mathbf{r}) = \sigma_f(\mathbf{r}) + \delta\sigma_0(\mathbf{r})$, where $\sigma_f(\mathbf{r})$ is the conductivity of a flaw and $\delta\sigma_0(\mathbf{r})$ is the perturbation in conductivity

$$\delta\sigma_0(\mathbf{r}) = \begin{cases} 0 & \text{outside the defect} \\ \sigma_f(\mathbf{r}) - \sigma_0(\mathbf{r}) & \text{inside the defect} \end{cases} \quad (24)$$

Substituting this conductivity function into the diffusion equation for $\mathbf{E}$, we see that the discontinuity in conductivity introduces a perturbation source term in the equation as

$$\mathbf{J}_f = [\sigma_f(\mathbf{r}) - \sigma_0(\mathbf{r})]\mathbf{E}(\mathbf{r}) \quad (25)$$

$$\nabla^2 \mathbf{E} - j\omega\mu\sigma_0\mathbf{E} = j\omega\mu\mathbf{J}_0 + j\omega\mu[\sigma_f(\mathbf{r}) - \sigma_0(\mathbf{r})]\mathbf{E} \quad (26)$$

The implementation of this method for the axisymmetric geometry consisting of an excitation and pickup coil pair moving in the interior of a tube consists of the following steps.

1. Compute the Green's function for the desired positions of the source. For the given geometry, $G_{ij}(\mathbf{r}, \mathbf{r}')$ corresponds to a field at $\mathbf{r}$ in region i due to a filamentary coil at $\mathbf{r}'$ in region j . The regions are labeled as 1, tube interior; 2, tube wall; 3, tube exterior.
2. Calculate the unperturbed field for the defect-free medium as

$$\mathbf{E}_0(\mathbf{r}) = \int_{\text{coil}} G(\mathbf{r}, \mathbf{r}') \mathbf{J}_0(\mathbf{r}') d\mathbf{r}' \quad (27)$$

3. Calculate the equivalent current source $\mathbf{J}_f(\mathbf{r})$ for the flaw by solving

$$\mathbf{E}_f(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int_{\text{flaw}} G_{12}(\mathbf{r}, \mathbf{r}') [\sigma_f(\mathbf{r}') - \sigma_0(\mathbf{r}')] \mathbf{E}_f(\mathbf{r}') d\mathbf{r}' \quad (28)$$

4. Compute the perturbation field at the coil using the current source from step 3:

$$\mathbf{E}_f(\mathbf{r}) = \mathbf{E}_0(\mathbf{r})|_{\text{coil}} = \int_{\text{flaw}} G_{12}(\mathbf{r}, \mathbf{r}') \mathbf{J}_f(\mathbf{r}') d\mathbf{r}' \quad (29)$$

5. Compute the emf induced in the coil using

$$\text{emf} = n \int_{\text{coil}} \mathbf{E}_f(\mathbf{r}) \cdot d\mathbf{l} \quad (30)$$

The eddy current probe impedance is then calculated using $Z = \text{emf}/I$.

Numerical Models

Analytical models are desirable because they provide exact closed-form solutions that provide a functional relation between experimental parameters and the measured eddy current signal. However, the relative complexity of such methods and their limitations in terms of their inability to deal with nonlinearities and the restricted classes of problem geometries that can be handled have limited their usefulness. In contrast, numerical solutions are not limited by geometrical complexities or material nonlinearities. The high-speed computational power and storage capabilities of modern-day computers have led to the extensive development and use of numerical techniques for solving and simulating a wide variety of physical situations. The principal numerical methods that are in use in electromagnetic NDT are the finite-difference, finite-element, and hybrid techniques such as boundary-element methods.

Finite-Difference Formulation. The finite-difference method is based on replacing domains and differential operators by a discrete grid of nodes and difference quotients, respectively. Once the difference equations are obtained, the task of solving the problem numerically consists of writing the difference equations for each nodal point in the grid in terms of the values at the appropriate neighboring nodes. This results in a set of linear algebraic equations written in a matrix form. The matrix of coefficients is then inverted to obtain the unknown fields at the nodal points.

Dodd (17) solved the complete, nonlinear eddy current problem for the inhomogeneous material case using finite-difference operators. For axisymmetric geometries such as a coil above a conducting plane or a coil encircling a tube with sinusoidal excitation, the equation can be written as

$$\begin{aligned} \frac{\partial^2 A}{\partial r^2} + \frac{1}{r} \frac{\partial A}{\partial r} + \frac{\partial^2 A}{\partial z^2} - \frac{A}{r^2} \\ = -\mu J_s + j\omega\mu\sigma A - \mu \left[\frac{\partial}{\partial r} \left(\frac{1}{\mu} \right) \left(\frac{1}{r} \frac{\partial r A}{\partial r} \right) + \frac{\partial}{\partial z} \left(\frac{1}{\mu} \right) \frac{\partial A}{\partial z} \right] \end{aligned} \quad (31)$$

Using central difference operators for all differentiations except for permeability variations across interface boundaries where forward difference operators are used, the vector potential at a point (r, z) in terms of four nearest neighbors is

$$A_{r,z} = \left(\frac{1+a}{r} \frac{\mu_{r,z}}{\mu_{r+a,z}} A_{r+a,z} + A_{r-a,z} + \frac{\mu_{r,z}}{\mu_{r,z+a}} A_{r,z+a} + A_{r,z-a} + a^2 \mu_{r,z} J_{r,z} \right) + \left(z + \frac{a}{r} + \frac{a^2}{r^2} + \frac{\mu_{r,z}}{\mu_{r+a,z}} + \frac{\mu_{r,z}}{\mu_{r,z+a}} + ja^2 \omega \mu_{r,z} \sigma_{r,z} \right) \quad (32)$$

where $\mu_{r,z}$, $J_{r,z}$, and $\sigma_{r,z}$ have specified values at each nodal point in the two-dimensional rectangular grid. The nonlinearities associated with medium permeability are also incorporated in the formulation without additional computations.

The major disadvantage of the finite-difference method (17) is one of fitting a regular grid to geometries that have curved boundaries and interfaces. One has to resort to using a mixture of rectangular and cylindrical coordinates, resulting in added complexity of programming and implementation. Also, increasing the mesh density to satisfy the boundary requirements may result in a large grid size and computer requirements without contributing to the accuracy of the final solution.

Finite-Element Formulation. A more widely used numerical model employs finite-element (FE) analysis techniques and has its origin in the fields of solid mechanics, structural analysis, and heat transfer. Finite-element analysis techniques was first applied by Chari to electromagnetics (18) for studying the eddy current problem in magnetic structures and later by Brauer (19) for determining induced fields and currents in transformers. The modeling technique was then used extensively by Lord and associates for the investigation of various problems in electromagnetic NDT. Palanisamy and Lord (20,21) developed a two-dimensional axisymmetric finite-element model for predicting eddy current probe signals from defects in steam generator tubing. Ida (22) developed a three-dimensional finite-element model for eddy current nondestructive evaluation applications.

The finite-element technique is based on principles of variational calculus (23) in which the solution of a differential equation is obtained as the stationary value of a functional. In eddy current problems, the functional represents the energy of the system so that the stationary value is a minimum. The principles of variational calculus can be used to show that the differential equation is satisfied by a function that simultaneously minimizes the appropriate energy functional (24). A brief description of the finite-element formulation for an axisymmetric geometry of a differential coil probe inside a tube follows.

Assuming a linear, isotropic, and homogeneous medium and a sinusoidal excitation source, the governing diffusion Eq. (20) is

$$\frac{1}{\mu} \left(\frac{\partial^2 \mathbf{A}}{\partial r^2} + \frac{1}{r} \frac{\partial \mathbf{A}}{\partial r} + \frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{\mathbf{A}}{r^2} \right) = j\omega\sigma \mathbf{A} - \mathbf{J}_s \quad (33)$$

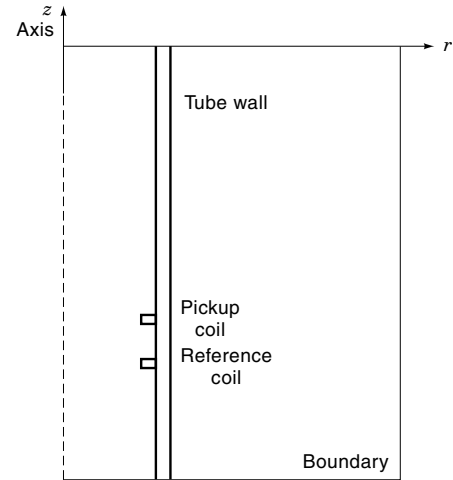
The corresponding energy functional obtained from variational principles is

$$F = \iint_R \left[\frac{1}{2\mu} \left(\left| \frac{\partial \mathbf{A}}{\partial r} + \frac{\mathbf{A}}{r} \right|^2 + \left| \frac{\partial \mathbf{A}}{\partial z} \right|^2 \right) + \frac{j\omega\sigma}{2} |\mathbf{A}|^2 - \mathbf{J}_s \times \mathbf{A} \right] r dr dz \quad (34)$$

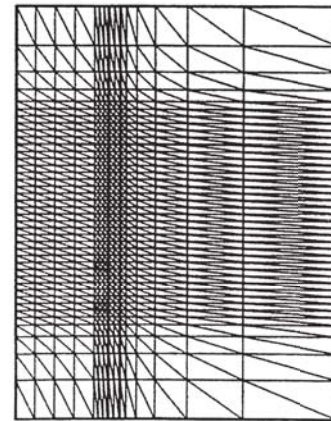
The terms in Eq. (34) can be recognized as the energy terms due to the magnetic field, eddy currents, and the input source.

The region of interest is discretized using a mesh consisting of triangular elements connected to each other at the nodes. A sample FE mesh for the axisymmetric geometry in Fig. 3 is shown in Fig. 11. Using linear or higher-order interpolation functions $[N]$ to express the magnetic vector potential $\mathbf{A}$ at any point within an element, the energy functional for an element can be written in terms of its unknown nodal values A_i, A_j, A_k as

$$F_e = \int_v \frac{1}{2} \left[\frac{1}{\mu} \left(\frac{\partial N}{\partial y} \right)^2 + \frac{1}{\mu} \left(\frac{\partial N}{\partial x} \right)^2 - 2J[N] \right] \begin{bmatrix} A_i \\ A_j \\ A_k \end{bmatrix} \quad (35)$$



(a)



(b)

Figure 11. (a) Two-dimensional region in the axisymmetric geometry; (b) FE mesh of the region in (a).

Minimizing F_e with respect to the nodal point values gives a set of linear algebraic equations

$$\frac{\partial F_e}{\partial A_l} = 0, \quad l = i, j, k \quad (36)$$

in three unknowns A_i, A_j, A_k . These individual element equations are combined into a single global matrix equation

$$\left[\begin{array}{c} \text{Diagram of a rectangular element} \end{array} \right] \begin{Bmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{Bmatrix} = \begin{Bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{Bmatrix} \quad (37)$$

Due to the banded, symmetric, and sparse nature of the matrix, direct inversion techniques such as the Gaussian elimination method can be used to solve for the unknown vector magnetic potentials at the nodal points in the mesh.

Calculation of Eddy Current Probe Impedance

In eddy current NDT, the physical quantity that is usually measured is the impedance of the probe. The eddy current probe impedance is derived from nodal values of magnetic vector potential. The impedance Z of a single-turn coil of radius r carrying an alternating current of I_s amperes is

$$Z = \frac{V}{I_s} \quad (38)$$

where V is the phasor voltage across the coil,

$$V = - \oint_c \mathbf{E} \cdot d\mathbf{l} \quad (39)$$

Using the Maxwell–Faraday law in terms of the vector magnetic potential, $\mathbf{E}$ can be expressed as

$$\mathbf{E} = \frac{-\partial \mathbf{A}}{\partial t} - \nabla \phi \quad (40)$$

which for sinusoidal excitation reduces to

$$\mathbf{E} = -j\omega \mathbf{A} - \nabla \phi \quad (41)$$

Since the induced voltage is independent of the scalar potential, Eq. (41) can be substituted in Eq. (39) to get

$$\mathbf{E} = -j\omega \oint_c \mathbf{A} \cdot d\mathbf{l} \quad (42)$$

The impedance of the coil is

$$Z = -\frac{j\omega}{I_s} \oint_c \mathbf{A} \cdot d\mathbf{l} \quad (43)$$

which for a single-turn coil is

$$Z = j \frac{2\pi r \omega A}{I_s} \quad (44)$$

Model predicted eddy current differential probe signals for both outside-diameter and inside-diameter defects in a steam generator tube compare very well with experimental measurements (20).

The finite-element method is a very powerful tool for analyzing nonlinear, inhomogeneous boundary-value problems. Nonlinearities and inhomogeneous material properties can be incorporated very simply by specifying the material constants appropriately in each element of the discretized domain. A major advantage of the finite-element technique is that irregularly shaped boundaries can be handled easily. It is therefore ideally suited for modeling NDT problems. The boundary conditions are introduced naturally in the process of formulation of the model. Also, since the nodes do not have to be equally spaced, the mesh density can be varied depending on the field gradient in a subdomain, resulting in computational efficiencies. Finally, the implementation is relatively simple and since the matrices involved in finite-element models are sparse, banded, symmetric and diagonally dominant, the inversion process is numerically stable and robust.

The finite-element eddy current model has been extended to simulate the pulsed eddy current phenomenon (25). In this method, the spatial solution of the conventional eddy current model is combined with a finite-difference formulation in the time domain to calculate the transient solution. The model is capable of simulating the diffusion of fields into the test sample, spatial broadening of the eddy current pulse, and a propagation delay time that is proportional to the distance traveled.

EDDY CURRENT INVERSE PROBLEM

The final goal of a nondestructive test is to solve the inverse problem. The complete solution to the inverse problem involves full reconstruction of defect profiles from the sensor measurements. In eddy current testing, the physical process associated with eddy current phenomenon is diffusive in nature. This renders the task of solving the inverse problem relatively difficult. The intractable nature of the inverse problem in eddy current nondestructive evaluation (NDE) has led to the development of several engineering approaches to address the issue. One of the earliest and most widely used approaches is the calibration method, in which features associated with the measured data are compared to those derived from a set of standard signals obtained from known defects. Calibration plots for eddy current probe signals were obtained by Auld et al. (26) in which features derived from an absolute coil probe signal are directly related to defect parameters. For instance, when the flaw dimensions are larger than the coil diameter, the signal exhibits distinct “shoulders” that can be mapped to edges of the defects. The calibration plot using the features in the signal for different defect widths and theoretical inversion charts using magnitude and phase of the signal to predict defect depths are also given in Ref. 26. These procedures characterize defects in terms of an equivalent depth, width, and angle or equivalent volume. It must be mentioned that calibration methods are useful only when the defect shapes are known a priori. The method becomes invalid if the defect shapes are substantially different from the shapes of the calibration defects.

With increasing availability of greater computing resources, more sophisticated approaches to numerical inversion algorithms have been developed. These algorithms can be

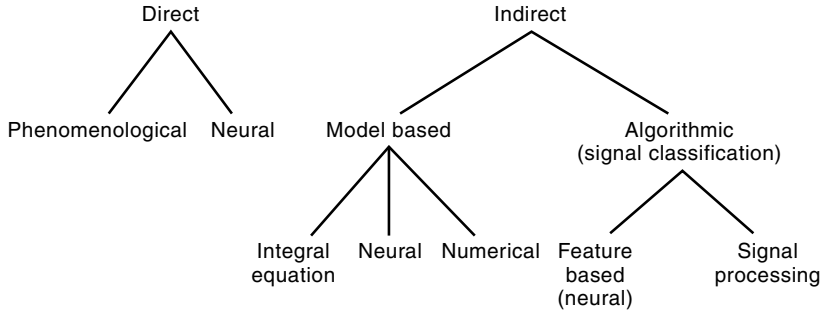


Figure 12. Taxonomy of inverse-problem solution methods.

categorized broadly either as phenomenological or model-based and nonphenomenological or signal-classification-based methods, based on the strategy used to arrive at the inverse solution. Phenomenological models are constructed on the basis of the underlying physical process. In contrast, signal classification methods use changes in the signal features to determine the defect class. A taxonomy of inverse problem solution methods in NDE is described in Fig. 12 and an overview of some of the key work in these classes is described.

MODEL-BASED METHODS

In this section, three representative model-based approaches are discussed for finding the inverse-problem solution in terms of the defect profile in a test object. The first method is based on a volume integral formulation. The second approach uses an error minimization approach that involves the use of a neural network for minimizing the quadratic error function. The third approach uses a finite element forward model in an iterative scheme to determine the defect profile.

Volume Integral Approach

This technique belongs to the category of direct approaches for the inverse-problem solution. A seminumerical approach proposed by Sabbagh et al. (15) uses an analytical formulation based on the Green's function where the inverse problem is reduced to solving an integral equation of the type

$$f(r) = \int_{\Omega} G(r, r') g(r') d\gamma' \quad (45)$$

for the source $g(r')$ from measurements $f(r)$, where the kernel $G(r, r')$ is the Green's function. For instance, in eddy current NDE, the volume integral equation for the electric field $\mathbf{E}$ is

$$\mathbf{E}(r) = \int_{\Omega} G(r, r') \mathbf{J}_s(r') dr' \quad (46)$$

A flaw of conductivity $\sigma_f(r)$ is modeled in a host material of conductivity $\sigma_h(r)$ by an equivalent current source

$$\mathbf{J}_s(r) = [\sigma_f(r) - \sigma_h(r)] \mathbf{E}(r) \quad (47)$$

The integral equation for the perturbation field $\mathbf{E}_p(r)$ due to the presence of the flaw is

$$\mathbf{E}_p(r) = \int_{\Omega_f} G(r, r') [\sigma_f(r) - \sigma_h(r)] \mathbf{E}(r') dv' \quad (48)$$

The discretization and the numerical solution of the resulting matrix give a solution of the inverse problem in terms of the conductivity profile $\sigma(r)$ of the test specimen.

Hopfield Network Approach

A novel strategy for solving Fredholm integral equations using the Hopfield neural networks was proposed by Elshafiey et al. (27). The major advantage of this method is the stability of the solution procedure that comes from the high degree of parallelism and interconnectivity used in the Hopfield network architecture.

In this algorithm, the integral in Eq. (45) is discretized using a basis-function expansion as

$$g(r) = \sum_{i=1}^n v_i R_i(r) \quad (49)$$

where $\{v_i\}$ are the expansion coefficients and R_i is a basis function, chosen to be a sine or cosine function. The integral equation is then written in matrix form as

$$FV = m + \eta \quad (50)$$

where $F_{ij} = \int_{\Omega} G(r, r') R_i(r') dr'$, $m_i = f(r_i)$, $i = 1, 2, \dots, n$ are the discrete measurements, and η represents measurement noise.

The solution vector $V = (v_1, v_2, \dots, v_n)$ is obtained by minimizing the cost function E written in matrix form

$$E = \frac{1}{2} (FV - m)^T (FV - m) + \lambda V^T DV + \lambda_1 (\Psi V - z_p)^T (\Psi V - z_p) \quad (51)$$

The first term in the cost function is the model error, the second term is a smoothness constraint included for regularization, and the third term represents the boundary conditions. Ψ is a matrix whose elements are basis functions, z_p is the system state at a given node p , and λ and λ_1 are Lagrange multipliers.

By comparing the error function in Eq. (51) to the energy function of the Hopfield neural network expressed in a similar quadratic form, the circuit parameters of the network are estimated. The network is then simulated for obtaining the solution to the integral equation. Once the Hopfield network converges to the solution vector $\mathbf{v}^* = (v_1^*, v_2^*, \dots, v_n^*)$, the defect profile can be computed using Eq. (49). The two methods just described are based on the solution of the integral equation, which, in turn, relies on the availability of the Green's function for the problem. Very often in many NDE test geometries,

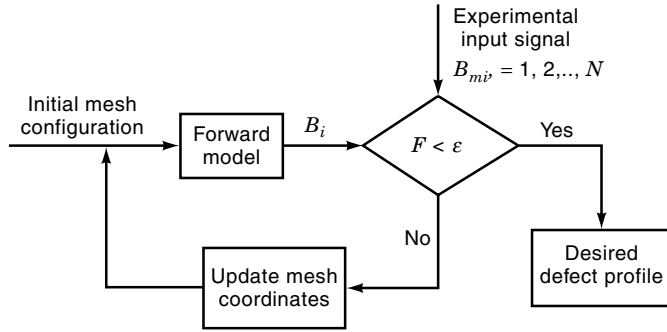


Figure 13. Schematic representation of the solution to the inverse problem.

the computation of Green's function poses a serious challenge, limiting the applicability of the method.

An alternate scheme for eddy current inversion that overcomes this difficulty is to use the finite-element model (FEM) in the iterative scheme shown Fig. 13. In this technique, the defect is parametrized in terms of the depth coordinates at different points in the flaw region and the solution minimizes the mean-square error between the measured signal and model prediction (28).

Figure 13 outlines the procedure employed for solving the inverse problem. The iterative process starts by selecting an initial mesh configuration. The mesh coordinates are then updated until the signal predicted by the finite-element model matches the measured signal in the least-squares sense. Specifically, the coordinates of the defect profile are defined by a vector $D = (d_1, d_2, \dots, d_n)$, which represents the depth values of the defect profile at each position on the defect boundary. The elements above and below the defect boundary are compressed or expanded accordingly. By altering the mesh coordinates, rather than the material properties of the elements, derivatives of the unknown material property values do not have to be taken into account explicitly.

The optimal values of D are determined by minimizing an objective function of the summed square error evaluated at N measurement points:

$$F(D) = \sum_{i=1}^N (Z_i - Z_{mi})^2 \quad (52)$$

where Z_i ($i = 1, \dots, N$) is the calculated coil impedance using the FE method and Z_{mi} ($i = 1, \dots, N$) is the measured eddy current signal.

The error function defined by Eq. (52) is minimized using the sequential quadratic programming (SQP) approach (29). This well-known optimization procedure employs gradient information, which is derived analytically to establish the search direction. Changes in defect geometry are mapped onto a change of stiffness matrix K . The corresponding change in the source vector S is also automatically taken into account.

Model-based methods are, in general, computation intensive and are capable of determining the defect profiles only if the solution space is constrained to a suitable subspace. A different class of inversion techniques that is widely used in practice is the nonphenomenological or signal-classification approach.

SIGNAL-CLASSIFICATION METHODS

In this class of techniques, the continuum of solutions to the inverse problem is quantized into a finite number of known classes of sources or defects. These sources may represent defect classes or benign geometrical sources that also produce an eddy current signal. Examples of the latter class are the eddy current signals from edges or support-plate signals obtained in the inspection of steam-generator tubes in nuclear power plants.

A characteristic feature of all signal-classification or pattern-recognition methods is that they rely on the ability to create a data bank of all expected defect types and the corresponding signatures. This collection of signals is referred to as the training database, which is then used for training an automatic signal-classification algorithm. The classification techniques that have received a lot of attention in this area are (1) matched filters and (2) feature-based pattern-recognition methods.

Matched Filters

The matched filter (30) is a linear filter the transfer function of which is matched to a particular input signal, achieved by selecting the impulse response $h(t)$ of the filter as a reversed, shifted and scaled version of the signal $s(t)$ to which it is matched:

$$h(t) = ks(t_0 - t) \quad (53)$$

where k and t_0 are arbitrary constants. The filter then behaves as a simple correlation detector with transfer function:

$$H(j\omega) = ke^{-j\omega t_0} S^*(j\omega) \quad (54)$$

Mucciardi and Shankar (31) use this technique for the interpretation of steam-generation inspection signals. An array of filters is built in which each one is matched to the signal from a known class of defects. An unknown signal is then automatically classified by observing the response of the filters in the array.

The matched-filter algorithm is simple and easy to implement but is also prone to errors in case of any fluctuations in probe speed. Also, this procedure requires that the entire signal be stored in contrast to feature-based signal-classification methods that store and use only distinguishing features in the signal.

Feature-Based Methods

Feature-based methods form the core of automated signal-classification systems that are becoming increasingly popular in many commercial applications. An overall schematic of this approach is shown in Fig. 14.

The approach consists of two steps:

1. The first step is feature extraction, from which characteristic features in the signal that carry discriminatory information are extracted.
2. The second step is classification of the feature vector via a clustering algorithm or a neural network.

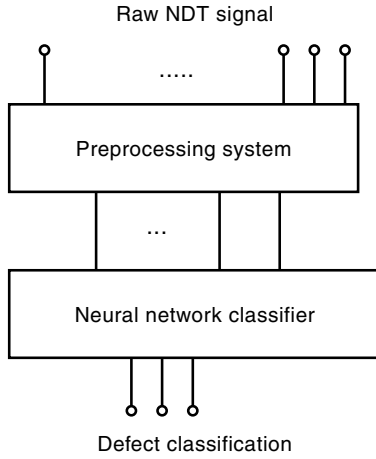


Figure 14. Schematic of the overall signal-classification procedure.

Feature Extraction. Feature extraction serves two major functions, namely, data compression and invariance processing. The vector of signal features referred to as the feature vector is extracted from a typically oversampled set of measurements. The feature extraction procedure is designed to eliminate the redundancy in the impedance plane trajectories. The entire signal is thus represented by a feature vector. The features are evaluated in the second step, called feature reduction, in which the goal is to identify those features containing the most amount of discriminatory information, in order to reduce the dimensions of the feature vector (32). One of the earlier methods used for feature reduction is the adaptive learning network, in which a polynomial representation is built from quadratic elements that use pairwise combinations of the features. The discriminatory performance of each polynomial fit is evaluated at each stage and only the “best” pairs are selected. The adaptive learning network was first introduced by Ivakhenko (33) and was applied for feature selection by Mucciardi and Shankar (31). An alternative method for feature reduction is the Fisher discriminant method (34), which calculates a statistical weight function for each feature as a measure of its ability to classify. The features are often picked from different domains, making it difficult to define a single quantitative measure for evaluating a figure of merit of the features. Furthermore, the time-domain features are sensitive to variations in probe speed and can consequently result in classification errors.

Parameter-Invariant Features. An example of features with invariance properties is obtained using a Fourier descriptor representation of eddy current impedance-plane trajectories. Fourier descriptors have been used for a long time for representing closed contours (35) in a variety of applications. This technique not only represents the signal by a few coefficients that are invariant under rotation, translation, and scaling of the eddy current impedance-plane trajectory, but also allows the resynthesis of the original signal from the stored coefficients. This results in data compression, and the error in the resynthesized signal can be used for a quantitative evaluation of the feature vector. The basic idea underlying the approach is to parametrize the impedance-plane trajectory such that the representation is periodic in the transformed domain. One such

function is the complex contour function (36)

$$u(l) = x(l) + jy(l) \quad (55)$$

where $x(l)$ and $y(l)$ represent the real and imaginary components of the impedance at a distance l arc length units away from an arbitrary starting point P_0 . Then $u(l)$ is periodic in l with period L ,

$$u(l + L) = u(l) \quad (56)$$

Here L represents the total arc length of the curve. The periodic nature of $u(l)$ allows its expansion in a Fourier series, that is,

$$u(l) = \sum_{n=-\infty}^{\infty} c_n e^{j(2\pi/L)nl} \quad (57)$$

where

$$c_n = \frac{1}{L} \int_0^L u(l) \exp\left(-\frac{j2\pi nl}{L}\right) dl \quad (58)$$

In order to obtain descriptors of the curve that are insensitive to drift and gain setting of the eddy current instrument, the coefficients in Eq. (58) are used to compute the descriptors

$$d_n = \frac{c_{l+n}c_{l-n}}{|c_1|^2} \quad (59)$$

The descriptors d_n , $n = 1, 2, \dots, N$ are insensitive to rotation, translation, and scaling of the curve. The implication of this invariance is that changes in the signal due to the instrument gain drift or gain settings will not affect the interpretation of the signal. The parametric representation $u(l)$ also ensures that variations in probe speed will not affect the final result.

The two procedures for feature extraction described previously illustrate the use of appropriate signal features for achieving data compression and invariance to selected experimental parameters. A K -means clustering or a neural network is used to classify the feature vector.

Classification Algorithms

The second step in signal-classification systems is the labeling of the feature vector. Two of the widely used techniques for classifying eddy current signals are (1) clustering algorithms and (2) neural networks.

K-Means Clustering. In a clustering algorithm, a feature vector of length n is treated as a point in an n -dimensional feature space. The set of feature vectors from a similar class of signals are expected to cluster together in the feature space. Clustering algorithms are capable of identifying the clusters in either a supervised (with training data) or unsupervised (without training data) mode. The K -means algorithm (34) partitions the set of feature vectors into K disjoint subsets such that a performance index is minimized. The performance index is typically chosen to be the sum of the squared Euclidean distances between a cluster center and all the points contained within that cluster. The algorithm is iterative in nature and during each cycle the cluster centers are updated. The performance of the algorithm can be improved significantly if the K cluster centers

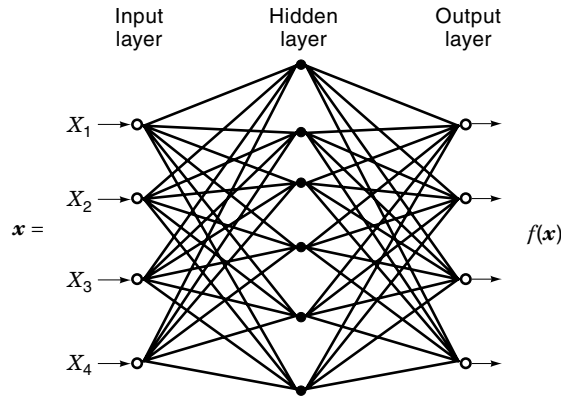


Figure 15. Multilayer perceptron neural network.

during the first iteration are drawn from K different classes of signals in the training data.

Neural Networks. Neural networks represent an attempt to mimic the biological nervous system with respect to both architecture as well as information-processing strategies (37). These networks consist of simple processing elements that are interconnected via weights. The network is first trained using training data and an appropriate learning algorithm for the estimation of the interconnection weights. Once the network is trained, unknown test signals can be classified. The class of neural networks used most often for classification tasks is the multilayer perceptron (MLP) network (37) shown in Fig. 15.

The multilayer perceptron network generally consists of an input layer of nodes, one or more hidden layers of nodes, and an output layer of nodes. The nodes within the same layer are not connected but each layer of nodes are fully interconnected to the nodes in the next layer. All units within a layer process data in parallel, but the outputs of different layers are calculated sequentially starting from the input layer and moving towards the output layer. Each node j in a layer $k + 1$ performs the following computations.

Step 1. The input to node j , x_j , is

$$x_j = \sum_{i=1}^{N_k} w_{ij} y_i \quad (60)$$

where y_i is the output of node i in layer k , N_k is the number of nodes in layer k , and w_{ij} are the interconnection weights.

Step 2. The output of node j , y_j , is

$$y_j = f(x_j) = \frac{1}{1 + \exp[-(x_j + \theta_j)]} \quad (61)$$

where θ_j is a bias variable. This nonlinear function is primarily used to limit the output of a node between the values of 0 and 1.

One of the most commonly used training algorithms is the backward error propagation algorithm (38) in which training patterns are sequentially applied to the network. The algorithm uses a gradient search technique for minimizing the squared

error between the actual output and the desired output by iteratively adapting the interconnection weights. The algorithm cycles through the training data until the error drops below a specified threshold value. Neural networks have been used with considerable success in the classification of eddy current and ultrasonic NDE signals (39,40). These networks are also capable of learning in an incremental mode and use prior knowledge for improving its performance with time.

PULSED EDDY CURRENT SIGNAL PROCESSING

The processing of pulsed eddy current signals for defect characterization is based on a calibration approach where the features are peak amplitude and zero crossover point of the waveform (41). The peak amplitude is proportional to the amount of metal loss and the zero crossover point, which is proportional to the propagation time, carries information about the depth of the flaw. Time gated images of the peak voltage, called C-scan images, provide information relating to material properties at different depths within the sample. The pulsed eddy current technique has proved to be particularly advantageous in the inspection of multilayer structures encountered in the aerospace industry.

BIBLIOGRAPHY

1. D. J. Hagemaiier, *Fundamentals of Eddy Current Testing*, Columbus, OH: American Society for Nondestructive Testing, 1990.
2. R. Halmshaw, *Non-destructive Testing*, London: Edwin Arnold, 1991.
3. J. D. Jackson, *Classical Electrodynamics*, New York: Wiley, 1975.
4. V. S. Cecco, *Eddy Current Manual*, Ontario: Chalk River National Laboratories, 1983.
5. S. Sharma et al., Finite Element Modeling of Eddy Current Probes for Edge Effect Reduction, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1997, vol. 16A, pp. 201–208.
6. W. Podney and J. Moulder, Electromagnetic Microscope for Deep, Pulsed, Eddy Current Inspections, in D. O. Thompson and D. E. Chimenti, (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1997, vol. 16A, pp. 1037–1044.
7. S. D. Brown, Multifrequency/Multiparameter Eddy Current Steam Generator NDE, in R. B. Clough (ed.), *Quantitative NDE in the Nuclear Industry*, Metals Park, OH: ASM, 1983.
8. J. Stolte, L. Udpa, and W. Lord, Multifrequency Eddy Current Testing of Steam Generator Tubes using Optimal Affine Transform, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1988, vol. 7A, pp. 821–830.
9. T. R. Schmidt, The remote field eddy current inspection technique, *Mater. Eval.*, **42**: 225–230, 1984.
10. W. Lord et al., A finite element study of the remote field eddy current phenomenon, *IEEE Trans. Magn.*, **24**: 435–438, 1988.
11. J. L. Fisher and R. E. Beissner, Pulsed Eddy Current Crack Characterization Experiments, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1986, vol. 5A, pp. 199–206.
12. Avanindra, *Multifrequency eddy current signal analysis*, M.S. Thesis, Iowa State University, Ames, IA, 1997.
13. C. V. Dodd and W. E. Deeds, Analytical solutions to eddy current probe coil problems, *J. Appl. Phys.*, **39** (6): 2829–2938, 1968.
14. B. A. Auld, F. Muennemann, and M. Riazat, Quantitative modeling of flaw and responses in Eddy Current Testing, in R. S. Sharpe

- (ed.), *Research Techniques in Nondestructive Testing*, New York: Academic Press, 1984, vol. 7.
15. H. A. Sabbagh and D. L. Sabbagh, *Development of a System to Invert Eddy Current Data and Reconstruct Flaws*, Final Report, Contract No. N60921-81-C-0302 with Naval Surface Weapons Center Code R34, White Oaks Laboratories, June, 1982.
 16. P. E. Wannamaker et al., Electromagnetic modeling of three dimensional bodies in layered earths using integral equations, *Geophysics*, **49** (1): 60–74, 1984.
 17. C. V. Dodd, *Solutions to electromagnetic induction problems*, Ph.D. Dissertation, Univ. Tennessee, Knoxville, 1967.
 18. M. V. Chari, Finite element solution of the eddy current problem in magnetic structures, *IEEE Trans. Power Appar. Syst.*, **PAS-93**: 62–72, 1974.
 19. J. R. Brauer, Finite element analysis of electromagnetic induction in transformers, presented at the *IEEE Winter Power Meet.*, New York, January, 1977.
 20. R. Palanisamy and W. Lord, Finite element analysis of eddy current phenomena, *Mater. Eval.*, **38** (10): 39–43, 1980.
 21. R. Palanisamy and W. Lord, Finite element analysis of axisymmetric geometries in quantitative NDE, in *Proc. ARPA/AFML Review of Progress in Quantitative NDE*, San Diego, July 1979.
 22. N. Ida, *Three dimensional finite element modeling of electromagnetic NDT phenomena*, Ph.D. Dissertation, Colorado State Univ., CO, 1983.
 23. L. J. Segerlind, *Applied Finite Element Analysis*, New York: Wiley, 1976.
 24. O. C. Zienkiewicz, *The Finite Element Method in Engineering Science*, London: McGraw-Hill, 1971.
 25. B. Allen, *Finite element modeling of pulsed electromagnetic NDT phenomena*, M.S. Thesis, Colorado State University, CO, 1983.
 26. B. A. Auld et al., Improved Prove-Flaw Interaction Modeling, Inversion Processing and Surface Roughness Clutter, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative NDE*, New York: Plenum, 1984, vol. 2.
 27. I. Elshafiey, L. Udpa, and S. S. Udpa, Solution of inverse problems in electromagnetics using Hopfield networks, *IEEE Trans. Magn.*, **31**: 852–861, 1995.
 28. M. Yan et al., Solution of inverse problems in electromagnetic NDE using finite element methods, *IEEE Trans. Magn.*, **34**: September 1998.
 29. K. Shittkowski, On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search functions, in *Series Optimization*, Mathematische Operationsforschung und Statistik, vol. 14, 1983.
 30. G. L. Turin, An introduction to matched filters, *IRE Trans. Inf. Theory*, **IT-6**: 311–329, 1960.
 31. A. N. Mucciardi and R. Shankar, *Signal Processing for ISI*, EPRI NP-121, Sec. 4, May 1980.
 32. P. A. Doctor et al., Pattern Recognition Methods for Classifying and Sizing Eddy Current Using Eddy Current Data, in G. Birnbaum and G. Free (eds.), *Eddy Current Characterization of Materials and Structures*, ASTM STP 722(ASNT), Philadelphia: ASTM, 1981.
 33. A. G. Ivakhenko, The group method of data handling—A revival of the method of stochastic approximation, *Sov. Autom. Control*, **13** (3): 43–55, 1968.
 34. J. T. Tou and R. C. Gonzales, *Pattern Recognition Principles*, Reading, MA: Addison-Wesley, 1974.
 35. E. Persoon and K. S. Fu, Shape discrimination using Fourier descriptors, *IEEE Trans. Syst. Man Cybern.*, **SMC-7**: 170–179, 1977.
 36. C. T. Zahn and R. Z. Roskies, Fourier descriptors for plane closed curves, *IEEE Trans. Comput.*, **C-21**: 269–281, 1972.
 37. R. P. Lippman, An introduction to computing with neural nets, *IEEE ASSP Mag.*, **4** (2): 4–21, 1987.
 38. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning Internal Representations by Error Propagation, in D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Cambridge, MA: MIT Press, 1986.
 39. L. Udpa and S. S. Udpa, Eddy current defect characterization using neural networks, *Mater. Eval.*, **48** (3): 342–347, 1990.
 40. J. Chao, L. Udpa, and S. S. Udpa, Ultrasonic Signal Analysis Using Wavelet Transform, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1993, pp. 735–742.
 41. J. A. Bieber et al., Time-Gating of Pulsed Eddy Current Signals for Defect Characterization and Discrimination in Aircraft Lap-Joints, in D. O. Thompson and D. E. Chimenti (eds.), *Review of Progress in Quantitative Nondestructive Evaluation*, New York: Plenum, 1997, vol. 16, pp. 1915–1921.

SATISH S. UDPA
LALITA UDPA
Iowa State University

IMPULSE TESTING

All electrical equipment is subjected to voltage surges (impulses) during operation. These surges vary in magnitude and duration depending on their origin and the system parameters in which the equipment is installed. Impulse voltages generated in a piece of equipment or a system are termed overvoltages because they usually exceed the normal operating voltage.

In designing electrical equipment, two main factors are taken into consideration. The first is the insulation, which depends not only on the operating voltage but also on the voltage surges that it may be subjected to during operation. Consequently, the insulation of a piece of electrical equipment should provide complete isolation between electrically live parts and other parts, either live or dead, for the safety of the equipment and personnel. The second factor is the equipment current. It should not produce excessive heating, which may shorten the life of the equipment or damage its conducting and insulating materials.

Overvoltages in a piece of electrical equipment installed in a power system are generated either from external or internal sources. Lightning is an external source of overvoltage to power systems, which is usually characterized by large magnitudes and a short duration. Switching operations in power systems are internal sources of overvoltages and are characterized by lower magnitudes and a longer duration.

Electrical equipment is subjected to impulse testing to verify its ability to withstand surge voltages or currents to which it may be subjected during operation.

LIGHTNING OVERVOLTAGES

Upward humid air drafts from the earth's surface reaching high altitudes are subjected to relatively low temperatures where their moisture content condenses into large water drops. These drops disintegrate into smaller ones. The process of disintegration produces electrification where the small drops carry positive charges and the negative charges are carried by the

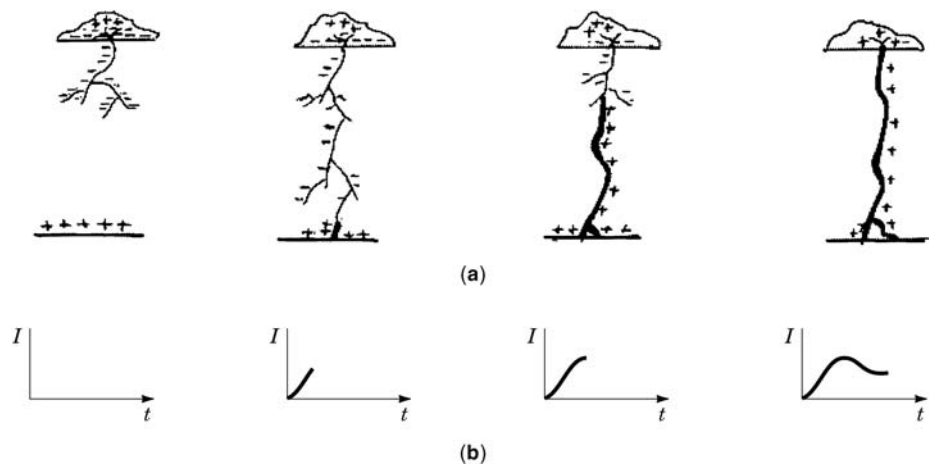


Figure 1. Direct lightning strokes. (a) Development of a lightning stroke. (b) Accompanied discharge current.

ambient air. The head of a cloud is positively charged, whereas its lower part is negatively charged (1). The process of water drops disintegration may be repeated, thus creating many charges. These charges produce potentials on the order of millions of volts within the cloud and voltage gradients on the earth surface on the order of a few tens or even hundreds of kilovolts per centimeter, depending on the amount of charge contained in the cloud and its height from ground surface. Under some favorable conditions, a direct lightning stroke may occur between the cloud and the earth's surface, buildings, or electrical installations including power transmission lines. The process of lightning strokes are rather complicated in their growth, and numerous theories are found in the literature (2,3). Direct lightning strokes are characterized by large and fast rising discharge currents. The amplitude of discharge current is on the order of few tens of kiloamperes and in severe cases may reach 200 kA. Its rise time is few microseconds with an average steepness of few kiloamperes per microsecond. The process of developing of lightning stroke is illustrated in Fig. 1. Direct lightning strokes may cause severe damage to buildings and power systems if they are not properly protected. Protection of buildings and similar structures is usually carried out by means of metal masts (Franklin rods). Power transmission lines are protected by earth wires strung between towers above the line conductors.

SWITCHING OVERVOLTAGES

Power systems are subjected to switching overvoltages during their operation. These voltages are classified into two main categories, according to the switching operation and the switched current. Table 1 shows the main switching operations in power systems.

Compared to lightning surges (impulses), switching surges are characterized by their long duration and lower magnitudes. Figure 2 shows the duration of lightning and switching over-

voltages as well as their magnitudes per unit of system working voltage.

GENERATION OF IMPULSE VOLTAGES

For test purposes, impulse voltages, simulating voltage surges from lightning or switching, are generated in high-voltage test laboratories.

An impulse voltage is a unidirectional voltage that rises to peak value in a short time and decays slowly to zero value. For the purpose of standardizing high-voltage tests, the impulse voltage shape should be precisely defined. Figure 3 shows an impulse voltage wave simulating a lightning surge. The wave shape, in general, is defined by its time to crest (wave front T_1) and its decay time to 50% of crest value (wave tail T_2).

Usually generated impulse voltages have a low rate of rise at their initial and final stages of buildup. On the other hand, a considerable percentage of the peak value builds up in a relatively short time compared to the total front time. To overcome this, the wave front and tail times are evaluated, as shown in Fig. 3. Impulse voltage waveshapes are defined by the values T_1 and T_2 in microseconds (T_1/T_2).

Impulse voltage in the range of 100 kV are generated by single-stage impulse generators and higher voltages, up to about 6 MV, are generated by multistage generators.

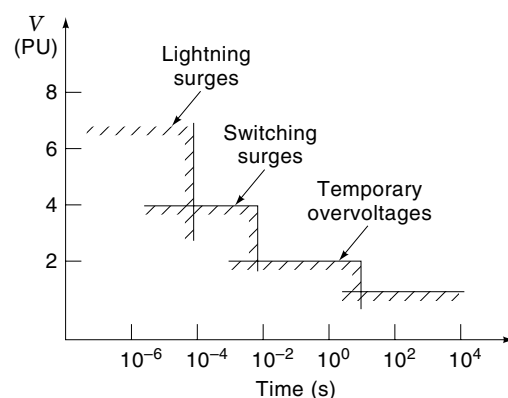


Figure 2. Magnitudes and duration of lightning and switching voltage surges.

Table 1. Switching Overvoltages

Temporary Overvoltages	Switching Overvoltages
Load rejections	Short circuit currents
Energization of unloaded lines	Inductive currents
Earth faults	Capacitive currents
Resonance effects	Short lines

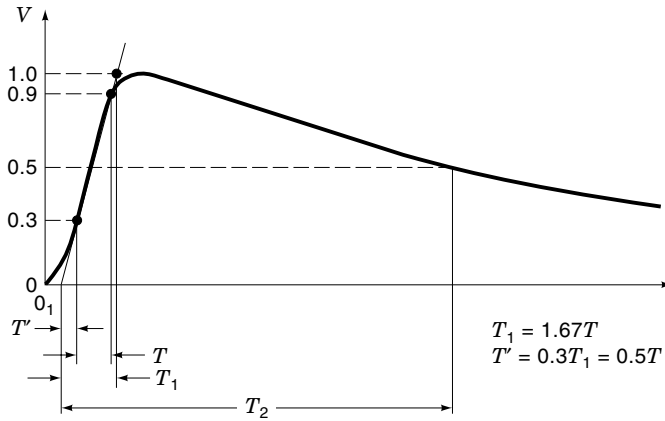


Figure 3. Lightning impulse voltage wave. Exact evaluation of wave front and tail times are shown.

Single-Stage Impulse Generators

Impulse voltages are simply generated by charging a condenser from a dc source and suddenly discharging it through a resistor by closing a fast-acting switch. In actual impulse generators, this switch is in the form of a sphere gap, controlled to close the circuit and start operation when required. Figure 4 shows a single-stage impulse generator. Construction of the impulse generator's control circuit (trigatron) is not shown; more details are given in Refs. 4 and 5. The voltage E is the input dc voltage usually obtained from a variable step-up high-voltage transformer with full-wave rectifier and smoothing circuit. The resistor R is for charging the stage capacitor C_1 ; R_1 and R_2 control the wave front and tail times, respectively; and C_2 is the test object capacitance. The output voltage $e(t)$ can be obtained by Laplace transform and simple circuit equations, which is in the form

$$e(t) = \frac{E}{R_1 C_2 (\alpha_2 - \alpha_1)} [\exp(-\alpha_1 t) - \exp(-\alpha_2 t)] \quad (1)$$

where

$$\alpha_1, \alpha_2 = \frac{a}{2} \mp \sqrt{\left(\frac{a}{2}\right)^2 - b} \quad (2)$$

$$a = \left(\frac{1}{R_1 C_1} + \frac{1}{R_1 C_2} + \frac{1}{R_2 C_1} \right)$$

$$b = \frac{1}{R_1 R_2 C_1 C_2}$$

Equation (1) is the difference between two exponential voltage waves decreasing with time. The first one decays slowly with

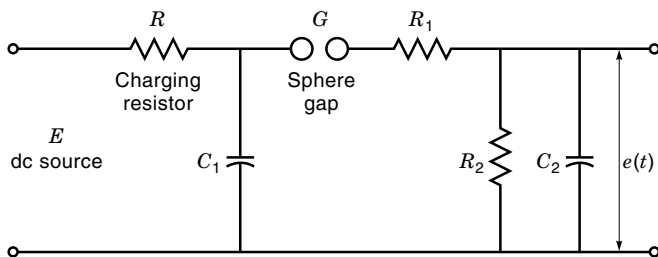


Figure 4. Single-stage impulse generator circuit.

time, and the second decays relatively faster. Examination of the two constants α_1 and α_2 in Eq. (2) shows that the values of R_1 and R_2 control the wave front and tail times. Complete details of other circuits arrangements, calculation of the wave front and tail times, and efficiency of the generator are fully given in Ref. 5.

In calculating the parameters of the impulse wave, no consideration was taken for the inherent inductance of the circuit elements and connecting leads. However, the value of this inductance has a great effect on its shape.

For the purpose of impulse voltage testing, standard impulse waves are recommended by different national and international standards. The most common impulse waves are 1.2/50 in Europe and 1/40 in North America.

Multistage Impulse Generators

For technical reasons, the dc charging voltage is usually limited to voltages around 200 kV. To obtain impulse voltages in the range of 6 MV multistage impulse generators are constructed. The basic idea of these generators is to charge several condensers in parallel from a moderate voltage dc source (say 200 kV) and then simultaneously discharging them in series through resistors. Figure 5 shows the circuit of a multistage generator in which an impulse voltage of approximately ten times the charging voltage is obtained. Full details of multistage impulse generators circuit, operation, and control are given in several high-voltage engineering textbooks (4–6).

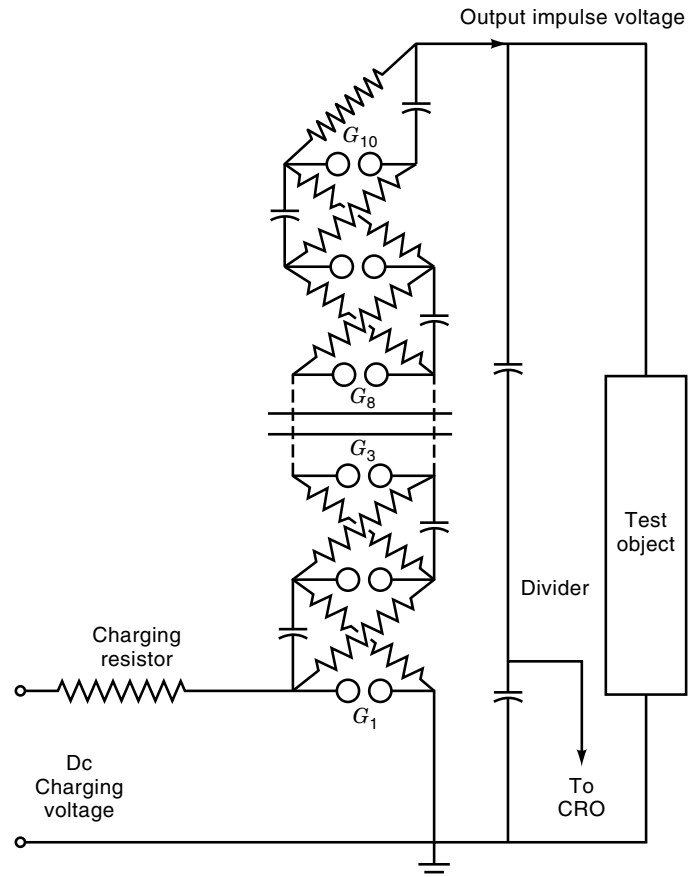


Figure 5. Multistage impulse generator circuit. Charging voltage = 200 kV, number of stages = 10, output impulse voltage $\approx$ 2 MV.

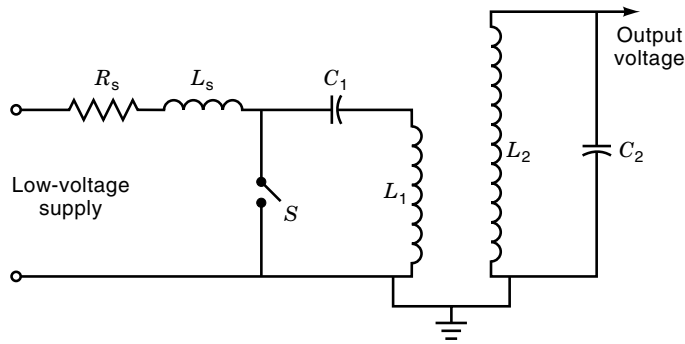


Figure 6. Switching impulse voltage generation by Tesla coil.

GENERATION OF SWITCHING SURGES

Switching voltage surges are characterized by their long front and tail times and high energy. The definition of the wave front and tail times is more or less similar to that of lightning impulse voltages (7). Switching impulse voltage wave shapes in common use for testing purposes have wave front and tail times on the order of several hundred microseconds (250/2500 and 500/2500 μ s). These voltages can be generated from normal impulse generators with some modifications in their circuit elements to control the wave front and tail times to the required values. Alternative methods for generating such voltages are by producing voltage oscillations in the low-voltage winding of a step-up transformer or coil (Tesla coil). These oscillations are then transferred to the high-voltage side. Adjustment of the low- and high-voltage side parameters (inductance and capacitance) of the transformer or coil circuit and the instant of switching operation on the low-voltage side control the generated voltage shape. Figure 6 shows the well-known circuit of a Tesla coil. Interested readers may refer to Refs. 4 and 5 for more details on circuit construction, operation, and merits and demerits of different systems.

IMPULSE VOLTAGE MEASUREMENT

Impulse voltage measurement needs special measuring equipment and techniques to give true measure of the peak value and wave shape. When measuring high-impulse voltages on the order of megavolts and high rising rates, the measuring process becomes rather complicated, and the accuracy is highly reduced. Whatever the method used to measure impulse voltages, tedious precautions should be taken to reduce measuring errors. These errors are mainly introduced by external interference and response of the measuring system to such single-event voltages.

The main measuring methods used in impulse testing and research work are (1) sphere gaps and (2) resistive or capacitive voltage dividers.

Sphere Gaps

Measurement of the peak value of impulse voltages by means of sphere gaps, especially at extra high voltages, is very common in impulse testing. They consist of two identical metallic spheres separated by an air gap. The basic principle of measurement by sphere gaps is that the air in the separating gap breaks

down in response to the peak value of the applied voltage, provided that its rise time is not less than 1 μ s. To obtain accurate measurement by sphere gaps, several precautions should be taken. These precautions include sphere diameter selection, separation, and mounting; surface conditions; and connection to the measured voltage and nearby objects. Standard sphere diameters in common use range between 6.25 cm and 200 cm.

Measurement of impulse voltages by sphere gaps is the simplest method, with an accuracy of $\pm 3\%$, provided all recommendations given by IEC (8) are strictly followed. In laboratory testing work, it is difficult to control the temperature and pressure to standard values (20°C and 101 kPa) because of the large space and size of equipment. A correction factor k is used to obtain actual values corresponding to ambient conditions. This factor depends on the relative air density δ , which is given by

$$\delta = \frac{P}{760} \frac{(273 + 20)}{(273 + t)} = \frac{0.386P}{273 + t} \quad (3)$$

where t and P are the ambient temperature and pressure in degrees Celsius and millimeters of mercury, respectively. If V_s is the breakdown voltage at standard conditions given in IEC (8), then the actual voltage measured is kV_s . The values of δ and k are very close to each other (8).

Resistive Dividers

Resistive voltage dividers are commonly used in measuring ac and dc high voltages, and their construction is relatively simple. When measuring high-impulse voltages on the order of megavolts having short rise time, these dividers need special construction to give a true picture of the measured signal in magnitude and shape. Theoretically, resistive dividers consist of a high resistance in series with a low-value resistor. The combination is connected directly on the test object with the low-voltage terminal earthed. The division of voltage between the high- and low-resistance sections of the divider is not simply in the ratio of their values. In ac and dc measurement by resistive dividers, the error is usually much less than when measuring impulse voltages. The difficulty with impulse voltage dividers is that their transfer characteristics introduce large errors in both magnitude and shape if they are not properly constructed and sized. More details on the transfer characteristics of voltage dividers will be presented later. The sample voltage signal on the low-voltage section of the divider can be measured and recorded by oscilloscopes, analog/digital conversion circuits, attenuators, or high-voltage probes. The accuracy of voltage measurement, of course, depends on the divider construction, electromagnetic interference from test equipment, connection leads, and effectiveness of the grounding system.

Capacitive Dividers

Similar to resistive dividers, capacitive dividers are constructed from a single capacitor unit or several units in series and connected across the test object. Construction of high-voltage capacitors needs high technical measures and experience. However, compressed gas (sulfur hexafluoride, SF₆) capacitors are now available with good performance characteristics. Again, the main difficulty with capacitive dividers is their transfer characteristics and electromagnetic interference. It is obvious that the measured voltage signal across the low-voltage section

of the divider is not simply proportional to the inverse ratio of the divider capacitances. This is true because the equivalent circuit of the divider involves stray capacitances, inherent inductances, and resistors either actual or representing losses in the capacitors dielectric.

Impulse Currents

Lightning and voltage surges (impulses) are often accompanied by current surges that flow in a piece of equipment or a system subjected to such surges. Thus, impulse currents are generated in high-voltage laboratories for testing purposes. Surge arresters, used for protecting high-voltage equipment, are tested with impulse currents to verify their performance characteristics.

Similar to lightning and switching voltage impulses, two current wave shapes are used for testing electrical equipments. Figure 7(a) shows the wave shape recommended by IEC for testing under lightning impulse currents. Again, the wave shape is defined by the wave front and tail times in microseconds (T_1/T_2). The standard lightning impulse currents used in testing purposes are 1/20, 4/10, 8/20, and 30/80. Figure 7(b) shows the wave shape, recommended by IEC, for testing under switching impulse current. The wave shape is nearly rectangular in shape. The two time parameters defining the current wave T_d and T_t are shown in Fig. 7(b). Standard rectangular impulse currents have a time duration T_d of 500, 1000, or 2000 μ s.

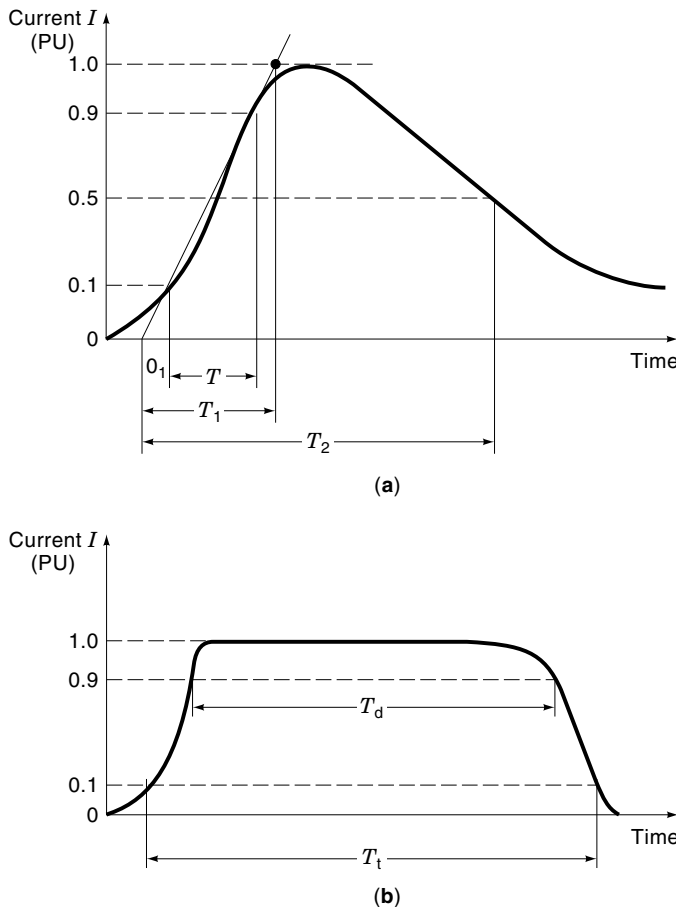


Figure 7. Impulse current wave shapes: (a) exponential and (b) rectangular.

Measurement of Impulse Current

Fast-rising high-impulse currents up to few hundreds kiloamperes (200 kA) are usually encountered in testing surge arresters and lightning discharges. The main requirements from the measuring system are: (1) accuracy, (2) fast response, (3) simple construction, and (4) size.

Measurement of impulse currents relies on its basic effects; voltage drops when passing in a resistive element, associated magnetic field, and optical phenomenon. Resistive shunts are commonly used for high-impulse current measurement having ohmic resistance on the order of few milliohms. The voltage drop across the shunt (a few hundred volts) is measured by any voltage-measuring device such as a cathode-ray oscilloscope (CRO). The difficulties in designing resistive shunts are their inherent inductance, which should be minimal; the voltage across the shunts, which should be on the order of few hundreds volts; and their volume, which should be capable of absorbing the energy associated with the measured current without the temperature exceeding a permissible rise (20°C). Different designs and performance characteristics are given in Refs. 9 and 10. Methods based on the magnetic effect of electric currents use a single-turn coil placed perpendicular to the magnetic field. The induced voltage in the coil is a function of the measured current in magnitude and shape. The basic equations governing the measured current and induced voltage in the coil are

$$\oint_c H dl = I \quad (4)$$

$$\int \int B ds = \phi \quad (5)$$

$$e = -\frac{d\phi}{dt} \quad (6)$$

$$e = -\mu \frac{r^2}{R_o} \frac{di(t)}{dt} \quad (7)$$

where H is the magnetic field, ϕ is the flux linking the coil, r is the radius of the coil, and R_o is the distance between the coil center and the current-carrying conductor.

Pulse current transformers are sometimes used to measure impulse currents. However, their accuracy, lowpass band (a few kilohertz), and saturation problems limited their applications.

Optical methods used to measure impulse currents are based on two facts. The first is that, when a current flows through a light emitting diode, it generates light that depends on the current shape and magnitude. This light can be detected and transformed by photodetectors to an electrical signal, which is taken as a measure of the current (11). The second is that light beams show rotation when subjected to magnetic fields. Again, photodetectors are used to transfer this rotation to an electrical signal (12) proportional to the measured current.

ELECTRICAL DISCHARGE

Insulation media in electrical equipments may be a liquid, solid, gas, or combination of them. Gases are used to insulate electrical equipments such as circuit breakers, high-voltage capacitors, and other equipment. Gases, when compressed, are

excellent insulants and superior to other media. Under natural conditions, a gas constitutes neutral molecules and few free electrons (negative charge carriers) produced by natural sources such as cosmic rays. When this gas is subjected to a uniform electric field between two electrodes, the free electrons gain energy and move toward the positive electrode (anode). On their way to the anode, they collide randomly with the neutral gas molecules. If they have enough energy, they may ionize these molecules, thus producing new electrons and positive ions. The new electrons and the initiatory ones in their way to the anode may produce additional electrons. This process is a cumulative one, and a large number of electrons form an "electron avalanche" heading toward the anode. In the mean time, positive ions travel slowly (because of their relative large mass) toward the cathode. This process of ionization is called the Townsend primary ionization process. The flow of electrons between the electrodes is a flow of charge or a current. This current increases with the increase of electric field between the electrodes. Many other ionization and de-ionization processes are involved in gaseous discharge. How these processes occur, the factors governing their contribution to the electrical discharge, and the equations quantifying their contribution is a very broad subject, which is covered in a huge number of articles and reference books (2,4,5,13). The basic equation for current growth is

$$i = i_0 \frac{e^{\alpha d}}{1 - \beta(e^{\alpha d} - 1)} \quad (8)$$

where i is the current flowing between the electrodes, i_0 is the initiatory primary current, α is Townsend first ionization coefficient, β is Townsend secondary ionization coefficient, and d is the electrode spacing. Equation (8) does not take into consideration all secondary ionization and deionization processes such as photoionization in the gas and at the cathode surface, ionization by metastable atoms, thermal ionization, diffusion, and attachment. However, Eq. (8) was modified to be in the form

$$i = i_0 \frac{\frac{\alpha}{(\alpha - \eta)} e^{(\alpha - \eta)d} - \frac{\eta}{\alpha - \eta}}{1 - \frac{\gamma\alpha}{(\alpha - \eta)} [e^{(\alpha - \eta)d} - 1]} \quad (9)$$

where γ is the secondary ionization coefficient and may involve more than one secondary process and η is the attachment coefficient.

The coefficients α , η , and γ in Eq. (9) are functions of more than one of the parameters evolved in gaseous discharges (electric field, gas, electrode geometry, cathode material, and gas pressure and temperature).

The condition of breakdown in the gas is assumed when the current flowing between the electrodes is theoretically equal to infinity. This occurs when

$$\frac{\gamma\alpha}{\alpha - \eta} \{ \exp[(\alpha - \eta)d_s] - 1 \} = 1 \quad (10)$$

where d_s is the electrode separation at breakdown.

Electric discharge leading to breakdown as given by Townsend's equations has many limitations in explaining how breakdown occurs in long gaps and nonuniform field gaps resulting from the very short time observed during their breakdown. The streamer theory was then developed to account for

such limitations. In brief, the streamer theory is based on the fact that for breakdown to occur in a time on the order of 10^{-6} s to 10^{-8} s, there must be a buildup of a high electric field in the gap to accelerate the breakdown process. This field was attributed to the accumulation of the slow moving positive ions in the gap, thus leading to an external space charge field. Photoionization in the gas was also considered as an additional accelerating factor for breakdown. Several explanations and equations have been developed by Meek, Loeb, and Raether on the theory of streamer breakdown (2,3).

In nonuniform field electrode geometries (e.g., point to plane, parallel wires, and coaxial wire-cylinder), the electric field is highly localized at the point or the wires while it is relatively low at other places. This high field may start discharge at and around these electrodes while the rest of the gap is not. This phenomenon is known as Corona discharge, and it is characterized by its pale violet light, smell of ozone, and hissing noise. Corona represent continuous power losses on transmission lines, and its presence in electrical equipment should be avoided because it may lead to their failure.

IMPULSE BREAKDOWN

Breakdown under impulse voltages is different from those under ac and dc because of their short duration time. For breakdown to start, at least one electron should exist in the gap at a favorable location and time. Alternating (60 Hz or 50 Hz) and direct voltages can be considered constant during the buildup of the breakdown process. The probability of the appearance of an electron or electrons in the gap is high, and that explains the small scatter in the breakdown voltages under ac and dc. On the other hand, the appearance of an electron away from the peak of the impulse voltage has very low probability to initiate breakdown. Figure 8 shows how breakdown occurs under impulse voltages. If an electron appears at time t_1 or t_4 corresponding to voltages V_1 and V_4 on the wave front and tail, respectively, no breakdown can occur because the voltages at these instants are less than the static breakdown voltage V_s . For discharge to start and develop to breakdown, the initiatory electron should appear within a time between t_2 and t_3 , where the impulse voltage exceed V_s . The time interval between t_2 and t_b is divided into two time intervals:

1. Statistical time lag T_s , which is defined as the time that elapses from the instant the impulse voltage reaches the

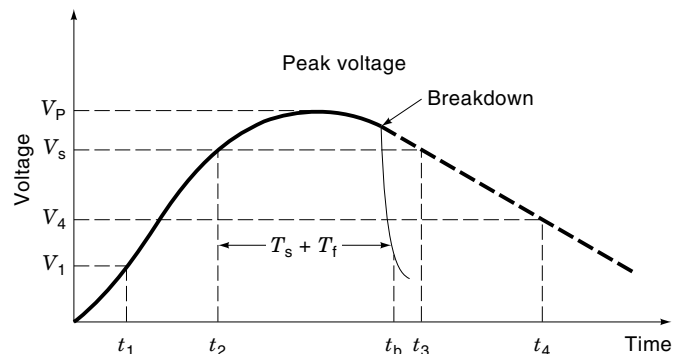


Figure 8. Breakdown under impulse voltage.

static breakdown value until an initiatory electron appears at a suitable location in the gap to start discharge.

2. Formative time lag T_f , which is defined as the time that elapses between discharge initiation until breakdown develops.

The statistical time lag is relatively long and is in the range of 10 s to 10^{-3} s. Short formative time lags in the 10^{-9} s range were recorded (5).

IMPULSE BREAKDOWN STRENGTH

When testing an insulation system under impulse voltage, large scatter in the results is usually observed. The probability of breakdown under impulse voltage depends on its peak value and duration time. A probability breakdown curve is usually drawn to define the impulse breakdown. Figure 9 shows the breakdown probability characteristics of a test object (e.g., an insulator string). This curve can be obtained by applying a certain number of shots for each voltage level and determining the ratio between the number of breakdowns and the total shots. This ratio is the probability of breakdown for a specified impulse voltage. It is clear that the breakdown probability increases with the increase of the applied voltage relative to the static breakdown voltage of the test object. To obtain representative values of the breakdown probability, large number of voltage shots (say 100 shots) at each voltage level are applied. In practical testing, this number is limited to 15 or 20 shots. Figure 9 shows two distinct voltages V_0 and V_{100} corresponding to zero and 100% breakdown probability. The question now arises which value is taken to represent the impulse breakdown voltage? The answer to this question is that it is the voltage corresponding to 50% breakdown probability and it is termed critical flash over voltage (CFO). The value of V_0 and V_{100} are of importance to the insulation design and insulation coordination, respectively. Exact determination of V_0 and V_{100} from Fig. 9 needs the application of large number of shots, which is time-consuming. In practice, more realistic procedures are used. These procedures depend on the fact that the breakdown probability function is Gaussian; interested readers may refer to Refs. 4 and 5.

TRANSFER CHARACTERISTICS

In measuring and recording impulse voltages by means of voltage dividers and CRO, errors as high as 10% or even higher are usually observed. The errors are not only in magnitude but

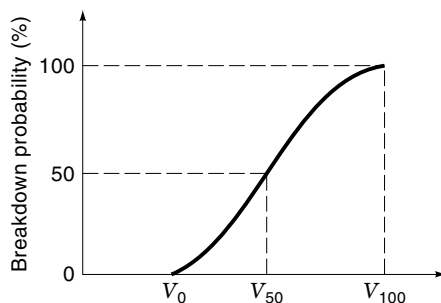


Figure 9. Impulse voltage breakdown probability.

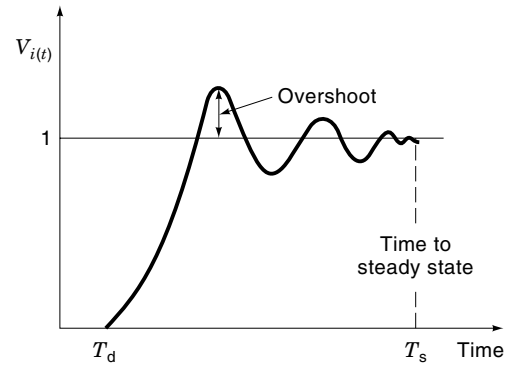


Figure 10. Response of voltage dividers to unit step function voltage.

also in phase. Whatever the equivalent circuit of voltage dividers is, they can be represented by a four-terminal network, and their transfer characteristics (output voltage/input voltage, v_o/v_i) can be evaluated. Practical impulse voltage shapes are the difference between two exponential voltages. The mathematical analysis for such waves is very lengthy and tedious. Checking the response of voltage dividers by applying a unit step function is recommended by IEC (14). If a unit step function in the form

$$v_i(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases} \quad (11)$$

is applied to a voltage divider, the output response will be $v_o(t)$. The response to another input voltage can be derived by Laplace transform, and hence the output in the time domain is obtained (4). Figure 10 illustrates how the output voltage of a divider can differ from the input in magnitude and shape. Good design of voltage dividers should provide (1) short delay time T_d , (2) low overshoot, and (3) short time to steady state.

In recording impulse voltage waves, the output of the divider is taken via coaxial cables to a CRO. Unless proper matching is provided, errors resulting from voltage reflections are introduced.

MEASUREMENT STANDARDS

All standard specifications for impulse voltage and current limit measurement accuracy. Table 2 shows the standard accuracy limits for impulse voltage and current measurement as well as acceptable tolerances in their generation (7).

Table 2. Standard Impulse Voltage and Current Tolerances and Measurement Accuracy

Quantity	Lightning Voltage (%)		Switching Voltage (%)		Impulse Current (%)	
	Tol.	Accu.	Tol.	Accu.	Tol.	Accu.
Peak value (full impulse)	±3	±3	±3	±3	±10	±3
Peak value (chopped)						
on front		≤±5				
on tail		≤±3				
Front time	±30	±10	±20	±10	±10	±10
Tail time	±20	±10	±60	±10	±10	±10

CALIBRATION

Regular calibration of the measuring system against a standard measuring device is a common practice in impulse testing and research work. The sphere gap is an acceptable standard measuring device for measuring the peak value of an impulse voltage with an accuracy of $\pm 3\%$. When recording an impulse wave, the recording system should be calibrated against a reference measuring system. The accuracy of the reference system ranges between ± 1 and $\pm 3\%$ for measuring the peak voltage or current and $\pm 5\%$ for measuring the time parameters (14).

APPLICATIONS

Electrical equipment is subjected to voltage and current surges during operation, and impulse testing is one of the main tests specified by all standard specifications. Circuit breakers, cables, transformers, high-voltage capacitors, switchboards, transmission line insulators, surge arresters, and other kinds of equipment are tested under impulse voltages and/or current. The main purpose of impulse tests is to validate the performance characteristics, proper design, and manufacture quality. Impulse voltages and currents have several applications other than for testing. They can be used for triggering electronic circuits and components. Electrostatic precipitators, used to collect dust particles from the exhaust of chimneys, use high-impulse voltages in their operation. The basic construction of such devices consists of a thin wire concentric with a metallic cylinder. When a high voltage is applied between the wire and the cylinder, corona discharge starts at and around the wire. The flow of charges between the wire and cylinder charges the particles, and they move toward the cylinder wall where they are collected. It is claimed that using impulse voltages instead of dc is more effective in collecting dust particles. In cars, an inductive electric circuit, supplied from a 12 V battery, is regularly switched off and on to produce impulse voltages of several kilovolts. This voltage is applied to the car's plugs to produce sparks, thus igniting the fuel.

Low-voltage impulse signals are used for fault detection and location in underground cables. This is based on the fact that when a voltage surge is applied at one end of a long conductor in air it travels to the other end with light velocity ($c = 3 \times 10^8$ m/s). In the case of underground cables, it travels with velocity $c/\sqrt{\epsilon_r}$, where ϵ_r is the relative permittivity of the cable insulating material. If there is a fault in the cable, between its ends, the voltage signal is reflected at the fault. The reflected signal is received at the voltage source on a recording device, such as a CRO. Knowing the velocity of the voltage signal and the time taken to reach the fault and return back to the source, one can then estimate distance. The reflected voltage signal may be positive or negative depending on whether the fault resides in an open cable's conductor or insulation. By this method fault location is determined with high accuracy and the type of fault is known, thus reducing the effort and time of digging and repairing the fault. Cable's fault detectors are now available in the marketplace and can detect and locate faults without any human calculations.

Fast-rising impulse voltages of a few tens of kilovolts are used to generate ozone for water treatment. Ozone generators are similar to electrostatic precipitators in their basic construction. Corona discharge is initiated by applying a repetitive voltage pulse to a thin wire in a duct where air flows axially or radially. Corona discharge starts at and around the wire surface, due to the high electric field therein, and consequently ozone is formed. The amount of generated ozone depends on the magnitude of the applied impulse voltage, its repetition rate, duration, wire diameter, and rate of air flow.

Lightning protection from thunderstorms is usually carried out by means of metallic rods (Franklin rods) installed on or around the protected objects. These rods have a protection zone where they are highly effective, and beyond it the degree of protection is reduced. To increase the protection zone and degree of protection and to decrease the height of the rod, the electric field at the tip of the protecting rod should be increased to help direct the lightning stroke to it. The field at the tip of the rod can be increased by applying to it a repetitive impulse voltage of a few tens of kilovolts from an external source.

BIBLIOGRAPHY

1. J. D. Cobine, *Gaseous Conductors*, 2nd ed., New York: Dover, 1958.
2. J. M. Meek and J. D. C. Craggs, *Electrical Breakdown of Gases*, Oxford: Clarendon Press, 1953.
3. R. H. Golde (ed.), *Physics of Lightning*, London: Academic Press, 1977.
4. E. M. Khalifa (ed.), *High-Voltage Engineering: Theory and Practice*, New York: Marcel Dekker, 1990.
5. E. Kuffel and W. S. Zaengl, *High Voltage Engineering: Fundamentals*, New York: Pergamon, 1984.
6. W. Hawley, *Impulse-Voltage Testing*, London: Chapman & Hall, 1959.
7. IEC Publication 60-1, *High-Voltage Test Techniques*, 2nd ed., Geneva: International Electrotechnical Commission, 1989.
8. IEC Publication 52, *Recommendations for Voltage Measurement by Means of Sphere-Gaps*, Geneva: International Electrotechnical Commission, 1960.
9. R. Malewski, Wirewound shunts for measurement of fast current impulses, *IEEE Trans. Power Appar. Syst.*, **PAS 103**: 2927–2933, 1984.
10. O. B. Oliveira, M. T. Silva, and F. A. Chagas, Design, construction, and response evaluation of shunts for the measurements of high impulse current, *7th Int. Symp. High-Voltage Eng.*, Dresden, 1991.
11. E. A. Ulmer, A high-accuracy optical current transducer for electric power systems, *IEEE Trans. Power Deliv.*, **5**: 892–898, 1990.
12. A. J. Schwab, *High-Voltage Measurement Techniques*, Cambridge, MA: MIT Press, 1972.
13. E. Nasser, *Fundamentals of Gaseous Ionization and Plasma Electronics*, New York: Wiley, 1971.
14. IEC Publication 60-2, *High-Voltage Test Techniques*, 2nd ed., Geneva: International Electrotechnical Commission, 1994.

ROUSHDY M. RADWAN
Cairo University