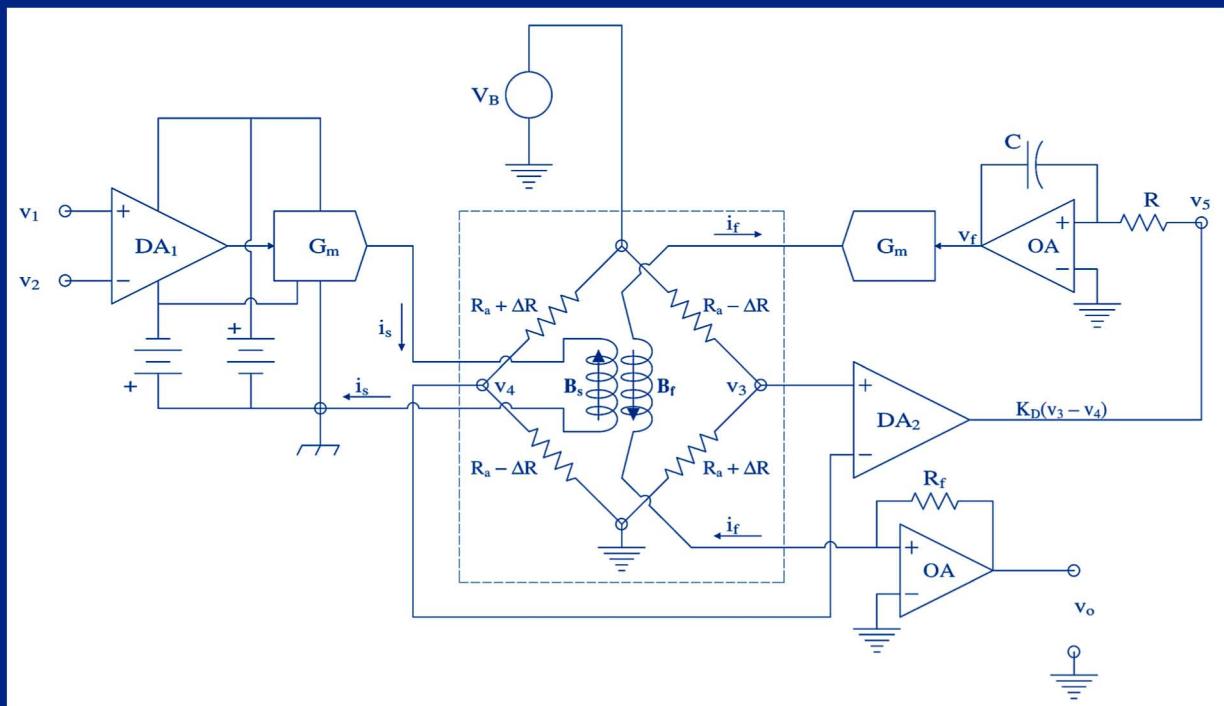


Introduction to INSTRUMENTATION AND MEASUREMENTS

Second Edition



Robert B. Northrop

Introduction to INSTRUMENTATION AND MEASUREMENTS

Second Edition

Introduction to INSTRUMENTATION AND MEASUREMENTS

Second Edition

Robert B. Northrop



Taylor & Francis
Taylor & Francis Group

Boca Raton London New York Singapore

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2005 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20110713

International Standard Book Number-13: 978-1-4200-5785-0 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

*I dedicate this second edition of IIM to my wife and daughters,
Adelaide, Anne, Kate and Victoria,
who have always believed in my writing.*

Preface

Purpose: This text is intended to be used in a classroom course for engineers that covers the theory and art of modern instrumentation and measurements (I&M). There is more than enough material to support two semesters' work. Thus, the instructor has the option of choosing those topics and the depth of coverage that suit his or her interests and curriculum. Due to its breadth, *Introduction to Instrumentation and Measurements, 2nd edition* will also be useful as a reference for the practicing engineer and scientist interested in I&M.

Why have a classroom course in I&M? Over the past decade or so, in the U.S., many EE departments have discontinued classroom courses on the theory and practice of instrumentation and measurements. In this period, we have also seen the swift development of new and exciting means of measurement using new technologies, the adoption of new standards and, concurrently, the lack of development of a coherent educational base to support their understanding and use. Using an instrument in the laboratory is not the same as understanding the physical and electronic principles underlying its design and functional limitations. Clearly, there is now more than ever a need for classroom experience in the new I&M that will give students the necessary technical background to use and design sensors, signal conditioning systems and I&M systems. We feel that this text supports that need.

This text was written based on the author's 40 years of experience in teaching a classroom course (EE 230), Electrical Instrumentation, to juniors and seniors in the Electrical and Computer Engineering Department at the University of Connecticut, Storrs.

Obviously, in 40 years we have seen the field of instrumentation and measurements evolve with the rest of electrical engineering technology. Due to the rapid pace of technical development, it has generally been difficult to find an up-to-date text for our Electrical Instrumentation course. After years of frustration trying to match a text to course content, I decided to write one that would not only encompass the "traditional" aspects of I&M, but also include material on modern IC and photonic sensors, microsensors, signal conditioning, noise, data interfaces and DSP.

Reader Background: Readers are assumed to have taken core EE curriculum courses, or their equivalents. The reader should be skilled in basic *linear circuit theory* (i.e., the reader has mastered Thevenin's and Norton's theorems, Kirchoff's Laws, superposition, dependent sources, and ideal op-amps, and should know how to describe DC and AC steady-state circuits in terms of loop and node equations). An introductory systems course should have given him/her familiarity with both time and frequency domain methods of describing linear dynamic systems characterized by ordinary, linear, differential or difference equations, including state space, Fourier, Laplace and z-transforms, transfer functions, steady-state frequency response of systems, and Bode plots. From physics or an EE course in electromagnetics, the reader should have a basic knowledge of electric and magnetic fields, inductance, capacitance, reluctance, etc. There should also be some familiarity with electromagnetic waves, Maxwell's equations, transmission lines

and polarization. From a first course in electronics, there should be basic knowledge of BJTs, JFETs, diodes, photodiodes and their simple linear circuit models.

Scope of the Text: A major feature of *Introduction to Instrumentation and Measurements*, 2nd edition is its breadth of coverage. Throughout the text, a high level of mathematical analytical detail is maintained. It is not a "picture book"; we assume that readers have already had contact with basic electrical instruments, including oscilloscopes and meters in their introductory EE and physics labs.

In the following paragraphs, we give an overview of the contents.

Chapter 1, "Measurement Systems," is introductory in nature. In it, we illustrate measurement system architecture and describe sensor dynamics, signal conditioning, data display and storage. Errors in measurements are discussed, including the meaning of accuracy and precision, limiting error, etc. The recent (1990) quantum standards adopted for the volt and the ohm are described, as well as other modern electrical and physical standards.

In Chapter 2, "Analog Signal Conditioning," we describe, largely at the systems level, the means of conditioning the analog outputs of various sensors. Op-amps, differential, instrumentation, auto-zero and isolation amplifiers are covered. Applications of op-amps in active filters, differential instrumentation amplifiers, charge amplifiers, phase sensitive rectifiers, etc. are shown. We also give practical considerations of errors caused by offset voltage, bias currents, input impedance, slew rate and gain bandwidth product etc. There is also a section on nonlinear signal processing with op-amps.

Noise and coherent interference in measurements are treated in depth in Chapter 3. A heuristic yet rigorous approach is used in which we define and use one-sided, noise voltage and current power density spectra to describe the effect of noise in instruments and measurement systems. Noise factor and figure are covered, and output signal-to-noise ratios are used to evaluate system noise performance. Examples are given of calculations of the noise-limited resolution of the quantity under measurement (QUM). Techniques are shown for the minimization of coherent interference.

The traditional topics of DC null measurements and AC null measurements are presented in Chapter 4 and Chapter 5, respectively. Wheatstone and Kelvin bridges, and potentiometers are described in Chapter 4, and the major AC bridges used to measure the inductance, Q , and capacitance, D , are treated in Chapter 5. New material added to this chapter includes a description and analysis of the Anderson Current Loop method of reading sensor outputs.

A survey of sensor mechanisms is presented in Chapter 6. This is a large and substantive chapter covering a broad range of sensor mechanisms and types. Of special note is the introduction of certain fiber optic and electro-optic sensors, as well as selected chemical and ionizing radiation sensors. The Sagnac effect is introduced and the basic fiber optic gyro is described.

New material in Chapter 6 includes a description and analysis of sensors based on the giant magnetoresistive effect and the anisotropic magnetoresistive effect. Pyroelectric IR sensors are also introduced. The various means of measuring the rotation of linearly polarized light is presented, as well as a substantive section on photomultiplier tubes and channel-plate photomultipliers. Finally, a new section has been added on electronic noses which are used to sense volatile organic compounds.

In Chapter 7, "Applications of Sensors to Physical Measurements," a detailed analysis of mechanical gyroscopes, clinometers and accelerometers is given, including new material on micromachined accelerometers and gyros. The Doppler effect in ultrasonic velocimetry and laser Doppler velocimetry are covered. Also new in Chapter 7 is a large section on the global positioning (GPS) system, a section on optical interferometry, and

an extensive introduction to spectrophotometry, sonoluminescence and surface plasmon resonance which are used for substance detection.

In Chapter 8, "Basic Electrical Measurements," the classic means of measuring electrical quantities are presented, as well as newer methods such as Faraday magneto-optic ammeters and Hall effect gaussmeters and wattmeters. Electronic means of measuring stored charge and static electric fields are described.

Digital interfaces on measurement systems are covered in Chapter 9. This chapter begins with a description of the sampling theorem, aliasing and quantization. The traditional topics of hold circuits, DACs and many types of ADC are covered. Also covered in Chapter 9 are data buses. New material includes a section on dithering as a means of reducing quantization noise, a section on delta-sigma ADCs and a section on the new USB. Virtual instruments and PXI systems are also introduced.

Since digitized, measured data is processed and stored on computers in modern instrumentation practice, Chapter 10, "Introduction to Digital Signal Conditioning," was written to acquaint the reader to this specialized field. The z-transform and its use in describing filtering operations on discrete, digitized data in the frequency domain is introduced. Examples of FIR and IIR digital filters are given, including numerical integration and differentiation routines, viewed both in the time and frequency domains. The discrete and fast Fourier transforms are covered and the effect of data windows on spectral resolution is discussed. Finally, the use of splines in interpolating discrete data sequences and estimating missing data points is described.

In Chapter 11, "Examples of the Design of Measurement Systems," four examples of complex measurement systems developed by the author and his students are given to illustrate design philosophy:

1. A self-nulling microdergree polarimeter to measure glucose concentration
2. A system to detect and locate partial discharges on underground, high-voltage power cables
3. Design of a laser velocity and distance measuring system
4. Design of capacitance sensors to detect hidden objects

Problems: Chapters 1 through 10 are followed by problems taken from the author's extensive classroom experience in teaching courses in instrumentation and measurement at the University of Connecticut. The problems are solvable; they are student-tested.

References and Bibliography: The references cited encompass a wide time span; from the 1950s to the present. There are many recent entries of review articles and specialized texts that should lead the reader interested in pursuing a specialized area of I&M further into that particular field.

Index: A complete index allows the reader to access topics, both featured and not featured in the Contents.

Features: Every chapter in the second edition of *Introduction to Instrumentation and Measurements* has been revised to reflect modern technology. In addition, many chapters contain all-new material which expands the scope of the text to include geophysical, chemical and photonic instrumentation. Some of this unique new material includes:

1. The Anderson Current Loop technology for conditioning the outputs of remote resistive and capacitive sensors (Chapter 4)
2. The design of optical polarimeters and their application to polarization responding sensors (Chapter 6)

3. Photonic measurements with photomultipliers and channel plate photon sensors (Chapter 6)
4. Introduction of the sensing of gas phase analytes; the vertebrate olfactory system is described, as well as various chemical sensors, and feature extraction of complex odorants (Chapter 6)
5. The Sagnac effect as a means of measuring vehicle angular velocity (Chapter 6)
6. Micromachined, vibrating mass and vibrating disk rate gyros. The Humphrey air jet gyro is also analyzed. Traditional pendulum, as well as fluid-filled clinometers, is described. Micromachined, IC accelerometers are also covered (Chapter 7)
7. Global Positioning System (GPS) and its various modifications to improve its accuracy (Chapter 7)
8. Substance detection using photons (Chapter 7); dispersive, non-dispersive, and Fourier transform spectroscopy are described, as well as sonoluminescence and surface plasmon resonance
9. Dithering, delta-sigma ADCs, data acquisition cards, the USB, and virtual instruments and PXI systems (Chapter 9)

The Author

Robert B. Northrop was born in White Plains, New York in 1935. After graduating from Staples High School in Westport, Connecticut, he majored in electrical engineering at MIT, graduating with a bachelor's degree in 1956. At the University of Connecticut, he received a master's degree in control engineering in 1958. As the result of a long-standing interest in physiology and biophysics, he entered a Ph.D. program at UCONN in physiology, doing research on the neuromuscular physiology of molluscan catch muscles. He received his Ph.D. in 1964.

In 1963, he rejoined the UCONN EE Department as a lecturer, and was hired as an assistant professor of EE in 1964. In collaboration with his Ph.D. advisor, Dr. Edward G. Boettiger, he secured a 5-year training grant in 1965 from NIGMS (NIH), and started one of the first interdisciplinary Biomedical Engineering graduate training programs in New England. UCONN currently awards M.S. and Ph.D. degrees in this field of study, as well as a B.S. specialization in BME.

Throughout his career, Dr. Northrop's areas of research have been broad and interdisciplinary and have been centered around biomedical engineering. He has done sponsored research on the neurophysiology of insect and frog vision and devised theoretical models for visual neural signal processing. He also conducted sponsored research on electrofishing and developed, in collaboration with Northeast Utilities, effective working systems for fish guidance and control in hydroelectric plant waterways on the Connecticut River using underwater electric fields.

Still another area of his sponsored research has been in the design and simulation of nonlinear, adaptive, digital controllers to regulate *in vivo* drug concentrations or physiological parameters, such as pain, blood pressure, or blood glucose in diabetics. An outgrowth of this research led to his development of mathematical models for the dynamics of the human immune system, which were used to investigate theoretical therapies for autoimmune diseases, cancer, and HIV infection.

Biomedical instrumentation has been another active research area: An NIH grant supported studies on the use of the ocular pulse to detect obstructions in the carotid arteries. Minute pulsations of the cornea from arterial circulation in the eyeball were sensed using a no-touch, phase-locked ultrasound technique. Ocular pulse waveforms were shown to be related to cerebral blood flow in rabbits and humans.

Most recently, Dr. Northrop has addressed the problem of noninvasive blood glucose measurement for diabetics. Starting with a Phase I SBIR grant, he developed a means of estimating blood glucose by reflecting a beam of polarized light off the front surface of the lens of the eye and measuring the very small optical rotation resulting from glucose in the aqueous humor, which in turn is proportional to blood glucose. As an offshoot of techniques developed in micropolarimetry, he developed a magnetic sample chamber for glucose measurement in biotechnology applications. The water solvent was used as the Faraday optical medium.

He has written eight textbooks on the subjects of analog electronic circuits, instrumentation and measurements, physiological control systems, neural modeling, signals and systems analysis in biomedical engineering instrumentation and measurements

in noninvasive medical diagnosis and analysis, application of analog electronic circuits to biomedical instrumentation, and a second edition of introduction to instrumentation and measurements. All except the first text have been published by CRC Press.

Dr. Northrop was on the Electrical & Computer Engineering faculty at UCONN until his retirement in June 1997. Throughout this time, he was program director of the Biomedical Engineering Graduate Program. He is a member of Sigma Xi, Eta Kappa Nu, Phi Kappa Phi, and Tau Beta Pi. As Emeritus Professor, he still teaches courses in biomedical engineering, writes, sails, and travels. He lives in Chaplin, Connecticut, with his wife, a cat, and a smooth fox terrier.

14 April 2005

Contents

Preface.....	vii
The Author.....	xi

Chapter 1 Measurement Systems

1.1 Introduction	1
1.2 Measurement System Architecture	1
1.2.1 Sensor Dynamics	3
1.2.2 Overview of Signal Conditioning.....	6
1.3 Errors in Measurements	7
1.3.1 Gross Errors.....	7
1.3.2 System Errors	7
1.4 Standards Used in Measurements.....	13
1.4.1 Electrical Standards.....	13
1.4.1.1 Potential Difference.....	14
1.4.1.2 Resistance.....	18
1.4.1.3 Current and Charge	23
1.4.1.4 Capacitance.....	26
1.4.1.5 Inductance	28
1.4.2 Time and Frequency	30
1.4.3 Physical Standards	31
1.4.3.1 Mass	31
1.4.3.2 Length.....	32
1.4.3.3 Temperature	33
1.4.3.4 The SI Base Units	33
1.5 Chapter Summary	34
Problems.....	34

Chapter 2 Analog Signal Conditioning

2.1 Introduction	39
2.2 Differential Amplifiers.....	39
2.2.1 Analysis of Differential Amplifiers.....	40
2.2.2 Common-Mode Rejection Ratio.....	41
2.2.3 Measurement of CMRR, A_D and A_C	41
2.2.4 Effect of Source Resistance Asymmetry on CMRR.....	42
2.3 Operational Amplifiers.....	44
2.3.1 Types of Op-amps	45
2.3.2 Basic Broadband Amplifier Design Using Op-amps.....	47
2.3.2.1 Non-Inverting Amplifier.....	48
2.3.2.2 The Inverting Amplifier and Summer	49

2.3.3	Current Feedback Op-amps	50
2.4	Analog Active Filter Applications Using Conventional Op-amps.....	54
2.4.1	Introduction.....	54
2.4.2	Analog Active Filter Architectures	55
2.4.2.1	Controlled Source Active Filters.....	56
2.4.2.2	Biquad Active Filters	58
2.4.2.3	Generalized Impedance Converter Active Filters	62
2.4.2.4	High-Order Active Filters.....	64
2.4.3	Operational Amplifier Integrators and Differentiators	66
2.4.4	Summary.....	67
2.5	Instrumentation Amplifiers	68
2.5.1	Instrumentation Amplifiers Made from Op-amps.....	68
2.5.2	Isolation Amplifiers	71
2.5.3	AutoZero Amplifiers	73
2.5.4	Absolute Isolation	74
2.5.5	Summary.....	74
2.6	Nonlinear Analog Signal Processing by Op-amps and by Special Function Modules.....	77
2.6.1	Introduction.....	77
2.6.2	Precision Absolute Value (Absval) Circuits.....	78
2.6.3	Multifunction Converters	80
2.6.4	True RMS to DC Converters	81
2.6.5	Square Root Circuits and Dividers	83
2.6.6	Peak Detectors and Track and Hold (T&H) Circuits.....	85
2.6.7	Log Ratio and Trigonometric ICs	88
2.6.8	Summary.....	90
2.7	Charge Amplifiers	90
2.7.1	Charge Amplifiers Used with Piezoelectric Transducers	90
2.7.2	Charge Amplifier as Integrating Coulombmeter	93
2.7.3	Summary.....	94
2.8	Phase Sensitive Rectifiers.....	94
2.8.1	Double-Sideband, Suppressed Carrier Modulation	94
2.8.2	Demodulation of DSBSCM Signals by Analog Multiplier	95
2.8.3	Other PSR Designs	96
2.8.4	The Lock-In Amplifier.....	96
2.8.4.1	Introduction.....	96
2.8.4.2	Calculation of the SNR Improvement Using a Lock-In Amplifier	100
2.8.5	Summary.....	104
2.9	Chapter Summary	104
	Problems.....	105

Chapter 3 Noise and Coherent Interference in Measurements

3.1	Introduction	113
3.2	Random Noise in Circuits	113
3.2.1	Probability Density Functions.....	114
3.2.2	The Power Density Spectrum	115
3.2.3	Sources of Noise in Signal Conditioning Systems	118
3.2.3.1	Noise from Resistors.....	118
3.2.3.2	The Two Source Noise Model for Active Devices	120

3.2.3.3	Noise in JFETs	121
3.2.3.4	Noise in BJTs	122
3.3	Propagation of Gaussian Noise through Linear Filters.....	124
3.4	Broadband Noise Factor and Noise Figure of Amplifiers.....	125
3.5	Spot Noise Factor and Figure	127
3.6	Transformer Optimization of Amplifier F_{spot} and Output SNR.....	129
3.7	Cascaded Noisy Amplifiers.....	130
3.8	Calculation of the Noise Limited Resolution of Certain Signal Conditioning Systems.....	131
3.8.1	Calculation of the Minimum Resolvable AC Input Voltage to a Noisy, Inverting Op-amp Amplifier	131
3.8.2	Calculation of the Minimum Resolvable DC Current in White and $1/f$ Noise.....	133
3.8.3	Calculation of the Minimum Resolvable AC Input Signal to Obtain a Specified Output SNR in a Transformer Coupled, Tuned Amplifier.....	134
3.8.4	Calculation of the Smallest $\Delta R/R$ in a Wheatstone Bridge to Give a Specified SNR _o	135
3.8.5	Determination of the Conditions for Maximum Output SNR Given a Simple Inverting Op-Amp Amplifier with Known e_{na} and i_{na}	137
3.9	Modern, Low Noise Amplifiers for Use in Instrumentation Signal Conditioning Systems.....	137
3.10	Coherent Interference and its Minimization	139
3.10.1	Sources of Coherent Interference	139
3.10.1.1	Direct Electrostatic Coupling of Coherent Interference	140
3.10.1.2	Direct Magnetic Induction of Coherent Interference	141
3.10.1.3	Ground Loops	142
3.10.2	Cures for Coherent Interference	144
3.10.2.1	Power Line Low Pass Filters.....	144
3.10.2.2	Transient Voltage Suppressors	145
3.10.2.3	Coherent Interference Induced in Coaxial Cables by Magnetic Coupling	147
3.10.2.4	Single Grounding of Coax Shields to Prevent Ground Loop Interference	150
3.10.2.5	Use of a Longitudinal Choke or Neutralizing Transformer to Attenuate Common-Mode Coherent Interference	150
3.10.2.6	Experimental Verification of Cabling and Grounding Schemes to Achieve Minimum Noise Pickup.....	152
3.10.2.7	Circuit Grounding.....	154
3.10.2.8	Ferrite Beads and Feed Through Capacitors.....	155
3.10.2.9	Interruption of Ground Loops by the Use of Isolation Transformers and Photo-Optic Couples.....	155
3.10.2.10	Photo-Optic Couplers	155
3.10.2.11	The Use of Guarding and Shielding to Reduce Common-Mode, Coherent Interference.....	157
3.10.3	Summary of Techniques for Coherent Noise Reduction	163
3.11	Chapter Summary	163
	Problems.....	164

Chapter 4 DC Null Measurement	
4.1 Introduction	171
4.2 Wheatstone Bridge Analysis	172
4.3 The Kelvin Bridge	174
4.4 The Anderson Constant Current Loop.....	176
4.4.1 Introduction	176
4.4.2 The Anderson Loop Applied to Groups of Sensors.....	178
4.4.3 Conclusion	183
4.5 Potentiometers.....	183
4.6 Chapter Summary	185
Problems.....	185

Chapter 5 AC Null Measurements	
5.1 Introduction	191
5.2 Inductor Equivalent Circuits	191
5.3 Capacitor Equivalent Circuits	193
5.4 AC Operation of Wheatstone Bridges.....	195
5.5 AC bridges.....	195
5.5.1 Bridges Used to Measure Capacitance.....	196
5.5.1.1 The Resistance Ratio Bridge.....	196
5.5.1.2 The Schering Bridge.....	197
5.5.1.3 The Parallel C Bridge	198
5.5.1.4 The De Sauty Bridge.....	199
5.5.1.5 The Wien Bridge.....	200
5.5.1.6 The Commutated Capacitor Bridge	202
5.5.2 Bridges Used to Measure Inductance and Mutual Inductance	203
5.5.2.1 The Maxwell Bridge.....	203
5.5.2.2 Parallel Inductance Bridge.....	204
5.5.2.3 The Hay Bridge	205
5.5.2.4 The Owen Bridge	205
5.5.2.5 The Anderson Bridge	206
5.5.2.6 The Heaviside Mutual Inductance Bridge.....	207
5.5.2.7 The Heydweiller Mutual Inductance Bridge	208
5.5.3 Null Method of Measuring Transistor Small Signal Transconductance and Feedback Capacitance	209
5.6 Chapter Summary	211
Problems.....	211

Chapter 6 Survey of Sensor Input Mechanisms	
6.1 Introduction	215
6.2 Categories of Sensor Input Mechanisms	215
6.3 Resistive Sensors.....	215
6.3.1 Resistive Temperature Sensors.....	216
6.3.2 Resistive Strain Gauges	218
6.3.3 Photoconductors	220
6.3.4 Conductive Relative Humidity Sensors	223
6.3.5 Use of Resistance Change to Sense Position or Angle.....	226
6.3.6 Giant Magnetoresistive Effect-Based Sensors	227

6.3.7	Anisotropic Magnetoresistance (AMR)	230
6.4	Voltage Generating Sensors	236
6.4.1	Thermocouples and Thermopiles.....	236
6.4.2	Photovoltaic Cells.....	239
6.4.3	Piezoelectric Transducers.....	241
6.4.4	Pyroelectric Sensors	245
6.4.5	Sensors Whose Voltage Output is Proportional to $d\Phi/dt$	249
6.4.5.1	The Variable Reluctance Phonograph Pickup.....	249
6.4.5.2	Electrodynamic Accelerometer	251
6.4.5.3	Linear Velocity Sensors	252
6.4.6	Sensors whose Output EMF Depends on the Interaction of a Magnetic Field with Moving Charges.....	252
6.4.6.1	Faraday Effect Flowmeters	252
6.4.6.2	Hall Effect Sensors	255
6.5	Sensors Based on Variable Magnetic Coupling	257
6.5.1	The LVDT.....	257
6.5.2	Synchros and Resolvers	259
6.6	Variable Capacitance Sensors	262
6.7	Fiber Optic Sensors	264
6.7.1	Magneto-Optic Current Sensors	265
6.7.2	Measurement of the Optical Rotation of the Linearly Polarized Light Output of Certain Optical Sensors.....	268
6.7.3	Fiber Optic (FO) Mechanosensors.....	278
6.8	Photomultiplier Tubes and Related Electron Multiplication Devices	283
6.8.1	Introduction.....	283
6.8.2	Operation of PMTs.....	285
6.8.3	The Single-Channel Photomultiplier	288
6.8.4	Microchannel-Plate Photomultipliers	289
6.8.5	Summary.....	290
6.9	Ionizing Radiation Sensors	291
6.9.1	Geiger-Muller Tube.....	292
6.9.2	Solid State Crystal Radiation Sensors.....	295
6.9.3	Scintillation Counters	296
6.10	Electro-Chemical Sensors	298
6.10.1	pH and Specific Ion Electrodes	298
6.10.2	Polarographic Electrodes	300
6.10.3	Fuel Cell Electrodes	302
6.11	Electronic 'Noses'	302
6.11.1	The Rationale for Artificial Noses.....	302
6.11.2	Olfactory Anatomy and Physiology	303
6.11.3	The Vertebrate Olfactory Chemoreceptor System	304
6.11.4	Sensors for Gas-Phase Analytes.....	307
6.11.5	Feature Extraction in E-Noses.....	318
6.11.6	Summary	318
6.12	Mechano-Optical Sensors	318
6.12.1	Optical Coding Disks	319
6.12.2	Sagnac Effect Sensing of Angular Velocity.....	320
6.12.3	Laser Doppler Velocimetry.....	323
6.13	Chapter Summary	329
	Problems.....	329

Chapter 7 Applications of Sensors to Physical Measurements

7.1	Introduction	343
7.2	Measurement of Angular Acceleration, Velocity and Displacement	343
7.2.1	Angular Acceleration Measurement.....	343
7.2.1.1	Angular Acceleration Measurement with a Constrained Mechanical Gyro.....	344
7.2.1.2	Simple Inertia Wheel-Spring-Dashpot Angular Accelerometer	345
7.2.1.3	A Servo Angular Accelerometer.....	346
7.2.2	Angular Velocity Measurement with Mechanical Gyroscopes	347
7.2.2.1	The Mechanical Rate Gyro	350
7.2.2.2	Sagnac Effect Fiber Optic Gyroscopes.....	353
7.2.2.3	The Vibrating Mass Rate Gyro	354
7.2.2.4	The Humphrey Air Jet Gyro.....	357
7.2.2.5	Oscillating Disk Rate Gyro	361
7.2.3	Angular Velocity Measurement with Tachometers	365
7.2.4	Angular Position Measurement with Gyroscopes	371
7.2.5	Angular Position Measurement with Clinometers	373
7.2.6	Angular Position Measurement of Shafts	377
7.3	Measurement of Linear Acceleration, Velocity, Displacement and Position.....	378
7.3.1	Linear Accelerometers	378
7.3.1.1	A Basic Newtonian Accelerometer	378
7.3.1.2	Servo Accelerometer Design	379
7.3.1.3	Piezoelectric Sensor Accelerometers	380
7.3.1.4	Micromachined, IC Accelerometers	382
7.3.2	Linear Velocity Measurement	384
7.3.2.1	Measurement of Fluid Velocity.....	384
7.3.2.1.1	Hot Wire and Hot Film Anemometers	384
7.3.2.2	Measuring Fluid Velocity and Flow—Other Methods	388
7.3.3	Measurement of Linear Position	397
7.3.3.1	Use of the Global Positioning System to Locate Objects	401
7.3.3.2	The Use of Optical Interferometry to Measure Δx	409
7.3.3.2.1	Measurement of Tympanal Membrane Displacement by Fiber Optic Fizeau Interferometer	415
7.3.3.2.2	Measurement of Skin Vibration by Optical Interferometry.....	416
7.3.3.2.3	Two Frequency, Heterodyne Interferometry	419
7.3.3.2.4	The Fabry-Perot Interferometer	421
7.3.3.3	Phase Lock Velocity and Range Sensing with Ultrasound and EM Radiation.....	423
7.4	Measurement of Force and Torque	430
7.4.1	Load Cells for Force Measurement.....	430
7.4.2	Torque Measurements	434
7.5	Pressure Measurements.....	437
7.5.1	High Pressure Sensors.....	438
7.5.2	Low Pressure Sensors.....	445

7.6	Introduction to Substance Detection and Measurement Using Photons	451
7.6.1	Introduction	451
7.6.2	Dispersive Spectrophotometry	454
7.6.3	Non-Dispersive Spectroscopy	460
7.6.4	Fourier Transform IR Spectroscopy	464
7.6.5	Single and Multiple Bubble Sonoluminescence.....	470
7.6.5.1	Single Bubble Sonoluminescence	470
7.6.5.2	Multiple Bubble Sonoluminescence	473
7.6.5.3	Summary.....	474
7.6.6	Surface Plasmon Resonance	474
7.7	Temperature Measurements	479
7.7.1	Temperature Standards	479
7.7.2	Some Common Means of Temperature Measurement.....	481
7.7.2.1	Mechanical Temperature Sensors	481
7.7.2.2	Electrical and Electronic Temperature Sensors.....	483
7.7.2.2.1	The Resistance Noise Thermometer	483
7.7.2.2.2	The Resistance Thermometer.....	484
7.7.2.2.3	Electronic IC Temperature Sensors.....	486
7.7.2.2.4	Optical Pyrometers	486
7.8	Chapter Summary	493
	Problems.....	494

Chapter 8 Basic Electrical Measurements

8.1	Introduction	501
8.2	DC Voltage Measurements	501
8.2.1	Electromechanical DC Voltmeters	502
8.2.1.1	D'Arsonval DC Voltmeters	503
8.2.1.2	The Capacitor, or Electrostatic Voltmeter	507
8.2.1.3	The Electrodynamometer Meter	508
8.2.2	Electronic DC Voltmeters.....	510
8.3	Static Electric Fields and the Potential of Charged Surfaces Measurement	513
8.4	DC Current Measurements.....	522
8.4.1	Electromechanical DC Ammeters.....	522
8.4.2	Electronic DC Ammeters	524
8.4.2.1	Error Analysis of the Shunt Picoammeter	525
8.4.2.2	Error Analysis of the Feedback Picoammeter.....	526
8.4.3	Magneto-Optic Current Sensors	528
8.5	AC Voltage Measurements	528
8.5.1	Electromechanical AC Voltmeters	529
8.5.1.1	The Dynamometer AC Voltmeter.....	529
8.5.1.2	Iron Vane Voltmeters	529
8.5.1.3	Rectifier/D'Arsonval AC Voltmeters.....	530
8.5.1.4	Vacuum Thermocouple/D'Arsonval AC Voltmeters.....	532
8.5.2	Analog Electronic AC Voltmeters.....	533
8.5.2.1	AC Amplifier-Rectifier AC Voltmeters.....	533
8.5.2.2	Peak Reading Electronic AC Voltmeters.....	534
8.5.2.3	True RMS AC Voltmeter of the Feedback Type.....	535

8.5.2.4	True RMS AC Voltmeters Using the Direct Conversion Approach.....	539
8.5.3	Measurements of Amplifier Noise Voltages, Noise Factor and Figure	540
8.6	AC Current Measurements.....	547
8.6.1	Electromechanical (Analog) AC Ammeters.....	547
8.6.2	Electronic and Magneto-Optical AC Ammeters.....	549
8.7	Magnetic Field Measurements	550
8.8	Phase Measurements.....	559
8.8.1	Analog Phase Measurements.....	561
8.8.2	Digital Phase Detectors	564
8.9	Frequency and Period (Time) Measurements	566
8.10	Resistance, Capacitance and Inductance Measurement	571
8.10.1	Resistance Measurements	571
8.10.2	Capacitance Measurements	576
8.10.2.1	The Use of Q-Meter for Capacitance Measurement	576
8.10.2.2	Capacitance Measurement by Q/V	580
8.10.3	Inductance Measurements	584
8.10.3.1	Voltmeter Method of Estimating Inductance	585
8.10.3.2	The Use of Q-Meter for Inductance and Q Measurement	586
8.11	Vector Impedance Meters.....	586
8.12	Chapter Summary	588
	Problems.....	589

Chapter 9 Digital Interfaces in Measurement Systems

9.1	Introduction	597
9.2	The Sampling Theorem	597
9.3	Quantization Noise	601
9.4	Dithering	603
9.5	Digital-to-Analog Converters	606
9.6	The Hold Operation.....	611
9.7	Analog-to-Digital Converters	612
9.7.1	Successive-Approximation (SA) ADCs	612
9.7.2	Tracking or Servo ADCs	614
9.7.3	Dual-Slope Integrating ADCs	616
9.7.4	Flash (Parallel) ADCs	617
9.7.5	Dynamic Range, Floating Point ADCs	620
9.7.6	Delta-Sigma ADCs	621
9.7.7	Data Acquisition Cards for PCs	625
9.8	The IEEE-488 Instrumentation Bus (GPIB).....	626
9.8.1	The GPIB Bus Structure	627
9.8.2	GPIB Operation.....	628
9.9	Serial Data Communications Links.....	630
9.9.1	The RS-232C and D Interfaces	630
9.9.2	The RS-422, RS-423 and RS-485 Interfaces.....	633
9.9.3	The Universal Serial Bus (USB)	634
9.9.3.1	Overview of USB Functioning	635
9.9.3.2	Summary.....	637

9.10	The CAMAC (IEEE-583) Modular Instrumentation Standard, and the VXI Modular Instrumentation Architecture	637
9.11	Effect of Transmission Lines on the Transfer of Digital Data	640
9.11.1	The Transmission Line Model.....	640
9.11.2	Reflections on an Improperly Terminated, Lossless Transmission Line.....	642
9.12	Data Transmission on Fiber Optic Cables.....	646
9.12.1	Fiber Optic Cable Basics	647
9.12.2	Semiconductor Sources and Detectors Used in FOC Data Transmission.....	653
9.12.3	FOC Systems	656
9.13	Virtual Instruments	657
9.13.1	Introduction	657
9.13.2	PXI Systems	658
9.13.3	Summary.....	659
9.14	Chapter Summary	659
	Problems.....	660

Chapter 10 Introduction to Digital Signal Conditioning

10.1	Introduction	667
10.2	Digital Filters and the z-transform	667
10.3	Some Simple DSP Algorithms.....	672
10.4	Discrete and Fast Fourier Transforms and their Applications.....	680
10.4.1	Use of Data Windows to Improve Spectral Resolution	682
10.4.2	Use of the DFT to Characterize Random Signals and Noise.....	685
10.4.3	The Fast Fourier Transform	687
10.4.4	DSP Coprocessor Boards.....	688
10.5	Digital Routines for Interpolating Discrete Data	689
10.5.1	Estimating Missing Data at the Sampling Instants.....	692
10.6	Chapter Summary	693
	Problems.....	694

Chapter 11 Examples of the Design of Measurement Systems

11.1	Introduction	697
11.2	A Self-nulling, Microdegree Polarimeter to Measure Glucose in Bioreactors	697
11.3	Design of a System to Detect, Measure and Locate Partial Discharges in High Voltage Coaxial Power Cables.....	707
11.4	Design of a Closed Loop, Constant Phase, Pulsed Laser Ranging System and Velocimeter	713
11.5	Design of Capacitive Sensors for the Detection of Hidden Objects.....	719
11.5.1	Introduction	719
11.5.2	Self-balancing Circuits Used to Measure $\Delta C/C_o$	721
11.5.3	Summary	724
11.6	Chapter Summary	725

	References and Bibliography.....	727
--	----------------------------------	-----

	Index	737
--	-------------	-----

1

Measurement Systems

1.1 Introduction

In this introductory chapter we will examine the architecture of typical measurement systems and discuss how noise, calibration errors, sensor dynamic response and non-linearity can affect the accuracy, precision and resolution of measurements. We will also discuss the modern, physical and electrical standards used by the U.S. NIST (National Institute of Standards and Technology, formerly the National Bureau of Standards) and discuss how these standards are used to create secondary standards used for practical calibration of measurement systems.

Measurement systems are traditionally used to measure physical and electrical quantities, such as mass, temperature, pressure, capacitance and voltage. However, they can also be designed to locate things or events, such as the epicenter of an earthquake, employees in a building, partial discharges in a high voltage power cable, or a land mine. Often, a measurement system is called upon to discriminate and count objects, such as red blood cells, or fish of a certain size swimming past a checkpoint. A measurement system is often made a part of the control system. The old saying ‘if you can’t measure it, you can’t control it’ is certainly a valid axiom for both the control engineer as well as the instrumentation engineer.

The reader should realize that the fields of instrumentation and measurements are rapidly changing and new standards, sensors and measurement systems are continually being devised and described in journal literature. The *IEEE Transactions on Instrumentation and Measurement*, the *Review of Scientific Instruments*, the *IEEE Transactions on Biomedical Engineering* and the *Journal of Scientific Instruments* are four of the important periodicals dealing with the design of new measurement systems, instruments and standards.

1.2 Measurement System Architecture

Figure 1.1 illustrates the block diagram of a typical measurement system. The quantity under measurement (QUM) is converted to a useful form, such as a voltage, current or physical displacement by an input transducer or sensor. (Note the distinction between a sensor and a transducer—both devices are sensors, however, transducers have the capability to convert an input signal to some analog quantity). An example of a transducer is a piezoelectric crystal. A mechanical displacement (or equivalently, a force or pressure) applied to the crystal produces an output voltage, while an input voltage to the crystal produces a mechanical displacement. Another example is a loudspeaker. A current through the voice coil causes cone displacement, the rate of mechanical displacement generates a proportional current output.

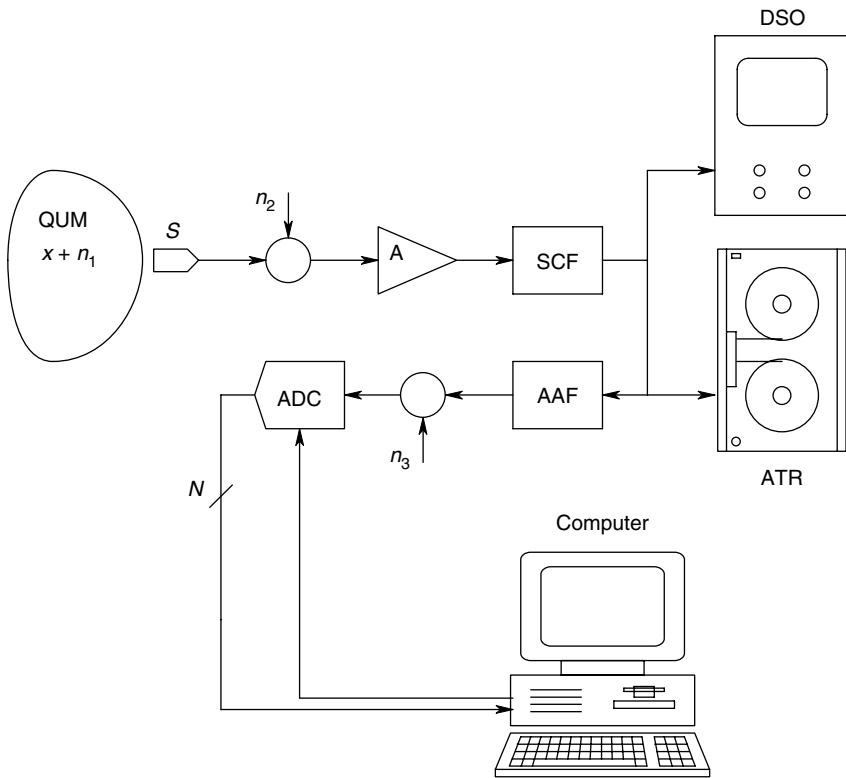


FIGURE 1.1

A generalized measurement system. Note: QUM = quantity under measurement, S = sensor, A = amplifier, SCF = signal conditioning filter, DSO = digital storage oscilloscope, ATR = analog tape recorder, AAF = anti-aliasing (lowpass) filter, n_1 = noise accompanying the QUM, n_2 = noise from electronics, n_3 = equivalent quantization noise, ADC = analog to digital converter, DC = digital computer, MON = monitor, KBD = keyboard.

The output of the sensor, here assumed to be a voltage, is amplified and filtered by a signal conditioning subsystem. The purposes of the signal conditioning subsystem are to amplify, to give a low or matched output impedance and to improve the signal-to-noise-ratio (SNR) of the analog signal proportional to the QUM. The conditioned analog signal, V_1 , can be distributed to various display and recording devices. V_1 can be displayed on an analog or digital oscilloscope, or a strip-chart recorder. It can also be recorded on a magnetic tape recorder, or it can be low-pass filtered to prevent aliasing and then periodically converted to digital words by an analog to digital converter (ADC).

The ADC output, D_N , is generally made an input to a digital computer through an appropriate interface. Once in the computer environment, the sampled and digitized QUM can be further (digitally) filtered, processed and stored on a magnetic hard and/or floppy disk, or on an optical disk in digital form. Often, the sensing instrument contains an ADC and is connected to a computer by a special instrumentation bus such as the GPIB or IEEE-488 bus (Hewlett-Packard calls the GPIB the HPIB). The GPIB is organized so that many GPIB-compatible sensing instruments can be managed by one master (host) computer. More information about the GPIB is available in Chapter 9 of this text.

Note that there are three major sources of noise in the measurement system shown in Figure 1.1.

1. Noise sensed along with the QUM (environmental noise)
2. Noise associated with the electronic signal conditioning system (referred to its input)
3. Equivalent noise generated in the analog to digital conversion process (quantization noise)

These noise sources generally limit the resolution and accuracy of the measurement system.

Many sensors provide the rate limiting element in a measurement system. That is, the response speed of the sensor is much slower than any of the other elements in the system. As a result of sensor dynamics, when there is a step change in the QUM, a measurement system may require a minimum settling time before a stable measurement can be made. Therefore, a settling time may be required for the sensor to reach a steady state output.

1.2.1 Sensor Dynamics

Sensor dynamics generally fall into one of three categories: First order low-pass, second order low-pass, and bandpass. A general example of a first order, low-pass sensor is shown in Figure 1.2A. The QUM is $x(t)$, and the sensor output is $V_x(t)$. The ordinary differential equation (ODE) relating V_x and x is:

$$\dot{V}_x(t) = -aV_x(t) + Kx(t) \quad (1.1a)$$

or

$$\dot{V}_x(t)\tau = -V_x(t) + K\tau x(t) \quad (1.1b)$$

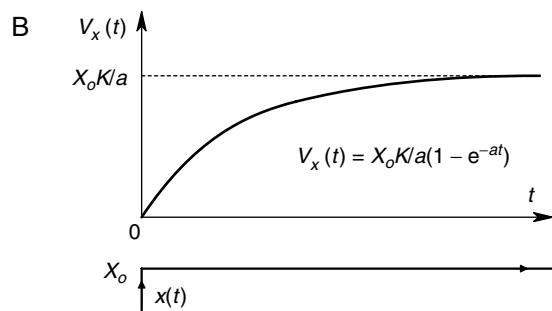
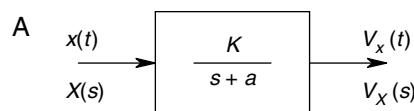


FIGURE 1.2

A. Transfer function of sensor with first order dynamics in Laplace format. $x(t)$ is the QUM, $X(s)$ is its Laplace transform, $V_x(t)$ is the sensor's voltage output, $V_X(s)$ is its Laplace transform, a = natural frequency of the first order sensor. B. Step response of the first order sensor.

Here, a is the natural frequency of the sensor in r/s, and $\tau=1/a$ is its time constant in seconds. The solution of the above ODE for a step input, $X(t)=X_o U(t)$, gives:

$$V_x(t) = \frac{X_0 K}{a} (1 - e^{-at}) = X_o K \tau (1 - \exp[-t/\tau]) \quad (1.2)$$

The sensor step response is plotted in Figure 1.2B. Note that the response is within 5% of the steady state value for $t=3\tau$ and within 2% of steady state for $t=4\tau$. The transfer function for the first order, low-pass sensor is found by carrying out the Laplace transform of the ODE in equation 1.1a,

$$\frac{V_x(s)}{X(s)} = \frac{K}{s + a} = H(s) \quad (1.3)$$

and its frequency response is found by letting $s=j\omega$ in equation 1.3.

Second order sensor dynamics fall into one of three categories, depending on the location of the roots of the characteristic equation of the ODE. These categories are *underdamped* (with complex conjugate roots in the left half s -plane), *critically damped* (with two, equal real roots on the negative real axis in the s -plane) and *overdamped* (with unequal real roots on the negative real axis in the s -plane). In all cases, the real parts of the ODE's roots are negative, as the sensor is assumed to be stable. The ODEs for the three damping conditions are:

$$\ddot{V}_x = -\dot{V}_x(2\zeta\omega_n) - V_x(\omega_n^2) + Kx(t) \quad (1.4a)$$

$$\ddot{V}_x = -\dot{V}_x(2a) - V_x a^2 + Kx(t) \quad (1.4b)$$

$$\ddot{V}_x = -\dot{V}_x(a+b) - V_x ab + Kx(t) \quad (1.4c)$$

In the underdamped case in equation 1.4a above, ζ is the damping factor and ω_n is the natural frequency in r/s. For an underdamped sensor, $0 \leq \zeta \leq 1$ and the roots of the ODE are located at $s = -\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$. The critically damped sensor has two roots at $s = -\omega_n$, and the over-damped sensor has roots at $s = -a$ and $s = -b$.

These conditions are shown in Figure 1.3. Using Laplace transforms, we can find the step response of the second order sensor for the above three conditions. The step response for the underdamped case is:

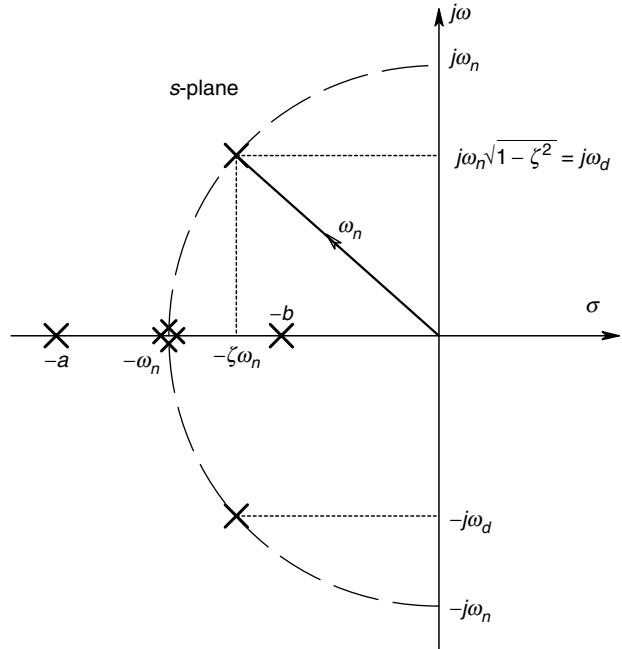
$$V_x(t) = \frac{KX_o}{\omega_n^2} \left\{ 1 - \frac{1}{\sqrt{1-\zeta^2}} \exp(-\zeta\omega_n t) \sin[\omega_n\sqrt{1-\zeta^2} t + \phi] \right\} \quad (1.5)$$

where

$$\phi = \tan^{-1} \left(\frac{\sqrt{1-\zeta^2}}{\zeta} \right) \quad (1.6)$$

The underdamped, second order system step response is illustrated in Figure 1.4A. The step response of the critically damped sensor ($\zeta=1$) is shown in Figure 1.4B and is given by:

$$V_x(t) = \frac{KX_o}{a^2} [1 - e^{-at} - ate^{-at}] \quad (1.7)$$

**FIGURE 1.3**

Pole locations of a second order sensor's transfer function for the cases of overdamping (poles at $s = -a, -b$), critical damping (two poles at $s = -\omega_n$) and underdamping (poles at $s = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$).

Finally, the step response of an overdamped sensor having two real poles is shown in Figure 1.4C and can be written as:

$$V_x(t) = \frac{KX_o}{ab} \left[1 - \frac{1}{b-a} (be^{-at} - ae^{-bt}) \right], \quad b > a \quad (1.8)$$

While the three types of low-pass sensor dynamics described above are most commonly encountered, other sensors which do not respond to a constant (dc) QUM also exist. The QUM must, in fact, be time-varying to produce a sensor output. Such sensors are said to have a bandpass frequency response characteristic and can be described by the ODE:

$$\ddot{V}_x = -\dot{V}_x(a + b) - V_x(ab) + \dot{K}x, \quad b > a \quad (1.9)$$

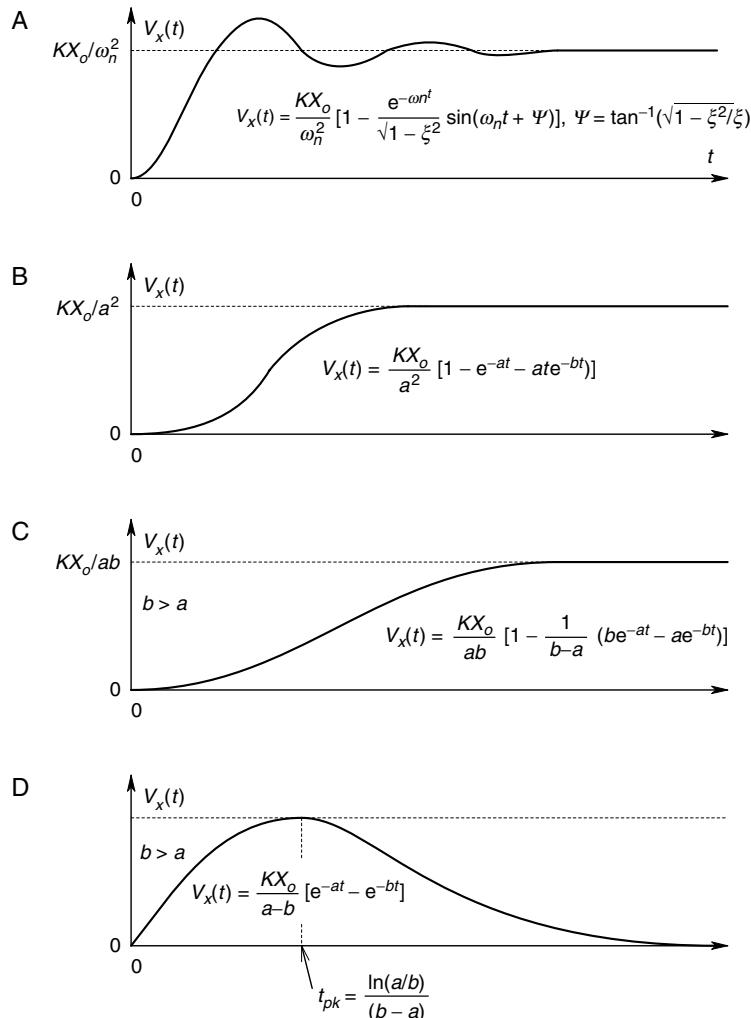
The corresponding transfer function for a bandpass sensor with two real poles is:

$$\frac{V_x(s)}{X(s)} = \frac{Ks}{(s+b)(s+a)} \quad (1.10)$$

The output voltage step response of a bandpass sensor rises from zero to a peak, then goes to zero in the steady state. This is illustrated in Figure 1.4D and is given by:

$$V_x(t) = \frac{KX_o}{b-a} (e^{-at} - e^{-bt}) \quad (1.11)$$

A piezoelectric crystal pressure transducer is an example of a bandpass sensor.

**FIGURE 1.4**

Step responses of various types of sensors. A. Underdamped, second order, low-pass sensor. B. Critically damped, second order, low-pass sensor. C. Overdamped, second order, low-pass sensor. D. Overdamped, second order, bandpass sensor.

1.2.2 Overview of Signal Conditioning

The voltages or currents obtained directly at the output of a sensor are generally low level and contain additive noise and coherent interference, picked up from the environment of the QUM and from the sensor itself. Sometimes the measurement process or the sensor introduces a nonlinear distortion of the QUM (as in the case of hot wire anemometers) which must be linearized. The analog signal conditioning module following the sensor must therefore amplify the sensor output voltage, as well as perform linear filtering on it in order to improve the SNR. Such amplification and filtering is usually performed by a low noise instrumentation amplifier, followed by op-amp active filters. Compensation for inherent nonlinearities, including corrections for temperature changes, is most easily done digitally, with a computer. Thus, the sensor output, after analog conditioning, is converted to digital form by an ADC and read into the computer as a binary number. If the digitization is done on a periodic basis, it is common practice to

use the analog signal conditioning system to low-pass filter the input to the ADC, so that the ADC input voltage contains no significant energy having frequencies in excess of one-half of the sampling frequency. This low-pass filtering before sampling is called *anti-aliasing filtering* and is necessary for accurate digital signal processing. Sampling and anti-aliasing filters are discussed in detail in Chapter 10.

Digital signal conditioning can also be used to effectively remove the coherent interference accompanying the QUM and to calculate functions of the sampled signal, such as its RMS value, its auto-correlation function, or its root power spectrum. It is now common practice to store digitized measurement records on floppy or hard magnetic disks, on optical disks, or on video magnetic tape in digital form.

1.3 Errors in Measurements

Errors in measurements can arise from many causes; there are remedies for some types of errors, but others haunt us as intrinsic properties of the measurement system under use, and often can be mitigated by system redesign.

1.3.1 Gross Errors

These errors in a measurement can arise from such human mistakes as:

1. Reading the instrument before it has reached its steady state. This type of premature reading produces a dynamic error
2. Not eliminating parallax when reading an analog meter scale, incorrect interpolation between analog meter scale markings
3. Mistakes in recording measured data and in calculating a derived measurand
4. Misuse of the instrument. A simple example of this source of error is when a 10 V full scale analog voltmeter of typical sensitivity ($20,000 \Omega/V$) is connected to a voltage source having a high Thevenin resistance ($100,000 \Omega$). Significant voltage is dropped across the internal (Thevenin) resistor

1.3.2 System Errors

These errors can arise from such factors as:

1. The instrument is not calibrated, and has an offset (i.e. its sensitivity is off, and it needs zeroing). Loss of calibration and zero error can occur because of long term component value changes due to ageing, or changes associated with temperature rise
2. Reading uncertainty due to the presence of random noise. This noise can accompany the measurand and can arise from the signal conditioning electronics in the system
 - (a) Noise from without is called environmental noise and can often be reduced by appropriate electric and magnetic shielding, proper grounding and guarding practices. Environmental noise is often coherent in nature and can

come from poorly shielded radio frequency sources, such as computers or radio stations. Powerline frequency electric and magnetic fields can be troublesome, as also ignition noise

- (b) Significant, internally generated random noise may be shown to arise in the first stage of an instrument's signal conditioning amplifier. Some of this noise comes from resistors (Johnson or thermal noise), some from the active elements (transistors) in the headstage and some from the quantization or rounding off that is inherent in the operation of ADCs in modern digital instruments
- 3. Slow, or long term drift in the system can destroy the certainty of static measurements (i.e. measurements of measurands which are not varying in time). Drifts can cause slow changes in system sensitivity and/or zero. Drift may arise as the result of a slow temperature change as a system warms up. Many parameter values can change with temperature, such as capacitance, resistance, inductance, the EMF of standard cells, the avalanche voltage of $p-n$ junctions, and so on. Humidity changes can also affect circuit parameter values, including resistance and capacitance. Good system design involves temperature regulation of the system, the use of low or zero tempco and sealed components. In integrators such as charge amplifiers, output drift can also be caused by the integration of the op-amp's dc bias current.

Drift or system offset can also arise from dc static charges, which may affect analog electronic circuits. In some cases, the face of an analog meter can become charged with static electricity, which then attracts or repels the meter pointer, causing an error.

Below, we discuss the concepts of *accuracy*, *precision*, *resolution*, *limiting error*, and various error statistics. To begin the quantitative discussion of errors in measurements, we define the *error* in the n th measurement as:

$$\varepsilon_n \equiv X_n - Y_n \quad (1.12)$$

and

$$\% \varepsilon \equiv |\varepsilon_n / Y_n| \times 100 \quad (1.13)$$

Here, Y_n is the actual, true, defined or calculated value of the QUM, and X_n is the n th measured value of the QUM. Philosophically, the use of Y_n in the above definitions may present an ontological problem because one can argue that the *true* value, which is the result of an ideal measurement process, can never be known. In which case, Y_n may be defined by a high resolution, primary standard. Nevertheless, error, as given in equation 1.12, is generally an accepted definition. In some cases, the absolute value signs in the $\% \varepsilon$ expression may be omitted and the $\% \varepsilon$ can be negative.

The *accuracy*, A_n , of the n th measurement is defined as:

$$A_n \equiv 1 - \left| \frac{Y_n - X_n}{Y_n} \right| \quad (1.14)$$

The percent accuracy is simply

$$\% A_n \equiv 100 - \% \varepsilon = A \times 100 \quad (1.15)$$

The *precision* of the n th measurement is defined as:

$$P_n \equiv 1 - \left| \frac{X_n - \bar{X}}{\bar{X}} \right| \quad (1.16)$$

\bar{X} is defined as the *sample mean* of N measurements, given by:

$$\bar{X} \equiv \frac{1}{N} \sum_{n=1}^N X_n \quad (1.17)$$

Low noise accompanying the QUM and in the signal conditioning system is required as a necessary (but not sufficient) condition for *precision measurements*. Precision measurements also require a measurement system capable of resolving very small changes in the QUM, say one part in 10^7 . Consequently, many precision, high resolution measurement systems rely on null methods, such as those used with ac and dc bridges, potentiometers, etc. The high resolution inherent in null measurements comes from the fact that the characteristics of the bridge components themselves are known to high precision and the null sensor is capable of resolving very small voltage or current differences around the null. The null condition equation(s) allow calculation of the QUM to high precision, given high accuracy components (parameters).

Accurate measurements require the use of a precision measurement system which is calibrated against a certified, accurate standard.

Other statistics used to describe the quality of a series of measurements are the deviation, d_n , the average deviation, D_N , and the standard deviation, S_N , defined as:

$$d_n \equiv X_n - \bar{X} \quad (1.18)$$

$$D_N \equiv \frac{1}{N} \sum_{n=1}^N d_n \quad (1.19)$$

$$S_N \equiv \frac{1}{N} \sum_{n=1}^N d_n^2 = \sigma_x \quad (1.20)$$

The variance of X is simply S_N^2 and can be shown to be equal to:

$$S_N^2 \equiv \frac{1}{N} \sum_{n=1}^N X_n^2 - (\bar{X})^2 = \sigma_x^2 \quad (1.21)$$

If it is assumed that the samples of the QUM, X_n , obey a normal or Gaussian probability density function, then the probable error, pe , in any one observation is defined such that there is a 50% probability that X_n lies between $\bar{X} - pe$ and $\bar{X} + pe$. From the normal probability density function, it is easy to show that in order to satisfy the above condition, $pe = 0.6745\sigma_x$.

A noisy set of measurements will have a small \bar{X}/σ_x , which may be considered to be an RMS SNR. If we increase the total number of measurements, N , then the SNR can be shown to improve by a factor of \sqrt{N} .

The *resolution* in measuring the QUM is related to the precision of the measurement and is basically the smallest unit of the QUM which can be reliably be detected.

Analog indicating instruments, with scales covering roughly a 90° arc range, provide a fundamental limitation to measurement system accuracy, even when null measurements are performed. Large, mirror scale analog meters with knife-edge pointers can usually be read to no better than 0.2%, or 2 parts/thousand. Oscilloscopes generally offer slightly poorer resolution. As a result of this physical limitation to resolution, direct reading instruments with digital outputs are used when accuracies better than 0.2% are needed. Null instruments have taps and verniers to permit high resolution.

Limiting Error (LE) is an important parameter used in specifying instrument accuracy. The limiting error, or guarantee error, is given by manufacturers to define the outer bounds or the expected worst case error. For example, a certain voltmeter may be specified as having an accuracy of 2% of its full scale reading. If on the 100 V scale, the meter reads 75 V, then the LE in this reading is $(2/75) \times 100 = 2.67\%$.

In many cases, such as in the determination of a resistor value by Wheatstone bridge, the QUM must be calculated from a formula in which various system parameters, each having a specified accuracy, appears. Thus, we must derive a formula for the determination of LE in the value of the calculated QUM. Let the QUM be a function of N variables, i.e.

$$Q = f(X_1, X_2, \dots, X_N) \quad (1.22)$$

Let us assume each variable, X_j , is in error by $\pm \Delta X_j$. Hence the calculated QUM will be noisy and will be given by:

$$\hat{Q} = f(X_1 \pm \Delta X_1, X_2 \pm \Delta X_2, \dots, X_N \pm \Delta X_N) \quad (1.23)$$

\hat{Q} can be expanded into a Taylor's series for N variables. For one variable,

$$f(X \pm \Delta X) = f(X) + \frac{df}{dX} \frac{\Delta X}{1!} + \frac{d^2f}{dX^2} \frac{(\Delta X)^2}{2!} + \dots + \frac{d^{n-1}f}{dx^{n-1}} \frac{(\Delta X)^{n-1}}{(n-1)!} + R_n \quad (1.24)$$

Hence for the case with N variables:

$$\begin{aligned} \hat{Q} &= f(X_1, X_2, \dots, X_N) + \left\{ \frac{\partial f}{\partial X_1} \Delta X_1 + \frac{\partial f}{\partial X_2} \Delta X_2 + \dots + \frac{\partial f}{\partial X_N} \Delta X_N \right\} + \\ &\quad \frac{1}{2!} \left\{ \frac{\partial^2 f}{\partial X_1^2} (\Delta X_1)^2 + \frac{\partial^2 f}{\partial X_2^2} (\Delta X_2)^2 + \dots + \frac{\partial^2 f}{\partial X_N^2} (\Delta X_N)^2 \right\} + \dots \\ &\quad + \frac{1}{3!} \left\{ \frac{\partial^3 f}{\partial X_1^3} (\Delta X_1)^3 + \dots \right\} + \dots \end{aligned} \quad (1.25)$$

The second and higher derivative terms are assumed to be numerically negligible. Thus the maximum or worst case uncertainty in Q can be finally approximated by:

$$\Delta Q_{MAX} = |Q - \hat{Q}| = \sum_{j=1}^N \left| \frac{\partial f}{\partial X_j} \Delta X_j \right| \quad (1.26)$$

As an example of the use of equation 1.26 for a worst case uncertainty, let us examine the LE in the calculation of the dc power in a resistor:

$$P = I^2 R \quad (1.27)$$

so

$$\Delta P_{\text{MAX}} = 2IR\Delta I + I^2\Delta R \quad (1.28)$$

and

$$\frac{\Delta P_{\text{MAX}}}{P} = 2 \left| \frac{\Delta I}{I} \right| + \left| \frac{\Delta R}{R} \right| \quad (1.29)$$

Thus, if the LE in R is 0.1%, the 0–10 A ammeter has 1% of full scale accuracy, the resistor value is 100Ω and the ammeter reads 8 A. and the nominal power dissipated in the resistor is 6,400 W, then the LE in the power measurement is:

$$\frac{\Delta P_{\text{MAX}}}{P} = 2 \frac{0.1}{8} + 0.001 = 0.026 \text{ or } 2.6\% \quad (1.30)$$

As a second example, consider finding the limiting error in calculating the Nernst potential for potassium ions, given by equation 1.31 below:

$$E_K = \frac{RT}{\mathfrak{F}} \ln \left(C_o / C_i \right) \quad (1.31)$$

Here, R is the gas constant, T is the Kelvin temperature, \mathfrak{F} is the Faraday number, C_o is the potassium ion concentration outside a semi-permeable membrane and C_i is the potassium concentration inside a compartment bounded by the membrane. To find the LE in calculating E_K , we differentiate equation 1.31 according to equation 1.26 above:

$$\Delta E_K = \left| \frac{R}{\mathfrak{F}} \ln \left(C_o / C_i \right) \Delta T \right| + \left| \frac{RT}{\mathfrak{F}} \frac{\Delta C_o}{C_o} \right| + \left| \frac{RT}{\mathfrak{F}} \frac{\Delta C_i}{C_i} \right| \quad (1.32)$$

Now if we divide equation 1.32 by equation 1.31, we find:

$$\frac{\Delta E_{K_{\text{MAX}}}}{E_K} = \left| \frac{\Delta T}{T} \right| + \left| \frac{\Delta C_o}{C_o} \frac{1}{\ln(C_o/C_i)} \right| + \left| \frac{\Delta C_i}{C_i} \frac{1}{\ln(C_o/C_i)} \right| \quad (1.33)$$

Let the temperature at 300 K be known to within $\pm 0.1^\circ$, and C_o be 100 mM, C_i be 10 mM and the concentrations be known to within $\pm 2\%$. The LE in E_K is thus,

$$\frac{\Delta E_{K_{\text{MAX}}}}{E_K} = 0.1/300 + 0.02/2.3 + 0.02/2.3 = 1.77 \times 10^{-2} \text{ or } 1.77\% \quad (1.34)$$

In conclusion, we have seen that the Taylor's series approach to the calculation of LE for a derived (calculated) measurand provides us with a systematic means of finding

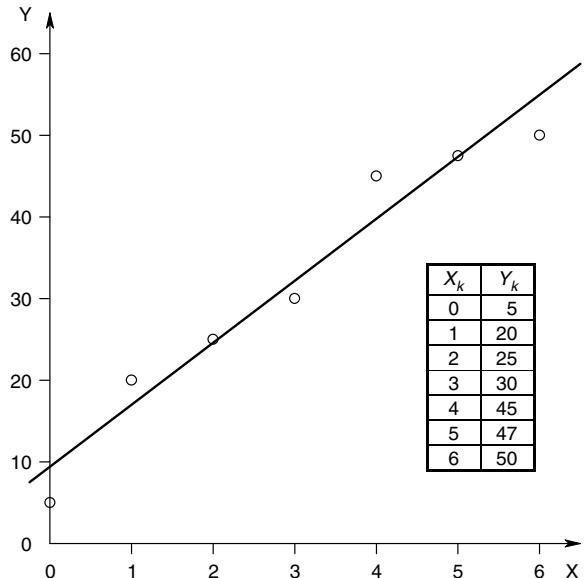
**FIGURE 1.5**

Illustration of least mean square error fitting of a number of data points with a linear regression line.

the worst case error in that quantity. In actuality, the uncertainties in the parameters may, fortuitously, combine algebraically to give a low degree of error. The details of this combination of errors is however, unknown. Therefore, we generally take the LE as the outer bound in evaluating the accuracy of a measurement.

Least mean square linear graph fitting, or linear regression, is a technique used to obtain an optimum linear fit of the form $y = mx + b$ to a noisy set of N measurement points, $\{Y_k, X_k\}, k = 1, \dots, N$ (such a fit is illustrated for a typical case in Figure 1.5). For example, we may wish to measure the optical rotation of polarized light, Y_k , as it passes through sugar solutions of various concentrations, X_k . In the optimum least mean square (LMS) error fitting of $y = mx + b$ to the data points, we wish to choose m and b to minimize the function, σ_y^2 :

$$\sigma_y^2 = \frac{1}{N} \sum_{k=1}^N [(mX_k + b) - Y_k]^2 \quad (1.35)$$

Equation 1.35 can be expanded to give:

$$\sigma_y^2 = \frac{1}{N} \sum_{k=1}^N [m^2 X_k^2 + b^2 + Y_k^2 + 2m X_k b - 2Y_k(m X_k + b)] \quad (1.36)$$

In order to minimize σ_y^2 , we set the derivatives equal to zero:

$$\frac{\partial \sigma_y^2}{\partial b} = 0 = 2Nb + 2m \sum_{k=1}^N X_k - 2 \sum_{k=1}^N Y_k \quad (1.37a)$$

$$\frac{\partial \sigma_y^2}{\partial m} = 0 = 2m \sum_{k=1}^N X_k^2 + 2b \sum_{k=1}^N X_k - 2 \sum_{k=1}^N X_k Y_k \quad (1.37b)$$

The two equations, 1.37a and b above, can be written as two simultaneous linear equations in m and b , and solved by conventional methods to obtain:

$$b = \frac{\bar{Y}\bar{X}^2 - \bar{X}R_{xy}(0)}{\sigma_x^2} \quad (1.38a)$$

$$m = \frac{R_{xy}(0) - \bar{X}\bar{Y}}{\sigma_x^2} \quad (1.38b)$$

for the y -intercept and the slope of the LMS linear fit, respectively. $R_{xy}(0)$ is the cross correlogram function evaluated at $\tau=0$, given by:

$$R_{xy}(0) = \frac{1}{N} \sum_{k=1}^N X_k Y_k \quad (1.39)$$

The goodness of the fit of $y=mx+b$ to the data set, $\{Y_k, X_k\}$, is given by the correlation coefficient for the LMS fit, r , defined as:

$$r \equiv [R_{xy}(0) - \bar{X}\bar{Y}] / \sigma_x \sigma_y, \quad 0 \leq r \leq 1 \quad (1.40)$$

and r^2 is called the *coefficient of determination* of the linear regression fit. Obviously, $r=1$ indicates a perfect fit, which means that all $Y_k=mX_k+b$.

It should be pointed out that not all measurands are linear functions of a single independent variable and the above analysis represents one of the simpler cases in statistical data analysis.

1.4 Standards Used in Measurements

One of the main concerns of all those who have occasion to make a measurement is whether the instrument is calibrated or not. As seen above, calibration is necessary, along with precision, to enable accurate measurements to be made. Calibration implies observing the instrument's performance when measuring a standard of some sort. Major changes have taken place in the definitions of standards in the late 1980s, and on 1 January 1990, several of these new standards were adopted by the international community. In the United States, the responsibility for maintaining primary and secondary standards lies with the National Institute for Standards and Technology (NIST). NIST also actively seeks to establish new and more accurate standards as well as the means of transferring their accuracy when calibrating instruments.

1.4.1 Electrical Standards

A standards is a physical representation of the QUM whose true value is known with great accuracy. Standards can be classified as:

1. International standards
2. Primary standards

3. Secondary standards
4. Working standards

International standards are defined by international agreement and are kept at the International Bureau of Weights and Measures in Sèvres, France. An example of an international standard is the *kilogram mass* (more will be said about the mass standard below). International standards are not available on a daily basis for calibration or comparison.

Primary standards are maintained in national standards laboratories in countries around the world. Primary standards, representing some of the fundamental physical and electrical units, as well as some derived quantities, are independently measured and calibrated at the various national laboratories and compared against each other. This process leads to grand or world average figures for the standards. Primary standards are used continually, but generally do not leave the national standards labs.

Secondary standards are reference standards which are initially calibrated from primary standards and then used in industry and research labs on a daily basis to calibrate their working standards, which are in turn used on a daily basis to check and calibrate working laboratory instruments.

1.4.1.1 Potential Difference

The defined international standard for the volt, which was in effect from 1908 to 1990, is based on the EMF of the saturated Weston Standard Cell, first developed in 1892 by Edward Weston. The Weston cell is an electrochemical battery which consists of two half-cell electrodes in a sealed, glass H-tube (Figure 1.6). In the normal Weston standard cell, the electrolyte is a saturated aqueous solution of cadmium sulfate (CdSO_4). The saturated CdSO_4 electrolyte gives the EMF of the normal cell a relatively high temperature coefficient (tempco), about $-40 \mu\text{V}/^\circ\text{C}$. However, this form of Weston cell shows higher, long term, output EMF stability than the cells with unsaturated electrolyte. The EMF of the saturated Weston cell drifts about $1 \mu\text{V}/\text{year}$ and the useful life of a well treated normal cell is 10–20 years [Helfrick and Cooper, 1990].

Under nearly open circuit conditions, the EMF of the normal cell in absolute volts is given by:

$$E(\Delta T) = E_{20} - 4.6 \times 10^{-5}(\Delta T) - 9.5 \times 10^{-7}(\Delta T)^2 - 1 \times 10^{-8}(\Delta T)^3 \quad (1.41)$$

where E_{20} is the normal cell EMF at exactly 20°C , ($E_{20} \equiv 1.01858$ Absolute volts), and $\Delta T = (T^\circ\text{C} - 20)$. One volt was defined as $1/1.01858$ times the EMF of a normal Weston cell at exactly 20°C . Due to their high temperature sensitivity, saturated Weston cells are kept in temperature regulated baths, where the temperature is kept constant to within $\pm 0.000010^\circ$.

The present international standard for the volt was adopted on 1 January 1990 and is based on a quantum effect phenomenon which takes place at liquid helium temperatures, namely the *Josephson effect*. One form of a *Josephson junction* (JJ) consists of two thin films of superconducting lead separated by a 1 nm thick layer of lead oxide [Taylor, 1990]. Another JJ design uses superconducting niobium separated by a thin film of Al_2O_3 . Such JJs can be fabricated into series arrays of JJs using integrated circuit technology [Pöpel *et al*, 1991]. Pöpel *et al* reported on the performance of IC arrays

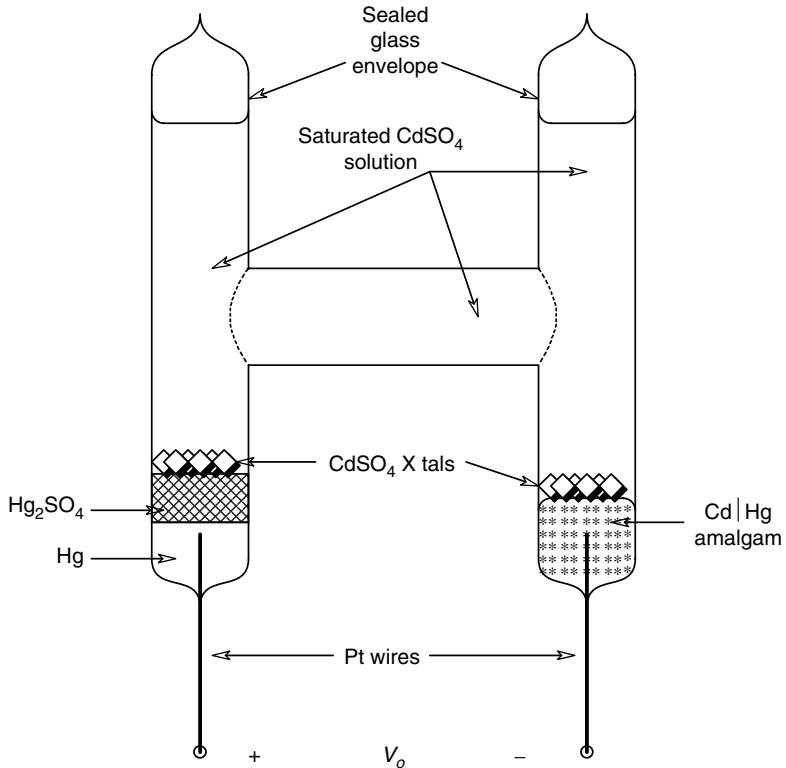
**FIGURE 1.6**

Diagram of a saturated Weston standard cell battery.

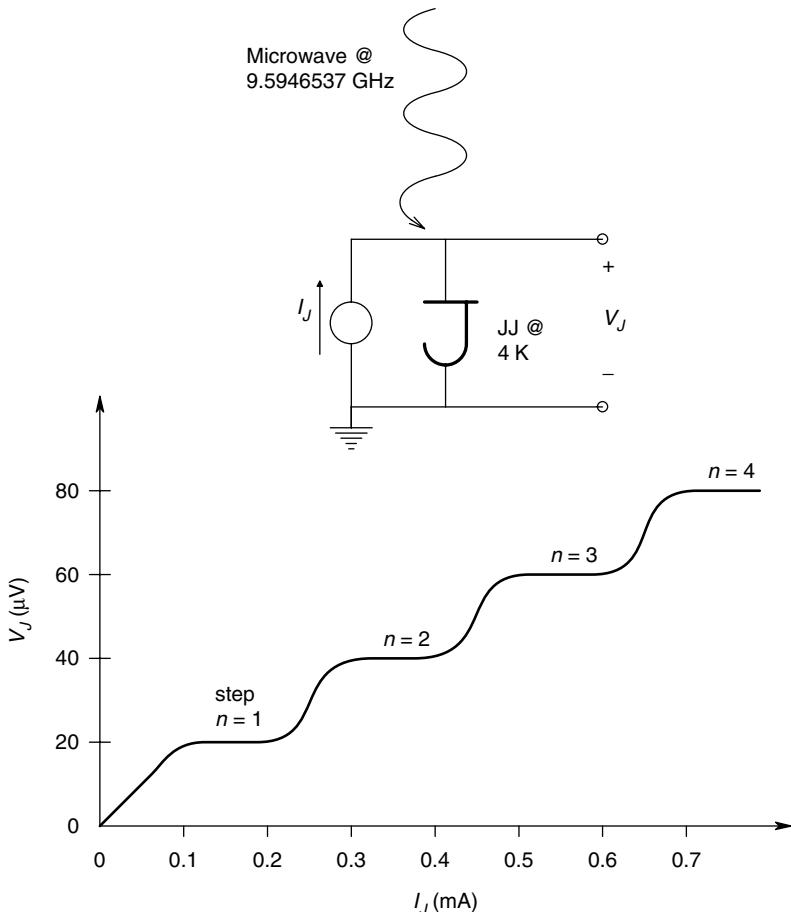
with 2,000 and 20,160 JJs, which were useful as standards in the range of 1 and 10 V dc, respectively.

A single, superconducting JJ has the unique property that when it is irradiated with microwave energy in the frequency range 9–100 GHz and biased with a dc current, a precisely known dc voltage appears across the dc portion of the JJ, having a stepwise volt-ampere curve, as shown in Figure 1.7. The voltage of each stable step, E_J , is given by

$$E_J = nf/(2e/h) \text{ V} \quad (1.42)$$

where n is the step number, f is the microwave frequency in GHz, e is the electron charge, and h is Planck's constant. The quantity $(2e/h)$ is known as K_J , the *Josephson constant*. The Comité International des Poids et Mesures (CIPM) established a universal standard value for the Josephson constant, $K_{J-90} = 483,597.9 \text{ GHz/V}$ on 1 January 1990. For a typical microwave frequency of 94 GHz, derived from a Gunn oscillator, the step voltage is seen to be $194.4 \mu\text{V}$. Exactly one volt can be defined as the EMF of a JJ when it is irradiated with electromagnetic radiation at a frequency of 483,597.9 GHz. Of course, such a frequency is a mathematical convenience; the JJ would not work when irradiated with 620 nm light, which has that frequency.

When using a JJ array as a primary (SI) standard, the microwave frequency is seen to be quite critical. Thus, the Gunn oscillator is phase locked to an ultra-stable frequency source, such as a cesium beam clock. The frequency source which has been used

**FIGURE 1.7**

Volt-ampere curve of a typical Josephson Junction voltage standard. The microwave frequency was 9.5946537 GHz. V_J is approximately 1.9 mV at the 97th step, with $I_J = 24$ mA, at 4.2 K. (Source: Drawn from data given by Ibuka *et al*, 1983.)

for several designs of JJ array voltage standards has been a 50–60 mW, 93–95 GHz Gunn oscillator.

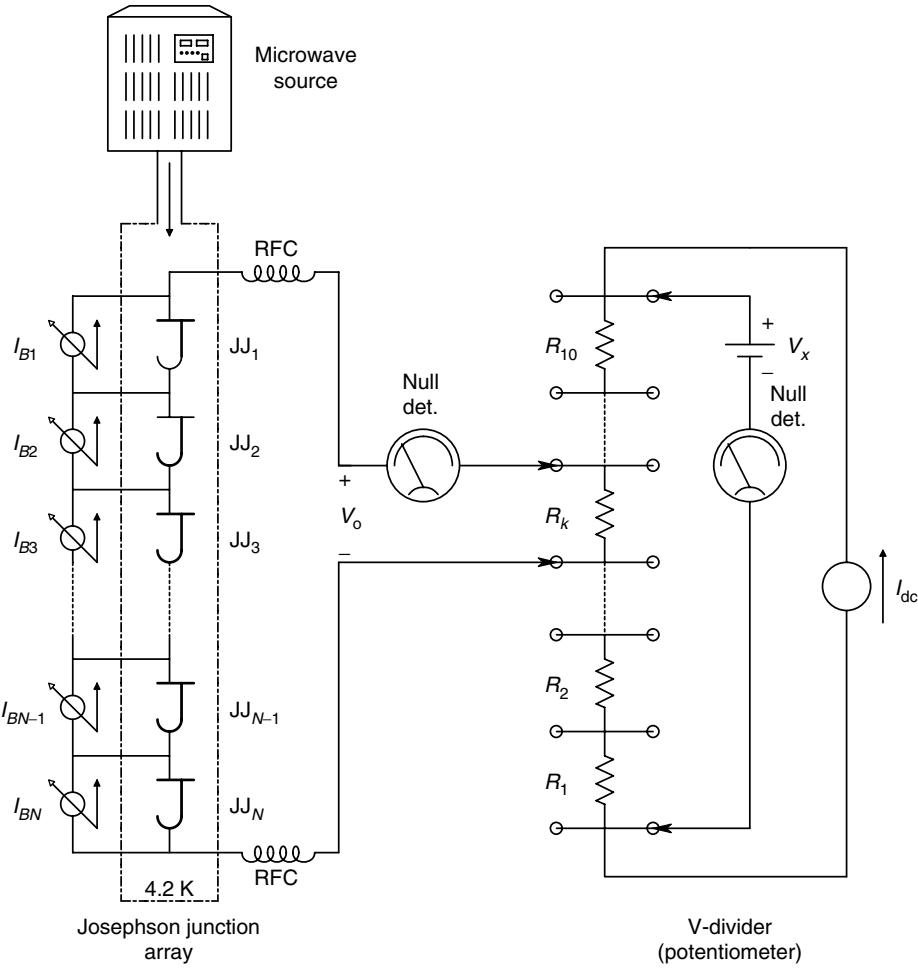
Figure 1.8 gives a simple schematic of a JJ array [Endo *et al*, 1983] which was used to calibrate other standard dc voltage sources and dc potentiometers. One of the problems in operating a JJ array is knowing which quantum step, n , the JJs are all on. One possible way of solving this problem is to obtain a stable array output voltage, then vary the dc bias current until the array jumps to the next step level and then adjust the microwave frequency down to f_2 until the same output voltage is obtained. Using this protocol, we can solve for n :

$$nf_1 = (n + 1)f_2 \quad (1.43)$$

so

$$n = \text{INT}[(f_1 - f_2)/f_2] \quad (1.44)$$

One can also simply count the steps as the bias current is slowly increased under conditions of constant microwave frequency and power.

**FIGURE 1.8**

Schematic diagram (simplified) of a precision potentiometer used to measure unknown EMFs. The potentiometer is calibrated with the standard voltage from the Josephson Junction array by adjusting I_{dc} .

The output of an array of N JJs in series, all having the same dc bias current and subject to the same microwave frequency and energy is:

$$V_o = N E_J(n, f) \quad (1.45)$$

Although the microwave frequency is the critical input parameter to a JJ array voltage standard, the microwave voltage and dc current must be kept within certain bounds for stable operation of the JJ array as a quantized voltage source. Hamilton *et al* (1991) have found relations for critical values of the dc current density, J_C , the length, l , and width, W , of a JJ for maximum stability of the voltage steps (to prevent noise induced jumps between steps). They have also given an empirical relation for the current range for the n th step, under optimum conditions. They have designed JJ arrays having stable operation at 24 GHz, producing up to 1.2 V.

Most of the Josephson array voltage standards which have been described in the literature (e.g. a number of papers on JJ standards in the April 1991 *IEEE Transactions on Instrumentation and Measurement*, 40(2)) have uncertainties around 2×10^{-8} .

TABLE 1.1

Specifications of PREMA® JJ primary Voltage Standard Chips

Parameter	1 V Chip	10 V Chip
Chip area	$19 \times 5 \text{ mm}^2$	$25 \times 11 \text{ mm}^2$
Number of JJs	2,400	13,920
Operating temperature	4.2 K	4.4 K
Operating frequency	70–75 GHz	70–75 GHz
Lead and bond wire resistance	<3 Ω	<3 Ω
Technology	Refractory, all niobium full wafer process	Refractory, all niobium full wafer process

Reymann (1991) describes a JJ array system which is used to measure Weston normal cell EMFs to $>1 \text{ nV}$ accuracy. A problem with the operation of JJ array voltage standards is that the noise induced instability of the voltage steps leads to jumps to unknown or uncertain step numbers. It is apparent that the JJ array provides an accurate, flexible, dc voltage standard which can be easily adapted for the calibration of secondary standards and dc voltmeters.

Primary Josephson voltage standards are now commercially available. PREMA® Semiconductor offers DC voltage standard systems based on their JVS-7001 and JVS-7010JJ chips. The specifications of these chips are given in Table 1.1.

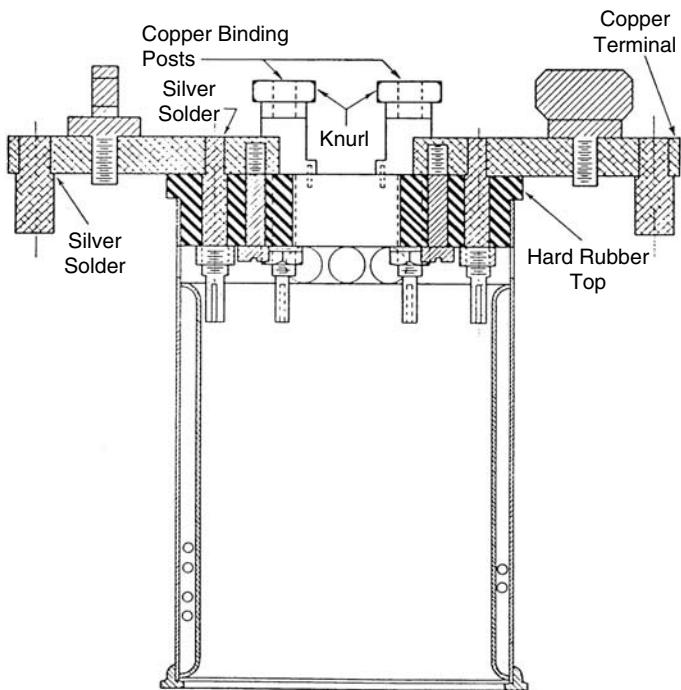
The Sandia National Laboratories' Primary Standards Lab has also developed a 10 V JJ Primary standard, and lists its accuracy as $\pm 0.017 \text{ ppm}$ (www.sandia.gov/pls/DC.HTML).

1.4.1.2 Resistance

In 1884, an international congress in Paris adopted a reproducible standard for the *legal ohm*, consisting of a column of ultra-pure mercury, 106 cm in length, and 1 mm^2 in cross-sectional area, measured at 0°C. In 1893, the Chicago Congress specified the length of mercury column of the *international ohm* to be 106.3 cm and specified the weight of the mercury to be 14.4521 g. The mercury ohm proved to be a poor standard. Mercury has a high temperature coefficient, requiring close thermoregulation of the standard in an ice bath. Slight impurities in the mercury also contributed to errors in certainty and variations in the design of the end terminals caused errors in the resistance. According to Stout (1950), another major cause of error in the design of the international ohm was nonuniformity in the diameter of the glass tubing. It was found that a 30 ppm change in the measured resistance could result from an undulation in the axis of the glass tube too small to be detected by eye.

In the 1920s, work was done on defining an *absolute ohm* as a derived quantity based on the fundamental units of length, mass and time. The absolute ohm was determined through the use of a rotating commutator or conductor, whose speed must be known, thus giving the time dependency. The determination also used either a self or mutual inductance, whose value was determined by calculations involving the dimensions. Not surprisingly, the value of the absolute ohm was found to differ significantly from that of the international ohm. One absolute ohm was equal to 0.9995052 international ohms. The absolute ohm and other absolute electrical units were formally adopted on 1 January 1948.

After 1892, working standard, wirewound one ohm resistors, made of the alloy *manganin* were developed, and in 1897, were calibrated in terms of the mercury ohm

**FIGURE 1.9**

Cross-section of a Thomas type standard resistor. (Source: Stout, Figure 4.2, 2nd ed., 1960.)

as defined in 1893 by the Chicago Congress. The U.S. National Bureau of Standards (NBS) adopted the Rosa design, one ohm, manganin, wirewound resistor in 1910 as the international ohm. In 1931, an improved standard resistor, designed by J.L. Thomas, was adopted at NBS. The cross-section of a Thomas type, one ohm standard resistor is shown in Figure 1.9. This resistor uses a double walled construction, with the resistance wire wound on the inner cylinder. The space between the cylinders is filled with dry air and sealed. The center cylinder is filled with oil for temperature stabilization and control. In the Rosa type resistor, the resistance wire windings are in direct contact with dry oil in a sealed can. Rosa resistors are available in decade sizes, typically 1, 10, 100, 1k and 10k ohms.

In 1966, Hewlett-Packard (HP) developed an improved standard resistor design, their 11100 series, which achieved superior temperature stability and precision through the use of new materials for insulation, resistance coil support and the resistance wire. HP used the alloy Evanohm for the coil (75% nickel, 20% chromium, 2.5% aluminum, 2.5% copper). According to HP, Evanohm has an α tempco of 0 to $+2 \text{ ppm}/^\circ\text{C}$, high resistivity, relatively flat resistance *vs* temperature curve and is insensitive to moisture. The Evanohm wire coils of the HP 11100 resistors are suspended in sealed, dry oil and supported by an inner and outer polyester coil form, rather than a metal form such as that used in the Rosa or Thomas resistors. HP claims the following specifications for its 11100 series resistors—LE at 25°C and 100 mW power dissipation is $\pm 6 \text{ ppm}$ (0.0006%) with reference to the legal ohm (1966) maintained by NBS; rated tempco is $\pm 4 \text{ ppm}$; drift per year is $< 20 \text{ ppm}$ ($100 \text{ k}\Omega$); power rating is 100 mW; internal trim resistance permits adjustment of $\pm 25 \text{ ppm}$. Connections are NBS type, 4-terminal, oxygen-free copper, nickel-rhodium plated.

Note that the HP 11100 standard resistors can be set to within $\pm 0.15 \text{ ppm}$ of their nominal (true) resistance, but the 1966 NBS primary standard has an uncertainty of $\pm 6 \text{ ppm}$.

At the present time, the Fluke 742 series, 4-terminal, working standard resistors are available in values ranging from $1\ \Omega$ to $-19\ M\Omega$. The Fluke 742 series resistors are used at ambient room temperatures, from 18 to -28°C , with resistance changes ranging from ± 1.5 (10k) to ± 4 (19M) ppm, showing excellent temperature stability—no oil baths are required. Calibration uncertainty at 23°C ranges from ± 1.0 ppm (for 1 – $100\ \Omega$ sizes) to ± 10.0 ppm ($10\ M\Omega$) and ± 20.0 ppm ($19\ M\Omega$). Stability over one year ranges from ± 8.0 ppm (1 – $10\ \Omega$) to ± 10.0 ppm ($19\ M\Omega$).

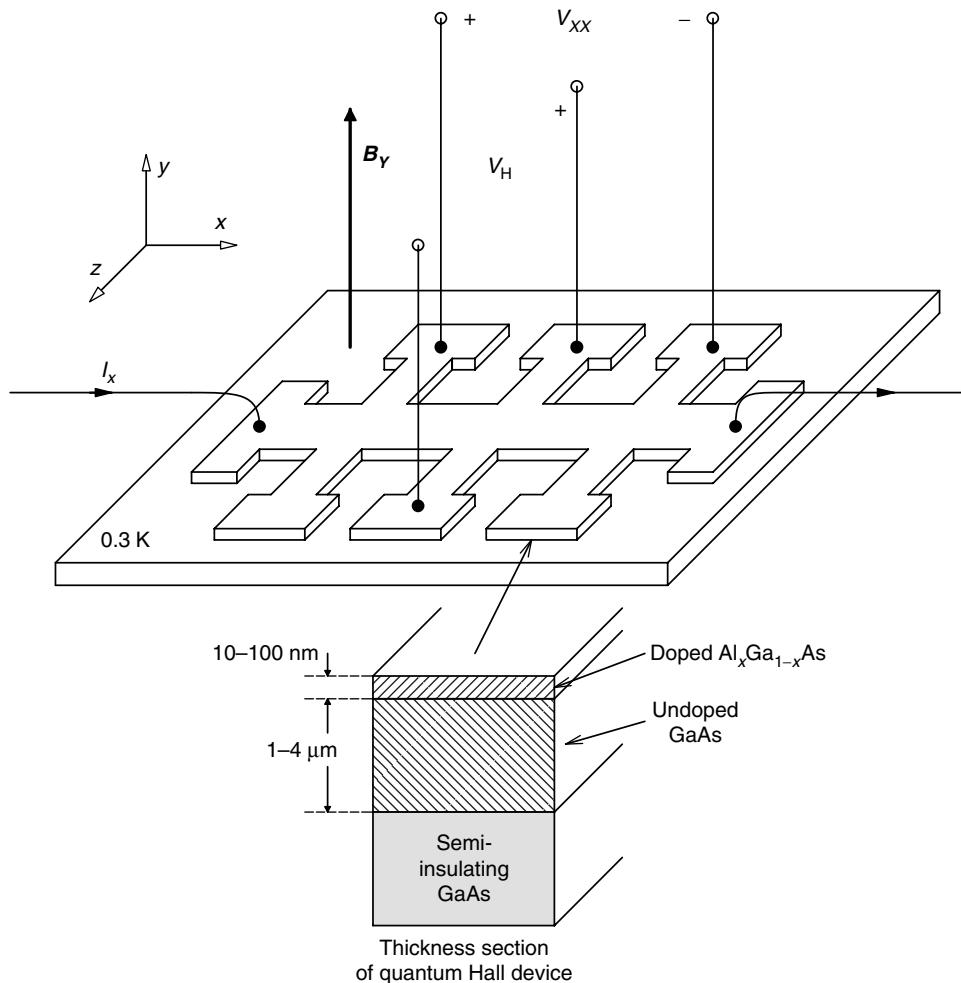
Electro Scientific Industries (ESI) also makes transfer standard resistors. The ESI model SR1060 resistance transfer system consists of six transfer standards in decades from $1\ \Omega$ – $100\ k\Omega$. Each decade standard is composed of 12 nominally equal resistors, matched to within ± 10 ppm. Each decade standard can be configured as 10 resistors in series, 10 parallel resistors or nine of the 10 resistors in series/parallel. By making a $1:1$ comparison with the 10th resistor, it is possible to resolve a series/parallel value to better than ± 1 ppm. All the standards, except the $100\ k\Omega$ standard, are immersed in an oil bath for temperature stability. The ESI SR1060 resistance transfer standard system is initially calibrated to ± 10 ppm, traceable to the NIST primary standard. Long term stability is ± 35 ppm for one year and tempcos are ± 15 ppm/ $^\circ\text{C}$ for $1\ \Omega$, and ± 5 ppm for $100\ \Omega$ – $100\ k\Omega$.

On 1 January 1990, the international instrumentation and measurement community adopted a standard for the ohm based on the *Quantum Hall Effect* (QHE). This definition of the ohm, like the JJ definition for the volt, is based on fundamental physical constants, rather than an artifact (such as a column of mercury). The QHE was first described by the noted physicist, Klaus von Klitzing in 1980. The operation of a basic Hall effect sensor at room temperature is described in Section 6.3.5.2 of this book. The ordinary Hall sensor is a 4-terminal device; one pair of terminals is used to inject current into the thin, doped semiconductor bar which is ordinarily perpendicular to a magnetic field, B . The second pair of terminals, as shown in Figure 6.22, is used to pick off the Hall EMF. It can be shown that the Hall EMF is given by:

$$E_H = R_H B_y I_x / h \quad (1.46)$$

where B_y is the B field component in the y direction, orthogonal to the current density vector, J_x , I_x is the injected current, h is the thickness of the doped semiconductor bar in the y direction and R_H is the Hall coefficient. $R_H = -1/qn$ for n -doped semicon, and $1 qp$ for p -doped semiconductors. The q is the magnitude of the electron charge in coulomb, n is the electron donor doping density, and p is the hole donor doping density. The Hall sensor was found to be useful in measuring magnetic fields, and indirectly, electric power. As will be illustrated below, the properties of the Hall sensor change markedly as its temperature approaches 0K and when it is subjected to strong, dc magnetic fields.

Figure 1.10 illustrates an isometric view of a quantum Hall resistor (QHR). Modern QHRs are generally fabricated from GaAs/AlGaAs heterostructures. The channel width may range from 150 to $250\ \mu\text{m}$ and the distance between lateral voltage contacts on the same side may range from 300 to $600\ \mu\text{m}$ [Piquemal *et al.*, 1991]. The thickness of the aluminum-doped conducting channel may range from 10 to $100\ \text{nm}$ [von Klitzing, 1986]. The QHR has six or eight terminals; two for current, two for voltage, and two or four for measuring body resistivity or voltage in the x direction. Under normal operation, the QHR is operated at a constant dc current bias, I_x , with values typically ranging 10 – $50\ \mu\text{A}$. A very strong, dc magnetic field, B_y , of the order of 4 – $15\ \text{T}$, is generated by a superconducting coil in close proximity to the QHR.

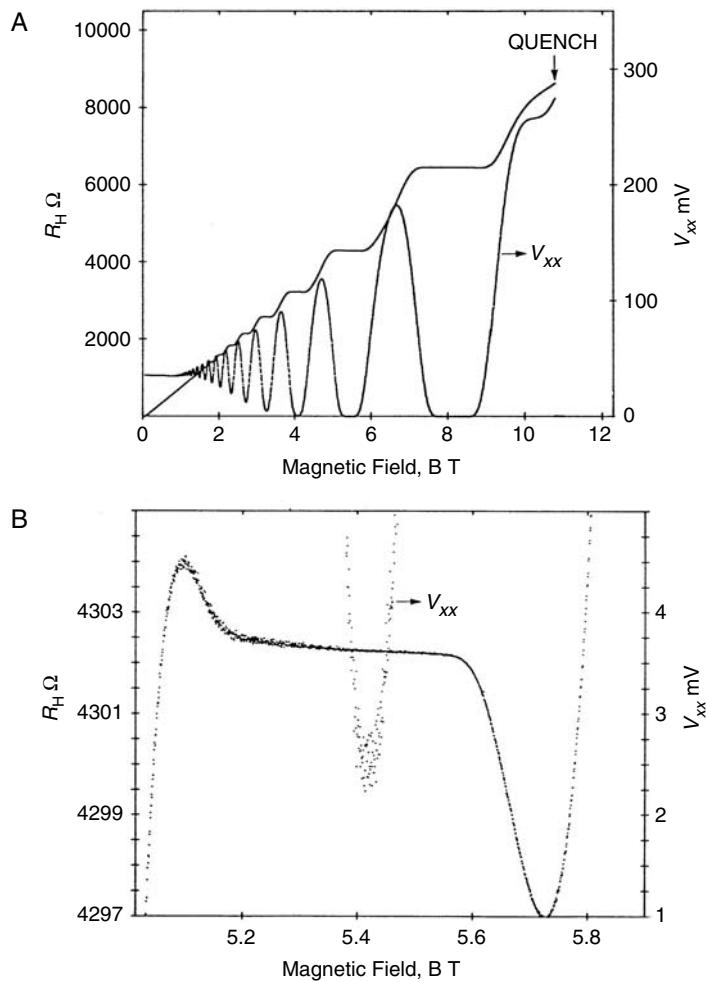
**FIGURE 1.10**

Isometric view of a Quantum Hall Device (adapted from von Klitzing, 1986). V_H is the Hall EMF, V_{xx} is the voltage drop due to the DC current, I_x , B_y is the strong, dc magnetic field. A gallium-aluminum-arsenide heterostructure is used. Operating temperature is typically 0.3 K.

A QHR is generally operated at temperatures ranging from 0.3 to 2 K, with 1.20 K being commonly used. As can be seen from Figure 1.11A, as the magnetic field is varied under conditions of constant dc bias current, the Hall resistance assumes a series of steps of precisely known values, regardless of the exact B_y , I_x , or temperature and material of the QHR semiconductor. The quantum Hall resistance, R_H , has been shown by von Klitzing to be given by:

$$R_H(j) = E_H/I_x = R_{K-90}/j = h/(q^2 j) \Omega \quad (1.47)$$

where h is Planck's constant, q is the electron charge, j is the step number (integer), R_{K-90} is the *von Klitzing constant* (universally adopted as 25,812.807 Ω on 1 January 1990),

**FIGURE 1.11**

A. Plot of the quantum Hall resistance, R_H and V_{xx} vs B_y . Note that there are plateaus of $R_H = E_H/I_x$ over a range of B_y on which R_H is precisely defined. In this case, $T=0.3\text{ K}$ and $I_x=50\mu\text{A}$. B. Detail of the $i=6$ quantum Hall step. Note that there is a distinct minimum of V_{xx} at the flattest part of the R_H curve. Conditions are same as in Figure 1.11A. (Source: Jaeger *et al*, 1991.)

and E_H is the measured dc Hall voltage. For all QHR devices, the quantum Hall resistances are:

$$R_H(1) = 25,812.807 \Omega$$

$$R_H(2) = 12,906.400 \Omega$$

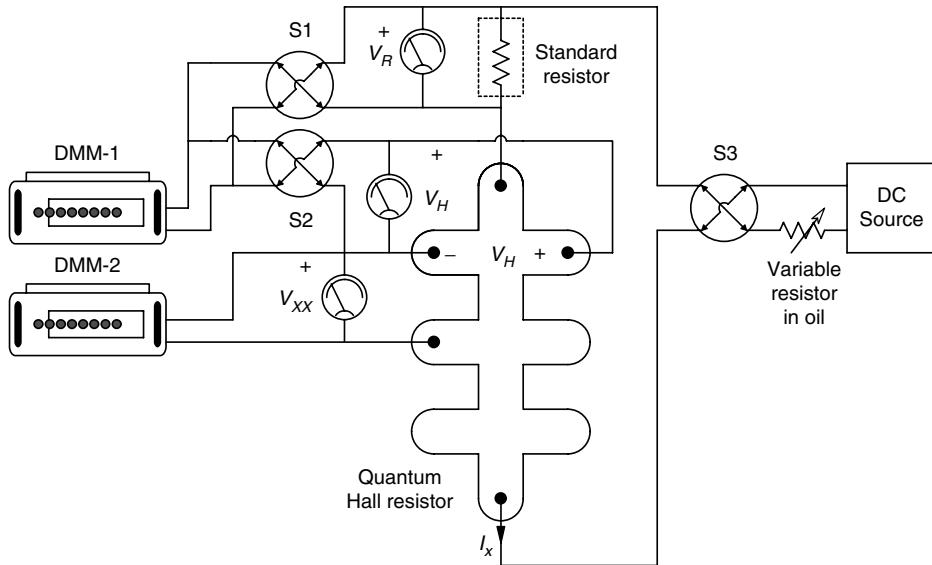
$$R_H(3) = 8,604.290 \Omega$$

$$R_H(4) = 453.2018 \Omega$$

$$R_H(5) = 5,162.5614 \Omega$$

$$R_H(6) = 4,302.1345 \Omega$$

The value of $R_H(1)$ is defined as one von Klitzing.

**FIGURE 1.12**

Schematic of a circuit used to compare a QHR with a secondary standard resistor. (Source: Jaeger *et al*, 1991.)

Also plotted in Figure 1.11A is the voltage drop in the x direction. Note that V_{xx} has minima which occur at the centers of the $E_H(B)$ or $R_H(j, B)$ steps. This is shown more clearly in Figure 1.11B, in which the value of V_{xx} is shown as B_y , which is varied over the range producing the $R_H(4)$ step [Jaeger *et al*, 1991]. Figure 1.12 illustrates the schematic of a simple series circuit which is used to compare the QHR with a secondary standard, using Ohm's law, precision dc nanovoltmeters and dc picoammeters.

Comparisons between quantum Hall resistance measurements made in various standards laboratories around the world suggest that the QHR method of defining the ohm has an uncertainty of less than one part in 10^8 .

1.4.1.3 Current and Charge

In a metallic conductor, such as a copper wire, the moving charge is electrons, whose average velocity is in the opposite direction to the direction of defined (in the U.S.) current flow. In semiconductors, two kinds of mobile charges are involved—holes and electrons. In p -semiconductor materials, the intrinsic semiconductor is doped with electron accepting atoms, and the majority charge carrier is holes. In n -material, electron donor atoms provide electrons as the majority carrier. In wet (chemical and biochemical) materials, the carriers are generally ions, although electrons are involved at metal/metal salt electrode interfaces.

In the SI system of measurements, the *ampere* is defined to be the basic unit of current and current is considered to be a fundamental unit of measurement. Before the SI protocols were adopted, current was considered to be a derived quantity, namely one coulomb of charge per second flowing past a perpendicular plane through a conductor. *Charge* was considered to be a fundamental quantity in the older, MKS system. In the SI system, charge is derived and has the units of current \times time. Therefore, one coulomb is the charge transferred when one ampere flows for exactly one second. Charge can also be measured by using the relationship that the voltage on a capacitor is equal to the charge divided by the capacitance, or that the change in voltage on a capacitor is equal to the change in charge divided by the capacitance.

Early definitions of the ampere made use of the electrolytic reduction of silver ions from a standard silver nitrate solution at a platinum cathode. Metallic silver is electroplated out on the cathode at a rate proportional to the electric (dc) current flowing in the external circuit. After a known time, the cathode is removed from the electroplating cell, dried and weighed. Since the gram molecular weight of silver is known, as is Avogadro's number, the accumulated weight of silver on the cathode is proportional to the total charge transferred in the circuit over the known time. At the London Conference of 1908, the *international ampere* was defined as that (dc) current which deposits silver at the rate of $0.00111800 \text{ g/second}$ from a standard AgNO_3 solution. This definition was based on a knowledge of the gram molecular weight of silver (107.868) and Avogadro's number, N_A , which is the number of molecules (hence Ag^+ ions) in one gram molecular weight of a substance. The value is presently known to be $6.022137 \pm 0.000007 \times 10^{23} \text{ mol}^{-1}$.

Due to the lack of precision and repeatability of the silver electroplating method, the international ampere was superceded in 1948 by the *absolute ampere*. The determination of the absolute ampere is made using a current balance which weighs the force between two current carrying coils. By definition, the SI absolute ampere is the constant (dc) current, which, if maintained in two, straight, parallel conductors of infinite length and negligible circular cross-section, placed one meter apart, *in vacuo*, will produce between these conductors a force of $2 \times 10^{-7} \text{ N/m}$ length. If the same current is traveling in the same direction in the two, parallel conductors, it may be shown that the force exerted between the conductors is attractive and is given by

$$F = \frac{\mu_0 I^2}{2\pi d} \text{ N} \quad (1.48)$$

where d is the separation of the conductors in meters, and μ_0 is defined as $4\pi \times 10^{-7} \text{ Wb/A m}$. Such a theoretical definition is unrealizable in practice.

Working standards to measure the absolute ampere generally involve a current balance, in which the force between a moveable coil or coils and a set of fixed coils or a permanent magnet is measured by a weighing balance. This type of determination of the ampere requires accurate measurement of force. The value of the mass used and g , the Earth's gravitational acceleration at the location of the balance, must be accurately known.

Kibble *et al* (1983) described the design and development of a moving coil apparatus for the determination of the ampere at the National Physics Laboratory (Britain). A large permanent magnet, with maximum flux density of about 0.7 T was used. Two series-connected, rectangular coils, each of 3,362 turns, were mounted one above the other on an armature which moves vertically in the magnet's air gap. The armature was attached by a vertical rod to the balance arm and passed coaxially through a force-producing solenoid coil at the bottom of the balance case. The current in the solenoid was controlled by a servo system so that it could generate axial forces to cause the measurement coils to move up and down through the magnetic field at a constant velocity, $u = 0.002 \text{ m/s}$. This motion generates an EMF in the series measurement coils, whose magnitude is given by $V = K_v u$ volts. The value of K_v depends on the value of B in the magnet's air gap, the number of turns and other geometrical factors. In the second phase of measuring the ampere, a dc current, I , is passed through the series measurement coils which are positioned at the center of the B field where the determination of K_v was made. A Lorentz force is generated, given by $F = K_F I$. F is measured with the balance. The value of I is

determined from the fact that in the MKS (SI) system, K_v is equal to K_F . If the relation $V = K_v u$ is divided by $F = K_F I$, we can obtain a relation equating electrical and mechanical power.

$$VI = Fu = Mgu \quad (1.49)$$

Now we assume that the generated electrical power is dissipated in a resistor R such that $V/R = I$. Thus

$$VI = I^2 R = Mgu \quad (1.50)$$

Solving for I , we find that

$$I = \sqrt{Mgu/R} \text{ SI amperes} \quad (1.51)$$

Although under development, Kibble *et al* reported that their moving coil current system for determination of the ampere had ± 1 ppm reproducibility and that they eventually expected to obtain 0.1 ppm accuracy.

A system utilizing the magnetic levitation of a superconducting mass to determine the SI ampere was proposed by Kibble (1983). This system is shown schematically in Figure 1.13. A current source, I causes current to flow in a superconducting coil surrounding a superconducting mass. Induced current in the mass produces a magnetic field which reacts with the field of the coil to produce an upward force until the mass reaches a position Y_1 , where the downward gravitational force, Mg , equals the upward levitation force, $F_L = I(d\Phi/dy)$. Now, if the mass is given a vertical velocity,

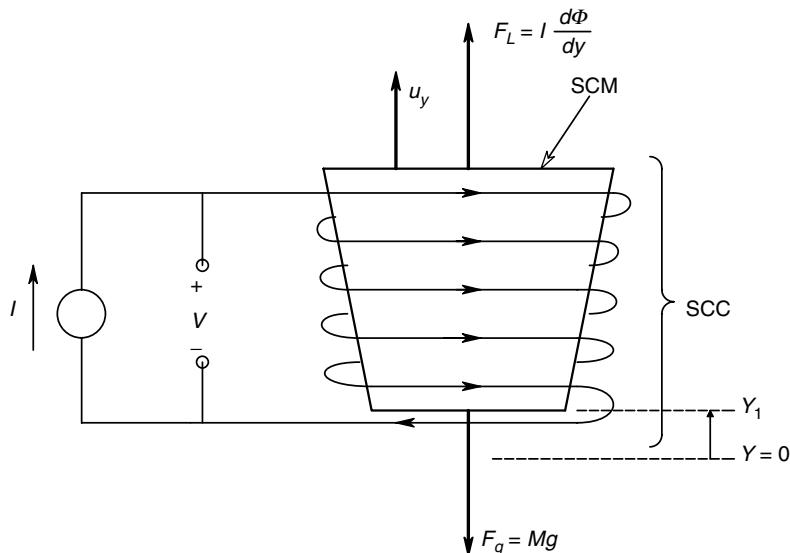


FIGURE 1.13

Schematic of a superconducting, mass balance means of measuring current proposed by Kibble (1983). Note: SCC = superconducting current carrying coil. SCM = superconducting mass, M . Means of inducing constant velocity, u_y , not shown. The induced voltage, V , is measured potentiometrically and is related to the current by equation 1.52.

$u_y = dy/dt$, around Y_1 , a voltage, $V = d\Phi/dt$, will be induced in the superconducting coil across the current source. We also note that $d\Phi/dt = (d\Phi/dy)(dy/dt)$. Now, from these relations, it is easy to see that

$$Mg/I = d\Phi/dy = V/u_y$$

Solving this relation for I gives:

$$I = (Mgu_y)/V \quad (1.52)$$

Kibble's proposed method of measuring dc current thus requires accurate knowledge of the mass, M , and the local acceleration of gravity, g . A constant, linear velocity of the mass, u_y , around its equilibrium displacement, Y_1 , must be generated, either by mechanical means or by electrostatic forces. A no touch means of measuring u_y could be designed using laser interferometry. Lastly, measurement of V can be done potentiometrically, by comparison with a Josephson voltage standard.

The reader should be aware that the bases for electrical and physical standards are constantly under development and are changing with advances in quantum physics, laser technology, etc. In metrology, we expect to see a trend away from complex, expensive, ultra-precision measurement systems towards simpler, more elegant solutions for the determination of standard quantities. Perhaps we should consider making the quantum Hall ohm and the Josephson volt the basic SI quantities, rather than the ampere.

Is charge a more easily measured quantity than current? Should we return to the earlier MKS system, where charge was a basic quantity, rather than current, which is used in the present SI definitions? There are no simple answers to these questions and the points of debate change constantly with evolving technology.

1.4.1.4 Capacitance

There are several ways of defining the basic unit of capacitance, the Farad. A capacitor is said to have a capacitance of one Farad if the potential across its plates is exactly one volt when it has been charged by one coulomb of electric charge. This definition of the Farad requires an accurate voltage measurement in which a negligible charge is exchanged with the capacitor, as well as an accurate measurement of the total charge put into the capacitor. The latter measurement is most difficult, since it involves precision integration of the capacitor's charging current.

Another way of measuring capacitance is through the use of a commutated capacitor bridge (cf Section 5.5.1.6 and Figures 5.9 and 5.10). This technique requires a precision dc null detector, accurate resistances and a switching clock with an accurately known period, T . Under the assumption that $R_4C_X \ll T/2 \ll R_2C_X$, we show in Section 5.4.1.6 that at null (set by R_1) is given by:

$$C_X = \frac{T(R_3 + R_1)}{4R_2(R_3 - R_1)} \quad (1.53)$$

Here, it is assumed that the switch has zero closed resistance and infinite open resistance, the waveforms at V_2 are exactly rectangular and triangular, and the clock duty cycle is exactly 50%. In practice, the duty cycle can be adjusted to be very close to 50%, taking into consideration the switching waveform rise and fall times, and the time delays in operating the MOS switch.

Probably, the most accurate means of measuring capacitance is by comparison of the capacitor to be measured with a *calculable capacitor*. Calculable capacitors are based on the electrostatics theorem developed by Thompson and Lampard (1956). A calculable capacitor is typically made from four parallel, hollow metal cylinders arranged in a square, with a movable shield device which effectively regulates the length of the active (exposed) cylinders. Capacitance is measured between opposite cylinders, *in vacuo*. Calculable capacitors have been made with capacitances ranging from 0.1 to 1.0 pF.

The Thompson and Lampard theorem states—*Let the closed curve S be the cross-section of a conducting cylindrical shell, whose cross-section has one axis of symmetry AC, but is otherwise arbitrary. Further, let this shell be divided into four parts by two planes at right angles, the line of intersection of the planes being parallel to the generators of the cylinder, and one of the planes containing the symmetry axis AC. Then the direct capacitance, per unit length of the cylinder, between opposing parts of the shell (for example, $\alpha\beta$ to $\gamma\delta$) due to the field inside (or outside) the shell, is a constant:*

$$C_o = \frac{\ln 2}{4\pi^2} \text{ e.s.u.} = 0.0175576 \text{ e.s.u./m} \quad (1.54)$$

Note that 1 pF of capacitance in SI units is equal to 0.898797 e.s.u. capacitance units. Thus, it is easy to show that $C_o = 1.953549 \text{ pF/m}$ (or $\text{aF}/\mu\text{m}$). A consequence of this theorem is that the absolute cross-sectional area of the closed cylindrical shell is not important. What is important is that the sides be perfectly parallel and that the structure has perfect symmetry. Figure 1.14A illustrates the geometry discussed in the theorem stated above and Figure 1.14B illustrates the shielded, four-tube, cross-sectional geometry of a practical calculable capacitor [Igarishi *et al*, 1968]. The length of the calculable capacitor at the U.S. NIST is measured interferometrically using a stabilized HeNe laser. Overall uncertainty in the measurement of the NIST Farad, based on the NIST calculable capacitor, is given as 0.014 ppm [Shields *et al*, 1989].

Calculable capacitor geometry has not been limited to four-tube devices. Delahaye *et al* (1987) have used five tubes in a pentagonal cluster, each 75.5 mm in diameter and 450 mm in length. A cylindrical central screen is displaced axially by a stepping motor and its

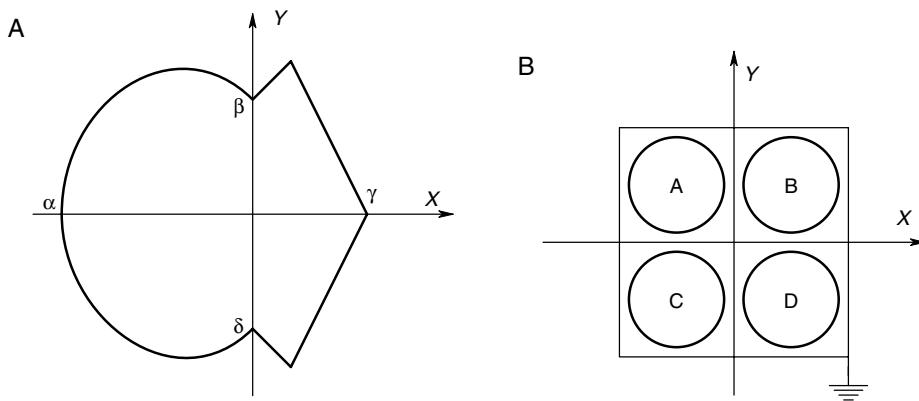
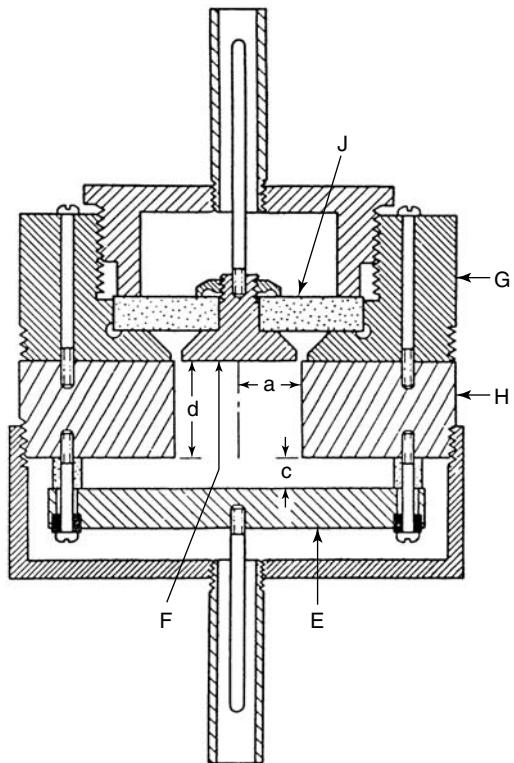


FIGURE 1.14

A. Cross-section of a symmetrical conducting tube relevant to the Thompson-Lampard theorem. The T-L theorem defines the capacitance between sections $\alpha\beta$ and $\gamma\delta$ on the tube as equal to $\ln(2)/(4\pi^2)$ e.s.u./m, regardless of the absolute cross-sectional area of the tube. B. Cross-section of a practical, four-tube, shielded calculable capacitor, such as that built by Igarishi *et al* (1968). Tube diameters were about 1 cm.

**FIGURE 1.15**

Cross-section through a secondary standard, air dielectric, guard-well capacitor. Note: F and E =insulated electrodes, J =borosilicate glass dielectric, G , H , K and L =grounded shield conductor pieces. Note that this capacitor is a three-terminal device; both top and bottom electrodes have capacitances to ground, as well as to each other. (Source: Stout, 2nd ed., 1960, with permission.)

position is measured with a laser interferometer. The calculable capacitor of Delahaye *et al* was used to establish a standard for the Farad and for the QHR, $R_H(2)$. The total one standard deviation uncertainty in the measurement of $R_H(2)$ was given as 2.2×10^{-7} .

Calculable capacitors have been used as standards by others to calibrate the quantum Hall ohm, inductors and transfer standard capacitors through the use of precision ac bridges [Shida *et al*, 1989; Dahake *et al*, 1983].

Secondary and transfer standard capacitors are generally of the order of 1–10 pF and are made with air, quartz (silica) or mica dielectrics. Such capacitors are generally made as three-terminal devices, so that the capacitance from either terminal to the grounded case can be compensated for. Air dielectric capacitors of the guard well geometry (Figure 1.15) have been constructed with values down to 1 fF. A 0.1 pF capacitor of this design has an accuracy of 0.1% [Stout, 1960]. For secondary standards with larger values, hermetically sealed, silvered mica capacitors offer long term stability and low tempcos. For example, a General Radio type 1409, 1 μ F standard silver mica capacitor has a guaranteed value up to $\pm 0.05\%$ and its tempco is $35 \pm 10 \text{ ppm}/^\circ\text{C}$. Silver mica capacitors have very low D values at 1 kHz. Most bridges used to transfer calibration from calculable capacitors to working and transfer standard capacitors are of the ratio transformer type, operated at frequencies of 2,500 and 5,000 r/s [Delahaye *et al*, 1987] and at 10,000 r/s [Igarishi *et al*, 1968; Dahake *et al*, 1983].

1.4.1.5 Inductance

The primary inductance standard is a derived quantity, measured using an ac bridge circuit and standard resistors and capacitors. We have seen how the quantum Hall ohm can be used to calibrate working standard resistors, and the calculable capacitor may be used to establish working standard capacitors. By using a bridge, such as the Maxwell

(cf Section 5.4.2.1), having three accurate resistors and an accurate capacitor, unknown inductances can be measured to a fair degree of accuracy.

Fixed value, standard inductors are generally wound in solenoidal form on glass or ceramic forms, having spiral grooves cut into their surfaces to precisely space the windings. Low value (μH) inductors have single windings, however, mH and larger inductors may have groups of series-connected windings set into evenly spaced slots in a ceramic core. Inductors with solenoidal windings are subject to the external influences of magnetic materials and time-varying magnetic fields. Therefore, care must be taken when using them not to place them on or near steel tables or cabinets, and to keep them well away from other current carrying solenoidal coils. Modern secondary standard inductors, such as the General Radio model 1490 series of switchable decade inductors, are wound on ferrite *toroidal cores*, from which there is negligible flux leakage and influence of nearby magnetic materials or time-varying magnetic fields. Another design which has been used for secondary standard inductors with fixed values in the mH range and larger, is the use of two series connected coils wound on semicircular, D-shaped forms, mounted so that one coil can be rotated with respect to the other in such a way that the area of overlap of the enclosed areas of each coil can be adjusted. This rotation allows variation of mutual inductances of the coils and provides a means of adjusting the series coil inductance in accordance with the relation:

$$L = L_1 + L_2 + 2M \quad (1.55)$$

assuming that the mutual inductance is aiding. The same principle of varying the mutual inductance of two series coils has also been used to make continuously variable inductors (Brooks inductometer). The Brooks inductometer's calibration is $\pm 0.3\%$ of its maximum value [Stout 1960].

A problem in the construction of all standard inductors is that their inductance depends on their dimensions, so that any small change in dimensions, caused by temperature changes, can lead to errors. Also, as we will see in Section 5.2, inductors, however made, are fraught with stray capacitance between their windings, and between layers and groups of windings. The wire from which they are wound also has a finite resistance. As the operating frequency of an inductor increases, the capacitive effects become more dominant and can cause significant errors in the apparent inductance. Thus all standard inductors are generally calibrated at a convenient frequency, such as 1 kHz, and are designed to be used at a constant temperature, such as 25°C. Inductors wound on laminated iron cores or sintered ferrite cores also have the additional limitations of core losses due to hysteresis, eddy currents and core saturation at magnetizing force (ampere turns).

It may certainly be said that of all the electrical quantities, natural inductance is the least pure, always being associated, at a given frequency, with a series resistance and distributed interwinding capacitance. Inductor equivalent circuits are described in Section 5.2 of this text. At superconducting temperatures, in air-core coils of certain metals, the coil resistance disappears, but the interwinding capacitance remains.

Another way of obtaining an extremely pure inductance, with values ranging from mH to kH in the audio frequency range, is by means of the *Generalized Impedance Converter* (GIC) active circuit element (Section 2.4.2.3). One end of the GIC circuit must be grounded. It can be shown that the GIC input impedance is given by:

$$Z_{11} = \frac{Z_1 Z_3 Z_5}{Z_2 Z_4} \quad (1.56)$$

Thus, if the circuit element for Z_2 is made a precisely known capacitor of high quality, such as a silver mica secondary standard, the remaining GIC impedance elements are resistors, calibrated from the quantum Hall ohm, and the op-amps are high quality ICs having large gain-bandwidth products, high input impedances and low noise, we will see a virtually lossless inductance, with value given by:

$$L_{eq} = \frac{R_1 C_2 R_3 R_5}{R_4} \text{ H} \quad (1.57)$$

It should be pointed out that the GIC circuit can also be used to rescale the size of capacitors. In this case,

$$Z_{11} = 1/\omega C_{eq}$$

It is easy to see that if Z_1 is a capacitor, then:

$$C_{eq} = \frac{C_1 R_2 R_4}{R_3 R_5} \text{ F} \quad (1.58)$$

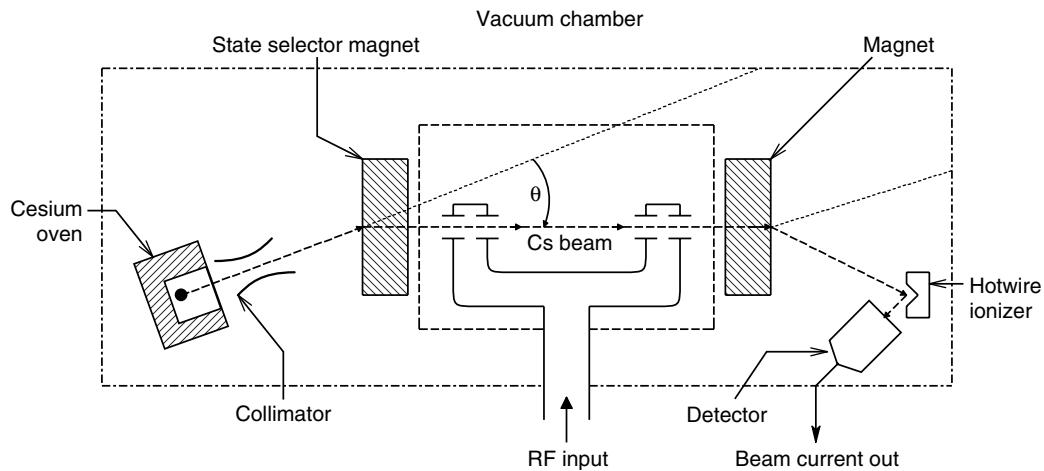
In using the GIC circuit, we must be careful to avoid op-amp voltage or current saturation, and to operate the circuit at frequencies well below the op-amps' f_T .

1.4.2 Time and Frequency

The basic SI and MKS unit of time is the second. The basic unit of frequency is the *hertz*, or cycle per second of a periodic waveform. Obviously, time and frequency are related, in a measurement sense. The earliest definitions of the second were based on the Earth's rotational velocity, which is now known to vary over the period of a year due to seasonal changes in the Earth's moment of inertia caused by the build-up of polar ice caps, etc. The *second* was first defined as 1/86,400 of the length of a mean solar day. The *mean solar day* is the average time of 365 consecutive days. Refinements in the measurement of the period of the Earth's rotation led to the definitions for *Universal Time* (UT). In a search for a more precise definition for the second, astronomers developed the units of *Ephemeris Time* (ET). The ephemeris second was defined as 1/31,556,925.9747 of the tropical year [Helfrick and Cooper, 1990]. The ET standard was impractical, as it took several years to determine the length of the tropical year and it required precise sightings of the positions of the sun and the moon.

The *primary time/frequency standard*, currently in use, is the cesium 133 atomic beam clock. A simplified diagram of the cesium clock is shown in Figure 1.16. To summarize the cesium clock's properties, it oscillates at 9.192631770 GHz, having an effective Q of 2×10^8 . Coordinated universal time (UTC), whose basic unit is the second, is defined as 9,192,631,770 periods of the cesium 133 beam oscillator. This international standard was adopted in October, 1967 and is more accurate than any clock calibrated by astronomical measurements [Oliver and Cage, 1971].

The Hewlett-Packard HP5051B (Opt 004) Cesium Beam Frequency Standard has a long term stability of $\pm 2 \times 10^{-12}$ over the life of the cesium beam tube. Accuracy is also $\pm 2 \times 10^{-12}$. The HP5051B cesium clock has sinusoidal outputs at 10, 5, 1 and 0.1 MHz. HP cesium clocks are used to calibrate and synchronize the LORAN-C and GPS radio navigation systems for boats and aircraft. Cesium beam clocks are ordinarily

**FIGURE 1.16**

Schematic cross-section of a Cesium ion clock.

used to adjust secondary rubidium and quartz oscillators, which are used as secondary standards for frequency or period determination.

Rubidium frequency standards are second in the hierarchy of accuracy. Similar to the operation of a cesium beam clock, the atomic resonance of a rubidium vapor cell is used to synchronize a quartz crystal oscillator in a frequency lock loop. The long term stability of the rubidium vapor oscillator is $\pm 1 \times 10^{-11}$ /month. It, too, has outputs at 5, 1 and 0.1 MHz.

1.4.3 Physical Standards

The physical standards discussed below include mass, length, volume and temperature. We have seen above that the present trend is to redefine standard quantities in terms of the (presumably) universal constants of quantum physics, rather than to use physical artifacts (e.g. the present use of the JJ to define the volt instead of the Weston normal cell).

1.4.3.1 Mass

The present (SI) unit of mass is the kilogram (kg). It is defined in terms of the standard kilogram artifact which has been kept at the *Bureau International des Poids et Mesures* (BIPM) at Sèvres, France, for over a hundred years. Physically, the standard kilogram is a cylinder of platinum containing 10% iridium. Comparison of a working standard kilogram mass with the primary standard must be made by means of a precision balance. The current gold standard balance is the NBS-2—single arm, two knife balance, which has a resolution of one μg or one part in 10^9 [Quinn, 1991]. It has been discovered over the years that the international prototype kilogram and its copies systematically loose mass when they are cleaned and washed in preparation for weighing. On the other hand, the masses tend to increase with time following cleaning and washing. These changes are of the order of tens of μg [Quinn, 1991]. The long term stability of the standard kilogram appears to be of the order of $5 \mu\text{g}/\text{year}$. This figure is based on the assumption that the international prototype and its copies are not drifting in mass by more than ten times the rate that they are drifting apart from each other, at about $0.5 \mu\text{g}/\text{year}$.

A number of workers are considering ways to replace the present artifact standard kilogram at BIPM with a mass standard based on the fundamental, physical constants which include the electron charge (q), Planck's constant (h), the speed of light *in vacuo* (c), and the permeability of vacuum (μ_0). Other constants which can be used are the von Klitzing constant (R_{K-90}), the Josephson constant (K_{J-90}), the fine structure constant (α), the Faraday number (\mathfrak{F}) and the Avogadro constant (N_A).

Taylor 1991 argued that the Avogadro constant, which is the number of molecules in a gram molecular weight, or mole, of some pure element (X), can be one determining factor, along with the molar mass of the element (M_X) in kg/mol. For example, $M(^{12}\text{C})$ is equal to 0.012 kg, by definition (^{12}\text{C} is the common isotope of the element carbon). Thus:

$$n_X = N_A / M_X \quad (1.59)$$

where n_X is the number of free Xs at rest required to make the mass of an international kilogram prototype. There are a number of interesting ways to determine the Avogadro number. Taylor 1991 proposed that N_A can be found from the X-ray crystal density of a pure, grown silicon crystal from the relation:

$$N_A = M(\text{Si}) / \left[\delta(\text{Si}) (d_{220} \sqrt{8})^3 / 8 \right] \quad (1.60)$$

where $M(\text{Si})$ is the mean molar mass of silicon in kg/mol, $\delta(\text{Si})$ is the crystal density in kg/m³ and d_{220} is the 2,2,0 silicon lattice spacing in meters. $M(\text{Si})$, $\delta(\text{Si})$ and d_{220} must be measured. In a practical determination of N_A , the ratio of the three naturally occurring isotopes (^{28}\text{Si}, ^{29}\text{Si}, ^{30}\text{Si}) in a crystal must be measured. Another way to find N_A indirectly is to use the relation:

$$N_A = K_{J-90} R_{K-90} \mathfrak{F} / 2 \quad (1.61)$$

where $K_{J-90} = (2q/h)$ is the Josephson constant, $R_{K-90} = (h/q^2)$ is the von Klitzing constant and the Faraday number, $\mathfrak{F} = 96,486.7$ (the coulomb charge of an Avogadro number of electrons).

Perhaps in the near future, progress will be made in replacing the standard kilogram artifact with a standard kilogram, defined by fundamental constants. Uncertainty in determining a new standard such as this may be of the order of 1×10^{-8} .

1.4.3.2 Length

The SI standard unit of length is the meter. Formerly, the meter was defined as one ten millionth of the arc distance of a meridian passing from the north pole through Paris, to the equator. It was represented by an artifact consisting of a platinum-iridium bar, with lines scribed on it one meter apart, kept at constant temperature at the BIPM near Paris. In 1960, the meter was redefined in terms of the wavelength of monochromatic light. One meter was exactly 1,650,763.73 wavelengths in vacuum of the orange radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the krypton-86 atom. However, this definition of the meter by wavelength assumes the permanence and invariance of the atomic energy levels, Planck's constant and the speed of light. These quantities are basic to the relation for wavelength:

$$\lambda = c/v = ch/E = ch/(E_2 - E_1) \quad (1.62)$$

To eliminate the requirement for h and ΔE , the invariant speed of light, c , was used in 1983, to redefine the standard meter as the distance that light travels in free space in 1/299,729,458 seconds (3.3363420 ns). Frequency stabilized laser sources are used to realize this definition [Sirohi and Kothiyal, 1991].

Gauge blocks are used as secondary standards for length. These are generally platinum alloy or stainless steel cubes or blocks, whose dimensions are established interferometrically to within a fraction of a wavelength of light, and whose faces are polished to optical smoothness. Gauge blocks are used to calibrate micrometers and to set up precision mechanical and optical systems.

1.4.3.3 Temperature

The SI unit of temperature is the kelvin. The standard reference temperature is defined by the *triple point* of water, at which the pressure and temperature are adjusted so that ice, water and water vapor exist simultaneously in a closed vessel. The triple point of pure water occurs at +0.0098°C (273.16 K) and 4.58 mmHg pressure. The kelvin was defined in 1967, at the thirteenth CGPM, to be the unit of thermodynamic temperature, equal to 1/273.16 of the thermodynamic temperature of the triple point of water. Other primary, fixed temperature points are also used for temperature calibration—the boiling point of O₂ (−182.97°C), the boiling point of sulfur (444.6°C), the freezing point of silver (960.8°C), and the freezing point of gold (1,063°C)—all at atmospheric pressure. Absolute zero is at 0 K, or −273.15°C. Clearly, Celsius and Kelvin temperature scales have the same slope.

Since so many physical and chemical phenomena are strong functions of temperature, there are many ways of measuring temperature, as will be seen in Chapters 6 and 7 of this text.

1.4.3.4 The SI Base Units

The fundamental SI units, determined by international treaty signed in 1960 are given in Table 1.2.

All other physical and electrical units are derived from the basic SI units (L, T, M, I, etc.). For example, charge, Q , has the dimensions of IT. Force in Newtons, F , has the dimensions of MLT^{−2}; this follows from Newton's $F=ma$. The volt (unit of potential difference) has the dimension of joule/coulomb; in SI units, the volt is ML²T^{−3}I^{−1}. The MKS units of the volt are ML²T^{−2}Q^{−1}. Current density, J , in amperes/meter has the SI dimensions of IL^{−1}; the MKS dimensions of J are QT^{−1}L^{−1}. Finally, the SI dimensions of the ohm are ML²T^{−3}I^{−2}.

TABLE 1.2

The Basic SI Units

Quantity	Unit	Approx. Uncertainty
Length	meter (m) {L}	3×10^{-11}
Time	second (s) {T}	1×10^{-13}
Mass	kilogram (kg) {M}	5×10^{-9}
Electric current	ampere (A) {I}	1×10^{-6}
Temperature	kelvin (K) {Θ}	2.5×10^{-4}
Luminous intensity	candela (cd)	1.5×10^{-2}
Amount of substance	mole (mol)	TBD

1.5 Chapter Summary

In this introductory chapter, we have introduced the architecture of typical measurement systems and we have shown that they not only contain sensors, signal conditioning operations and data display, storage and retrieval options, but also sources of noise, which can limit measurement precision. Measurement errors occur if sensor dynamics are not considered and the sensor is read before it reaches a steady state output, given a transient change in the value of the QUM.

The types of errors in measurements were considered. Elementary statistics were discussed and the important concept of limiting error in a measurement was introduced and derived. Examples were given of calculating limiting error.

Standards in electrical and physical measurements were described. The important SI standards for voltage (the JJ), resistance (the QHR) and capacitance (the Thompson-Lampard theorem) were discussed in detail.

The reader is advised that primary standards are constantly evolving with science and technology, and accuracies of one part in 10^7 are routine and in fact, many systems exceed this accuracy.

Problems

- 1.1 One of the earliest means devised for measuring moderate pressures was the U-tube, liquid filled manometer (Figure P1.1). Liquids used are typically mercury or water. The dynamics of a manometer can be described by a second order, linear ODE:

$$\Delta PA = \rho ghA + \left(\dot{h}/2\right)D(2\pi L) + \left(\ddot{h}/2\right)(\rho LA)$$

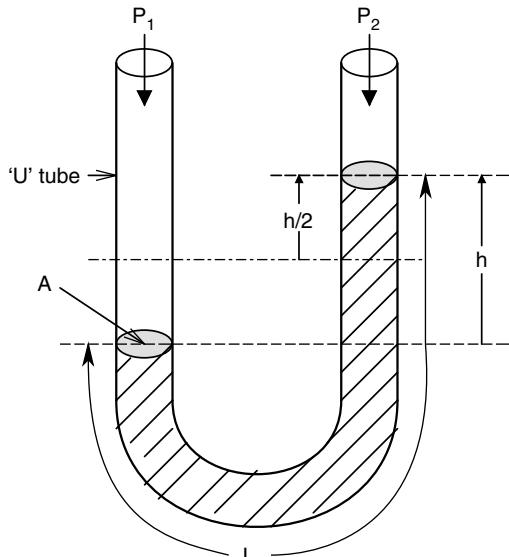


FIGURE P1.1

From Figure P1.1, it can be seen that L is the total length of liquid in the U-tube, A is the area of liquid exposed to pressures P_1 and P_2 ; $A = \pi r^2$; r = tube lumen radius. $\Delta P = P_1 - P_2$, ρ = liquid density in kg/m^3 , D = viscous damping constant, g = acceleration of gravity in m/sec^2 , h = height difference between the two sides of the manometer.

- Derive a transfer function for the manometer's dynamic response, $\frac{h}{\Delta P}(s)$ in time-constant form.
- Give expressions for the manometer's undamped natural frequency, ω_n and its damping factor, ξ .
- Calculate the numerical value for ω_n and D , given $\xi = 0.4$. Assume water is the liquid. Let $L = 1 \text{ m}$, $a = 3.142 \times 10^{-6} \text{ m}^2$.

1.2 An accelerometer has critically damped, second order, low-pass dynamics with the transfer function:

$$\frac{V_o}{\ddot{X}}(s) = \frac{K_A}{(\tau s + 1)^2}$$

where $\tau = 1 \text{ s}$, $K_A = 0.001 \text{ V}/(\text{m/s}^2)$.

Given a step input of acceleration at $t = 0$, find how long it will take $V_o(t)$ to reach: (A) 95% of its steady state value. (B) 99% of its steady state value. (C) 99.9% of its steady state value.

1.3 A piezoelectric pressure sensor and its signal conditioning amplifier have a transfer function approximated by:

$$\frac{V_o}{P}(s) = \frac{-10s}{s100 + 1}$$

The input is a 100 psi pressure step. (A) Sketch and dimension $V_o(t)$. (B) Find the peak $V_o(t)$. (C) Find the time for $V_o(t)$ to drop to 99.5% of its peak value.

1.4 An electronic temperature sensor converts temperature in kelvin to output microamperes. When it is immersed in water, its transfer function is:

$$\frac{I_o}{T}(s) = \frac{1.5 \times 10^{-2}}{(s + 0.3)(s + 0.05)} \mu\text{A/K}$$

(A) Sketch and dimension $i_o(t)$ when, at time $t = 0$, the sensor is taken out of a glass of 20°C water and immersed in a glass of 55°C water. (B) How long will it take $i_o(t)$ to reach 99.8% of its steady state value?

- The pressure sensor of Problem 1.3 is given an input of $p(t) = 100[1 + \sin(2\pi ft)]$ psi for a long time. Make a Bode plot of $dB = 20 \log_{10} |V_o/P(jf)|$ vs f (log scale).
- Calculate the net dc open circuit output voltage of a series array of $N = 550$ JJ at 4.2 K , given a series dc current of 24.00 mA and a microwave irradiation frequency of $f_o = 9.0646279 \text{ GHz}$. Operation is on the $n = 97$ th step. Give the output voltage to six significant figures.
- A QHR and a JJ array are put in series so that the dc current, I , that passes through the QHR also passes through the JJ array (Figure P1.7). The B field is adjusted on the

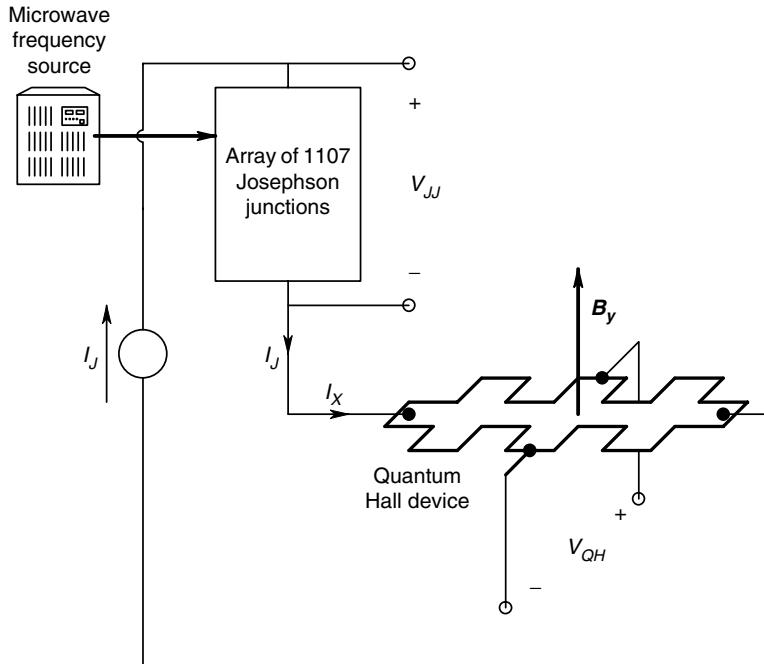


FIGURE P1.7

QHR so that it is on its $k=7$ th step. I_x is set at $150 \mu\text{A}$ and the JJ array is operated so that all the JJs are on step $n=1$. The microwave frequency irradiating the JJ array is adjusted to $f_o = 13.374664 \text{ GHz}$ to make $V_{QH} = V_{JJ}$.

- (A) Under the condition of $V_{QH} = V_{JJ}$, find a general expression for I_x in terms of q, h, n, k, f_o and N .
 - (B) Find the numerical value of the number of JJs, N , in the JJ array required to meet the conditions above.
 - (C) Find the numerical value of $V_{QH} = V_{JJ}$ to six significant figures. Note that the electron charge, $q = 1.60217733 \times 10^{-19} \text{ C}$, Planck's constant $h = 6.626\ 075\ 5 \times 10^{-34} \text{ Js}$.
- 1.8 Use MicroCap or PSPICE to simulate the frequency response of a GIC synthesized 1.000 H inductor. The schematic is shown in Figure P1.8. Note that the accuracy of the simulated inductor will degrade at high frequencies because of the finite gain-bandwidth and slew rate of the non-ideal (real) op-amps used. You are to use three $10,000 \Omega$ and one $1,000 \Omega$ resistors, a $0.001 \mu\text{F}$ silver mica capacitor and two OP-37 op-amps. The op-amp parameters are: $R_{in} = 6 \text{ M}\Omega$, $A_{OL} = 1.8\text{E}6$, $R_{out} = 70 \Omega$, $f_1 = 44.4 \text{ Hz}$, $f_2 = 3.7\text{E}7 \text{ Hz}$, $\eta = 2.0\text{E}7 \text{ V/sec}$, $V_{OS} = 10 \mu\text{V}$, $I_B = 10 \text{ nA}$, $I_{OS} = 7.0\text{E}-7$, ($f_T = 80 \text{ MHz}$). The VCCS is used to make simulation with MicroCap easier; note that $Z_{11} \equiv V_1/I_1$.
- (A) Make Bode and magnitude and phase plots of $Z_{11}(jf)$. Indicate the range of frequencies over which the GIC inductor simulation is valid.

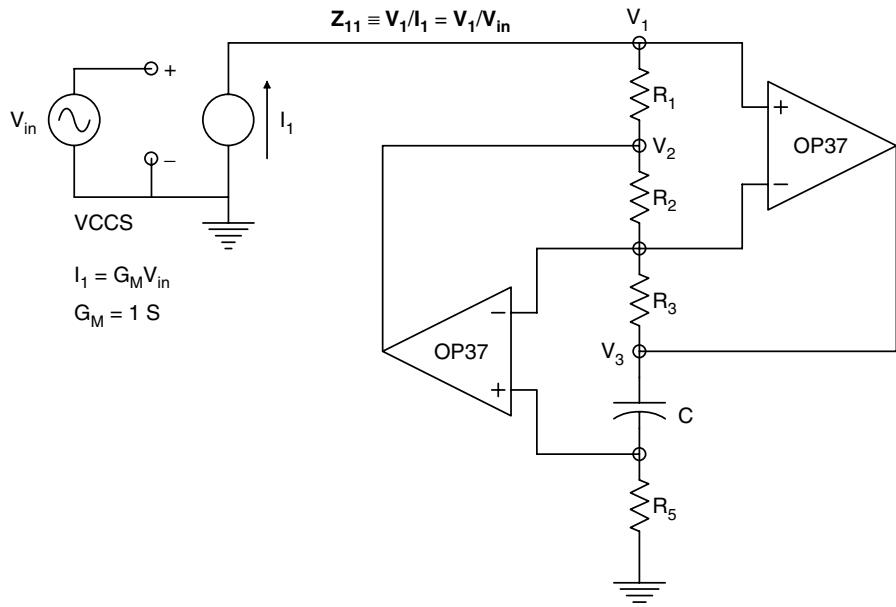


FIGURE P1.8

- (B) Find the frequency where the phase of $Z_{11}(j\omega)$ departs from 90° by $\pm 0.1^\circ$. What is the equivalent resistance that appears in series with the simulated inductor at this frequency?
 (C) Give the simulated inductor's Q at this frequency.
- 1.9 Find the limiting error (LE) in the mid-frequency L_{eq} synthesized by the GIC circuit of Figure P1.8. The inductance can be shown to be $L_{eq} = C_4 R_1 R_3 R_5 / R_2$. C_4 is known to 0.05%, R_2 to 0.01% and R_1 , R_3 , and R_5 to 0.002%.

2

Analog Signal Conditioning

2.1 Introduction

Practically all instrumentation systems require some type of analog signal conditioning between the analog input transducer and the data display, processing and storage systems. In its simplest form, analog signal conditioning can be voltage amplification, with a change in impedance level between the conditioning amplifier's input and output.

Analog signal conditioning may also involve linear filtering in the frequency domain, such as bandpass filtering to improve signal-to-noise ratio (SNR) at the amplifier's output. In other cases, the analog input to the signal conditioning system may be processed nonlinearly. For example, depending on system requirements, the output of the analog signal conditioner may be proportional to the square root of the input, to the RMS value of the input, to the logarithm of the input, or to the cosine of the input, etc.

Analog signal conditioning is often accomplished by the use of the ubiquitous operational amplifier, as well as special instrumentation amplifiers, isolation amplifiers, analog multipliers, and dedicated nonlinear processing ICs.

In the following sections of this chapter, we examine the properties of the analog integrated circuit systems used in analog signal conditioning in instrumentation systems, beginning with the differential amplifier which has wide use in the headstages of nearly all types of op-amps, instrumentation amplifiers, isolation amplifiers, analog multipliers, oscilloscopes etc.

2.2 Differential Amplifiers

Differential amplifiers (DAs) are widely used as input stages in a variety of amplifier types, including op-amps, analog comparators, analog multipliers, instrumentation amplifiers, CRT oscilloscope vertical and horizontal amplifiers and certain specialized IC's. There are two good reasons for their wide use in signal conditioning system design. The first and foremost is the ability of the DA to respond to the difference in the input signals ($V_1 - V'_1$) and to discriminate against noise or interference that may be common to both inputs. Such a common-mode (CM) input voltage is often a dc level hum, or other coherent interference, which is desired to be eliminated. Another reason for using a DA headstage is that it inherently discriminates against changes in the amplifier's dc power supply voltages.

2.2.1 Analysis of Differential Amplifiers

Figure 2.1 illustrates the most general form of differential amplifier. Note that including ground, it is a four-port circuit. Most practical DAs have only a single-ended output, V_o , and are thus three-port circuits. The deflection amplifier circuits of most analog oscilloscopes are an exception to this rule, preserving the differential circuit architecture from the input to the output, to the CRT deflection plates.

In the analysis of DAs, it is expedient to define *difference-mode* and *common-mode* input and output signals.

$$V_{id} \equiv (V_i - V'_i)/2 \quad (2.1)$$

$$V_{ic} \equiv (V_i + V'_i)/2 \quad (2.2)$$

$$V_{od} \equiv (V_o - V'_o)/2 \quad (2.3)$$

$$V_{oc} \equiv (V_o + V'_o)/2 \quad (2.4)$$

The input/output relationships for a general DA can be expressed in terms of the Middlebrook equations:

$$V_{od} = A_{dd} V_{id} + A_{dc} V_{ic} \quad (2.5)$$

$$V_{oc} = A_{cd} V_{id} + A_{cc} V_{ic} \quad (2.6)$$

The four gains of the voltage controlled voltage sources (VCVSs) in equations 2.5 and 2.6 are written in vector format to stress that they are functions of frequency ($j\omega$).

It is easy to show that the single ended outputs are given by:

$$V_o = (A_{dd} + A_{cd}) V_{id} + (A_{dc} + A_{cc}) V_{ic} \quad (2.7)$$

$$V'_o = (A_{cd} - A_{dd}) V_{id} + (A_{cc} - A_{dc}) V_{ic} \quad (2.8)$$

The single-ended output, V_o , can also be written in terms of the actual DA's inputs:

$$V_o = \frac{A_{dd} + A_{cd} + A_{dc} + A_{cc}}{2} V_i + \frac{A_{cc} + A_{dc} - A_{dd} - A_{cd}}{2} V'_i \quad (2.9)$$

Often, it is expedient to write:

$$V_o = A_D V_{id} + A_C V_{ic} \quad (2.10)$$

where A_D and A_C are given by the corresponding terms in equation 2.7. In an ideal DA, $A_{cc} = A_{dc} = A_{cd} = 0$

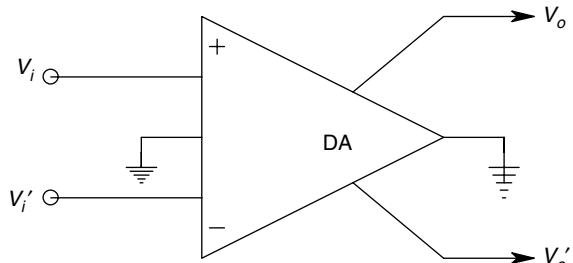


FIGURE 2.1

Generalized differential amplifier with differential output. Note that is a four-port circuit.

So, in equation 2.10 above, $A_C = 0$ and $A_D = A_{dd}$.

2.2.2 Common-Mode Rejection Ratio

The common-mode rejection ratio (CMRR) is a figure of merit for DAs, usually given in decibels. It describes how well the behavior of a real differential amplifier approaches that of an ideal DA. The scalar CMRR, measured at some frequency, is defined as:

$$\text{CMRR} \equiv (|V_{ic}| \text{ to give a certain output magnitude}) / (|V_{id}| \text{ to give the same output}) \quad (2.11)$$

Using the definition for CMRR and equation 2.7, we find:

$$\text{CMRR} = |(A_{dd} + A_{cd}) / (A_{cc} + A_{dc})| \quad (2.12)$$

If the DA circuit is perfectly symmetrical, then $A_{dc} = A_{cd} = 0$, and the CMRR magnitude at a given frequency reduces to:

$$\text{CMRR}(f) = |A_{dd}/A_{cc}| = |A_D + A_C| \quad (2.13)$$

State-of-the-art instrumentation and isolation amplifiers may have CMRRs of 120 dB or more ($> 10^6$), under ideal conditions. Generally, CMRR decreases with frequency, and as we will see, an imbalance in Thevenin source resistances at the input of a DA can lead to either a loss or increase in CMRR over the manufacturer's specified value.

2.2.3 Measurement of CMRR, A_D and A_C

Often, in order to save money, an engineer puts together a differential instrumentation amplifier from op-amp 'building blocks'. (The design of such amplifiers is covered in Section 2.5.1). In order to measure the CMRR of the DA, we make use of equation 2.11. The two inputs are shorted together and a sinusoidal signal of frequency f and amplitude V_{s1} is applied to them, generating a CM input signal, V_{s1c} which is adjusted to give an output, V_o . Next, the negative input is grounded, so $V'_i = 0$. The sinusoidal signal source of frequency f and amplitude V_{s2} is now connected to the + input terminal of the DA. This procedure is necessitated because of the difficulty in generating a perfectly balanced DM input signal. The single-ended input generates both DM and CM signals.

$$V_{id} = V_{s2}/2 \text{ and } V_{1c} = V_{s2}/2$$

V_{s2} is adjusted to make $V_{o2} = V_o$ of the CM case. Thus, from equation 2.10,

$$A_C = V_o/V_{s1} \quad (2.14)$$

and also, considering the single-ended input, V_{s2} , we have:

$$V_o = A_D V_{s2}/2 + A_C V_{s2}/2 \quad (2.15)$$

If equation 2.14 for A_C is substituted into equation 2.15, we may solve for the DM gain for A_D as:

$$A_D = V_o(2/V_{s2} - 1/V_{s1}) \quad (2.16)$$

Hence using relations 2.14, 2.16 and 2.13, the CMRR that is experimentally found can be expressed as equation 2.17:

$$\text{CMRR} = A_D/A_C = (2V_{s1}/V_{s2} - 1) \quad (2.17)$$

if the amplifier's scalar A_D is very large, a precision attenuator may be needed to reduce the output of the signal source to an appropriately small V_{s2} .

2.2.4 Effect of Source Resistance Asymmetry on CMRR

Not only does CMRR decrease with input signal frequency (largely due to an increase of A_C with frequency), but it is also severely affected by unbalance in the Thevenin resistance (or impedance) of the sources connected to the DA's input leads. Figure 2.2 illustrates the dc equivalent input circuit of a typical DA. Manufacturers typically specify a common-mode input resistance, R_{ic} , measured from one input lead to ground under CM excitation, and a difference-mode input resistance, R_{id} , measured under DM excitation from either input to ground.

By Ohm's Law, the DM current into the non-inverting input node is:

$$i_d = 2V_{id}/R_1 + V_{id}/R_{ic} \quad (2.18)$$

From which we can write

$$i_d/V_{id} = 1/R_{id} = 2/R_1 + 1/R_{ic} \quad (2.19)$$

Solving for the shunting resistance in equation 2.19,

$$R_1 = 2R_{id}R_{ic}/(R_{ic} - R_{id}) \quad (2.20)$$

In many differential amplifiers, $R_{ic} > R_{id}$. If $R_{ic} = R_{id}$, then from equation 2.20, $R_1 = \infty$.

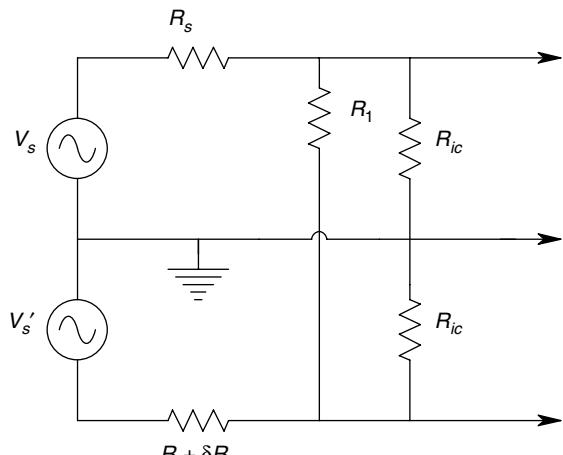


FIGURE 2.2

Differential amplifier input circuit at dc showing source resistance unbalance.

Let us assume that $R_{ic} = R_{id}$. Thus R_1 may be eliminated from Figure 2.2, which illustrates two Thevenin sources driving the DA through unequal source resistances, R_s and $R_s + \Delta R$. Using superposition and the definitions in equations 2.1 and 2.2, it is possible to show that a purely CM excitation, V_{sc} , produces an unwanted difference-mode component at the DA's input terminals.

$$V_{id}/V_{sc} = R_{ic} \Delta R / 2(R_{ic} + R_s)^2 \quad (2.21)$$

Also, V_{sc} produces a large CM component; the ΔR term is numerically negligible.

$$V_{ic}/V_{sc} = R_{ic}/(R_{ic} + R_s) \quad (2.22)$$

for purely DM excitation in V_s , we can also show that:

$$V_{id}/V_{sd} = R_{ic}/(R_{ic} + R_s) \quad (2.23)$$

and

$$V_{ic}/V_{sd} = R_{ic} \Delta R / 2(R_{ic} + R_s)^2 \quad (2.24)$$

In order to find the CMRR of the circuit of Figure 2.2, we will use equation 2.10 for V_o and the definition for CMRR. Thus,

$$V_o = A_D V_{sc} R_{ic} \Delta R / 2(R_{ic} + R_s)^2 + A_C V_{sc} R_{ic} / (R_{ic} + R_s) \quad (2.25)$$

and

$$V_o = A_D V_{sd} R_{ic} / (R_{ic} + R_s) + A_C V_{sd} R_{ic} \Delta R / 2(R_{ic} + R_s)^2 \quad (2.26)$$

After some algebra, we find that the circuit's CMRR, (i.e. CMRR_{sys}), is given by

$$\text{CMRR}_{\text{sys}} = [A_D + A_C \Delta R / 2(R_{ic} + R_s)] / [A_D \Delta R / 2(R_{ic} + R_s) + A_C] \quad (2.27)$$

Equation 2.27 may be reduced to the hyperbolic relation:

$$\text{CMRR}_{\text{sys}} = [A_D + A_C + \Delta R / 2(R_{ic} + R_s)] / [A_D / A_C] \Delta R / 2(R_{ic} + R_s) + 1 \quad (2.28)$$

which can be approximated by equation 2.29,

$$\text{CMRR}_{\text{sys}} \cong \text{CMRR}_A / [\text{CMRR}_A \Delta R / 2(R_{ic} + R_s) + 1] \quad (2.29)$$

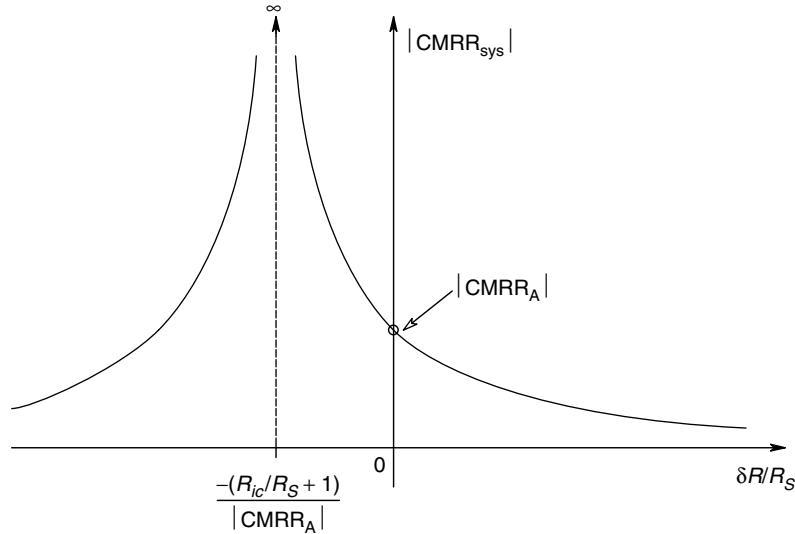
in which the $\text{CMRR} = \text{CMRR}_A = A_D / A_C$, and $\text{CMRR}_A \gg \Delta R / 2(R_{ic} + R_s)$ is specified by the manufacturer. A plot of CMRR_{sys} vs $\Delta R / R_s$ is shown in Figure 2.3. Note that when the Thevenin source resistances are matched,

$$\text{CMRR}_{\text{sys}} = \text{CMRR}_A$$

Also, when

$$\Delta R / R_s = -2(R_{ic} / R_s + 1) / \text{CMRR}_A \quad (2.30)$$

$$\text{CMRR}_{\text{sys}} \rightarrow \infty$$

**FIGURE 2.3**

Differential amplifier CMRR magnitude vs fractional unbalance in source resistance, $\Delta R_s / R_s$.

This implies that a judicious addition of an external resistance in series with one input lead or the other to introduce a ΔR may be used to increase the effective CMRR of the system. For example, if $R_{ic} = 100 \text{ M}\Omega$, $R_s = 10 \text{ k}\Omega$ and $\text{CMRR}_A = 100 \text{ dB}$, then $\Delta R/R = -0.2$, to give $\infty \text{ CMRR}_{\text{sys}}$. Since it is generally not possible to reduce R'_s , it is easier to externally add a similar δR to R_s .

Again, we stress that an amplifier's CMRR_A is a decreasing function of frequency because of the frequency dependence of the gains, A_D and A_C . Also, the ac equivalent input circuit of a DA contains capacitances in parallel with R_1 , R_{ic} and R'_{ic} , and the source impedances often contain a reactive, frequency dependent component. Thus in practice, in a given range of frequencies, CMRR_{sys} can often be maximized by the ΔR method, but can seldom be drastically increased because of reactive unbalances in the input circuit.

2.3 Operational Amplifiers

In this section, we review the systems properties of integrated circuit operational amplifiers of various types and discuss the many ways engineers use them for signal conditioning in the design of instrumentation systems.

A typical op-amp is a differential amplifier with a single output. It has a very high dc gain, K_{vo} , and high CMRR. Since op-amps are generally used with large amounts of negative feedback, the open-loop transfer function of most op-amps is designed to be of the form

$$A_D = V_o / (V_i - V'_i) = K_{vo} / [(\tau_1 s + 1)(\tau_2 s + 1)] \quad (2.31)$$

The second break frequency, $f_2 = 1/(2\pi\tau_2)$, is generally made to occur so that $|A_D(jf_2)| \ll 1$. This ensures closed-loop stability over a wide range of feedback conditions. The lower break frequency, $f_1 = 1/(2\pi\tau_1)$, occurs at a relatively low value.

However, a critical parameter governing the closed-loop, high frequency response of an op-amp circuit is the small signal gain \times bandwidth product (GBWP), which for most op-amps can be approximated by:

$$\text{GBWP} \cong K_{vo}/(2\pi\tau_1) \quad (2.32)$$

The symbol f_T is the unity gain or 0dB frequency of the open-loop op-amp. The f_T is approximately equal to the GBWP. The larger the GBWP, the higher the -3 dB cut-off frequency can be made in a broadband amplifier made with an op-amp.

Another parameter specified by the manufacturer, which determines the large signal dynamic behavior of an op-amp is its *slew rate*, η . The slew rate η , is defined as the largest possible attainable rate of change of the output voltage. Its units are generally given in volts/microsecond. The op-amp slew rates range from less than $1\text{ V}/\mu\text{s}$ in older models, such as the venerable LM741, to $3000\text{ V}/\mu\text{s}$ for the Comlinear CLC300A, which is intended to condition video frequency signals. As with the case of GBWP or f_T , the cost of op-amps increases with increasing slew rate. If an op-amp signal conditioner circuit is driven by a high frequency and amplitude sinusoidal source such that $|dV_o/dt|$ exceeds the amplifier's slew rate, then the output will appear to be a triangle wave with rounded (sinusoidal) peaks. When it occurs, slew rate limiting generates harmonic distortion in V_o . A prudent designer will always check to see that the maximum expected dV_o/dt in an op-amp circuit does not exceed η .

While the parameters K_{vo} , f_T , η , f_1 and f_2 govern an op-amp's dynamic behavior, manufacturers also give specifications on an op-amp's output resistance (R_o), CM and DM input resistances, dc bias current (I_B), dc offset voltage (V_{OS}), temperature coefficients (tempcos) for I_B and V_{OS} , short circuit voltage noise root power spectrum ($e_{na}(f)$), and equivalent input noise current root power spectrum ($i_{na}(f)$), to mention the most important ones.

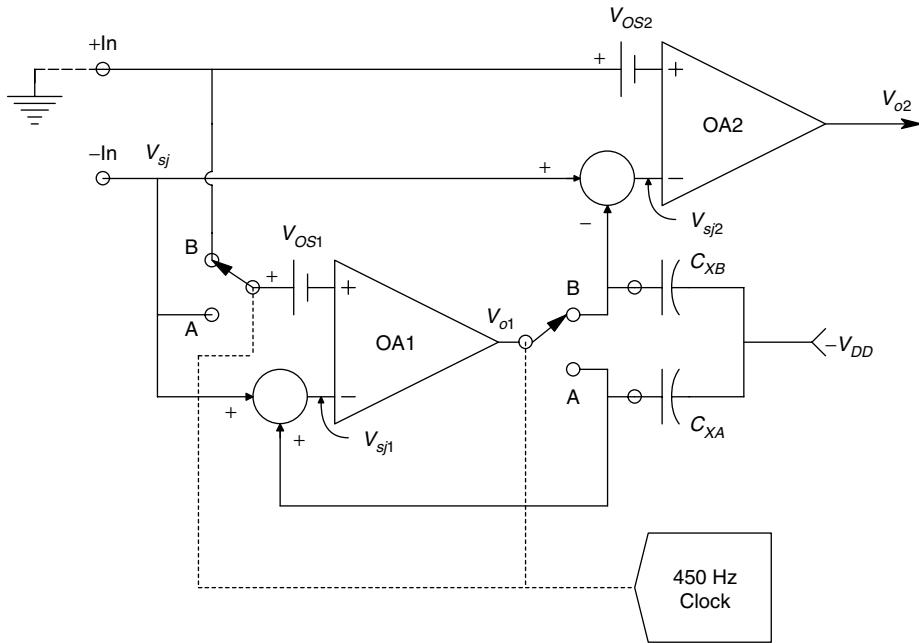
2.3.1 Types of Op-amps

Op-amps may be categorized by characteristics and applications. As will be explained in Chapter 3, low noise op-amps are best used in the headstages of analog signal conditioning systems, because their noise characteristics determine the overall noise performance of the system. When excellent dc stability and freedom from thermally caused dc drift are required, then designers often specify *chopper-stabilized op-amps*, provided that other criteria for dynamic response and noise are met.

Electrometer op-amps are used when input bias currents of 10^{-13} A or less and input resistances of greater than $10^{13}\Omega$ are required, with other conditions being met.

Fast op-amps are defined here as those having slew rates in excess of $25\text{ V}/\mu\text{s}$, and small signal GBWPs of 75 MHz or more. Most op-amps are designed to operate with ± 15 or $\pm 12\text{ V}$ dc supplies, and can source about $\pm 10\text{ mA}$ to a load. Modern op-amps have been designed to operate in portable instrumentation and communications equipment at supply voltages of $\pm 3.2\text{ V}$, $+5\text{ V}$ (single-ended), etc.

We define power and high voltage op-amps as those which can source in excess of 10 mA to a load, or which can operate with high voltage supplies of greater than $\pm 15\text{ V}$ (typically ± 40 to $\pm 150\text{ V}$). Some power op-amps can deliver as much as 30 A at 68 V (Apex PA03). In instrumentation systems, power op-amps can be used to drive strip-chart recorder pen motors, or other electromagnetic transducers. It is possible for an opamp to fall in two or more of the categories described above. For example, the Apex PA84

**FIGURE 2.4**

Simplified internal circuit of a Texas Instruments TLC2652 chopper-stabilized op-amp.

has a GBWP of 75 MHz, a slew rate of 200 V/ μ s, a peak output voltage of 143 V and a peak output current of 40 mA. Thus, it falls in the high voltage and power category, as well as the fast category.

Chopper-stabilized op-amps are used in applications where stable, low drift, dc amplification is required over a long period and under varying conditions of environmental temperature. Applications include weighing (electronic scales), electrochemical applications, and static photometric measurements.

CSOAs use a special, internal feedback architecture to continuously minimize the dc input offset voltage, V_{OS} , to compensate for the V_{OS} drift due to temperature changes or from power supply drift. Figure 2.4 illustrates the basic internal circuit architecture of a Texas Instruments TLC2652 series CSOA. Note that there are two op-amps and two SPDT MOS switches. The MOS switches are commutated synchronously by an internal 450 Hz clock, or in some versions, can be operated up to 1 kHz by an external clock. Not shown are the filters and circuits required to eliminate switching glitches and to shape amplifier bandwidths. The capacitors C_{XA} and C_{XB} are low loss dielectric, 0.1–1 μ F.

When the switches are in position A, the input to OA1 is shortcircuited, and it operates as a unity gain follower. Thus its output is its dc offset voltage, V_{OS1} , which is also the incremental voltage on capacitor C_{XA} . The incremental voltage on C_{XB} is also V_{OS1} , which is subtracted from V_{OS2} of OA2. When the switches are in position B, $V_{sj2} = (1 + K_1)V_{sj}$, and the two amplifiers are essentially in series— V_{OS1} is cancelled out and $V_{CXB} = K_1 V_{sj}$. The V_{OS2} is negligible because $K_1 \gg 1$.

Some key parameters for the TLC2652 series of CSOAs are:

$$V_{OS} = 0.5 \mu\text{V}$$

$$V_{OS} \text{ tempco} = 3 \text{ nV/}^\circ\text{C},$$

$$I_B = 4 \text{ pA}$$

$$\text{DC } A_{vo} = 150 \text{ dB}$$

$$\text{GBWP } f_T = 1.9 \text{ MHz}$$

$$\text{Open-loop break frequency } f_b \cong 0.06 \text{ Hz}$$

$$\eta = 3.1 \text{ V}/\mu\text{s}$$

A final type of op-amp which is widely encountered in preliminary circuit design practice, but unfortunately, is not commercially available, is the ideal op-amp. The ideal op-amp assumption is used in pencil-and-paper design calculations. It has the following properties: infinite differential gain, CMRR, input resistance, slew rate and GBWP, Zero noise, bias currents, offset voltage and output resistance. Probably the most important of these assumptions is that $A_D = \infty$. In order to realize a finite output voltage under this condition, $(V_i - V'_i) = 0$, or V_i must equal V'_i . In summary, an ideal op-amp is a differential VCVS, with infinite gain and frequency response.

The designer of an analog signal conditioning system for an instrumentation system faces a bewildering array of op-amp types and specifications. In the following sections, we try to give some rationale for choosing one op-amp over another. In making such choices, generally, the aim is to achieve a balance between meeting all minimum system specifications and dollar cost.

2.3.2 Basic Broadband Amplifier Design Using Op-amps

In many instances, the output signal of a sensor is too small to be effectively recorded, observed, or stored. The signal is assumed to be a voltage or current. Thus, the designer has the task of amplifying the weak signal, while introducing as little noise or distortion as possible. Often, impedance levels are a concern. For example, the transducer output may be represented by a Thevenin equivalent circuit in which the Thevenin (series) impedance may not be negligible compared to the input impedance of a signal conditioning amplifier. Thus, the design problem to be resolved includes the selection of an op-amp type and circuit which will minimize loading of the sensor output.

Another problem in designing high gain, broadband signal conditioners with op-amps is the gain-bandwidth trade-off. With simple op-amp amplifiers, the system's GBWP is approximately equal to the op-amp's GBWP. This implies that, in order to realize a signal conditioning gain of 1000 with a bandwidth of 100 kHz using low-noise op-amps with $f_T = 10 \text{ MHz}$, the strategy would be to cascade two amplifier circuits with their gains set to 31.62 (overall gain of 1000). The bandwidth (-3 dB point) of each stage is just $107/31.62 = 3.34 \text{ kHz}$. It can be shown that when N identical stages with transfer function:

$$A_v(s) = K_{vo}/(\tau s + 1) \quad (2.33)$$

are cascaded, then the -3 dB frequency of the cascade, f_c , is given by equation 2.34.

$$f_c = (1/2\pi\tau)(2^{1/N} - 1)^{1/2} \quad (2.34)$$

Thus, in the above example, $f_c = 203.5 \text{ kHz}$, and the frequency response design criterion is well met. If a single op-amp circuit were used, with a gain of 1000, then f_c would obviously be 10^4 Hz .

2.3.2.1 Non-Inverting Amplifier

Figure 2.5A illustrates the common, single ended, non-inverting amplifier configuration. Due to the op-amp's high K_{vo} , $V_i \cong V'_i$, the input impedance seen by the source circuit is R_{ic} , which is the op-amps CM input resistance (cf Figure 2.2).

In a modern, FET input op-amp, such as the Texas Instruments TL071, the R_{ic} can be of the order of $10^{12} \Omega$. Thus, in this case, unless the Thevenin source resistance approaches $10^8 \Omega$, loading effects and offset due to dc bias current (about 30 pA) will be negligible.

To simplify calculations, the R_{ic} at the V'_i node will be treated as infinite and the op-amp's output resistance considered to be negligible compared to R_F . The amplifier's closed-loop transfer function can be found by noting that

$$V_o = (V_s - V'_i) K_{vo} / (\tau s + 1) \quad (2.35)$$

and

$$V'_i = V_o R_1 / (R_1 + R_F) = \beta V_o \quad (2.36)$$

Substituting equation 2.36 into equation 2.35, and solving for V_o/V_s :

$$V_o/V_s = \frac{K_{vo}/(\tau s + 1)}{1 + \beta K_{vo}/(\tau s + 1)} \quad (2.37)$$

which reduces to

$$V_o/V_s = \frac{1/\beta}{s(\tau/\beta K_{vo}) + 1} \quad (2.38)$$

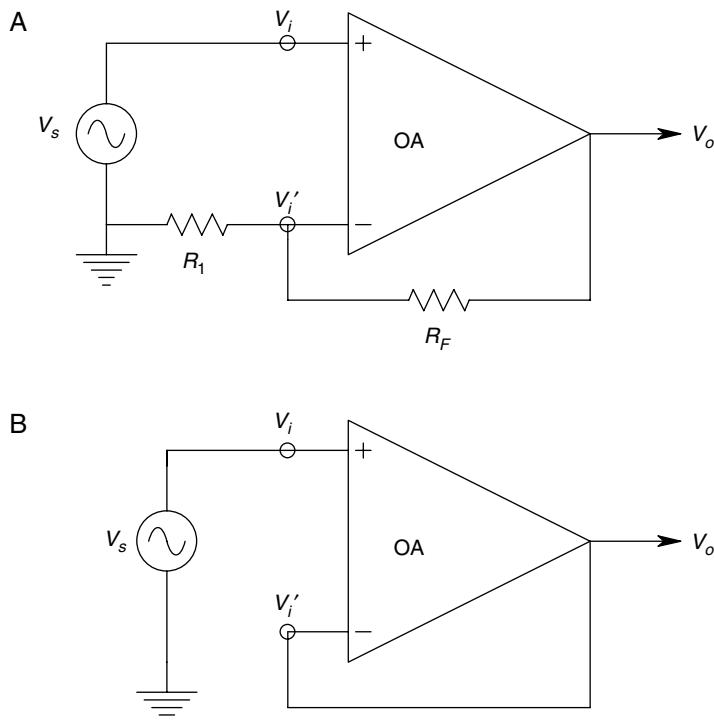


FIGURE 2.5

A. Non-inverting op-amp circuit, with gain set by R_1 and R_F . B. Unity gain buffer op-amp circuit.

if we assume that βK_{vo} is $\gg 1$, which is generally valid.

The GBWP for this amplifier is found from equation 2.37 to be:

$$\text{GBWP} = [K_{vo}/(1 + \beta K_{vo})][(1 + \beta K_{vo})/2\pi\tau] = K_{vo}/2\pi\tau \text{ Hz} \quad (2.39)$$

which is the same as for the op-amp by itself.

If R_1 is made infinite and R_F is made a shortcircuit, the non-inverting follower circuit of Figure 2.5B becomes a unity gain voltage follower. Unity gain followers are often used as output buffers to achieve impedance isolation.

2.3.2.2 The Inverting Amplifier and Summer

Figure 2.6 illustrates the common inverting amplifier configuration. Here, we have assumed two inputs, V_{s1} and V_{s2} . The amplifier output can be found by superposition. The output voltage can be written as:

$$V_o = (0 - V'_i) K_{vo}/(\tau s + 1) \quad (2.40)$$

V'_i is found by solving the node equation,

$$(V'_i - V_{s1}) G_1 + (V'_i - V_{s2}) + (V'_i - V_o) G_F = 0 \quad (2.41)$$

When V'_i is substituted into equation 2.40 and solving for V_o/V_{s1} , letting $V_{s2}=0$:

$$V_o = \frac{-V_{s1}(G_1 K_{vo} / \sum G) / (1 + K_{vo} G_F / \sum G)}{s[\tau/(1 + K_{vo} G_F / \sum G)] + 1} \quad (2.42)$$

where $\sum G \equiv G_1 + G_2 + G_F$. When it is assumed that $K_{vo} G_F / \sum G \gg 1$, then:

$$V_o/V_{s1} = \frac{-(G_1/G_F)}{s(\tau \sum G/K_{vo} G_F) + 1} \quad (2.43)$$

In general, the gain for the k th inverting input is, by superposition,

$$V_o/V_{sk} = \frac{-(G_k/G_F)}{s(\tau \sum G/K_{vo} G_F) + 1} \quad (2.44)$$

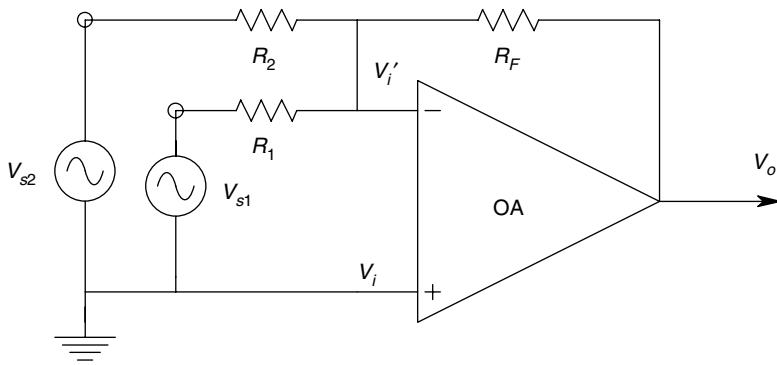


FIGURE 2.6

Summing inverting op-amp configuration.

The GBWP for the k th inverting input is easily seen to be:

$$\text{GBWP}_k = (K_{vo}/2\pi\tau)(G_k/\sum G) \text{ Hz} \quad (2.45)$$

Note that in the inverting amplifier case, the amplifier GBWP, given by equation 2.45, is not equal to the op-amp's GBWP, except for very large G_k .

2.3.3 Current Feedback Op-amps

Current feedback op-amps (CFOAs) constitute a relatively new class of op-amp that offers certain design advantages, in terms of independent control of gain and bandwidth, and it generally has very high slew rates. The internal systems architecture of a CFOA connected as a simple non-inverting amplifier is shown in Figure 2.7. With a nearly ideal unity gain, VCVS replaces the high impedance input circuit of a conventional op-amp. Thus, the non-inverting (V_i) terminal has a very high input impedance, but the V'_i node is the output of the VCVS and therefore, presents a very low input (Thevenin) impedance, in the order of 50Ω . The output current of the VCVS, I_n , is the input for a CCVS (Ω_s), which determines the CFOA's output voltage, V_o . The CCVS is designed so that:

$$\Omega_s = \Omega_o/(\tau s + 1) \Omega \quad (2.46)$$

Hence

$$V_o = I_n[\Omega_o/(\tau s + 1)] V \quad (2.47)$$

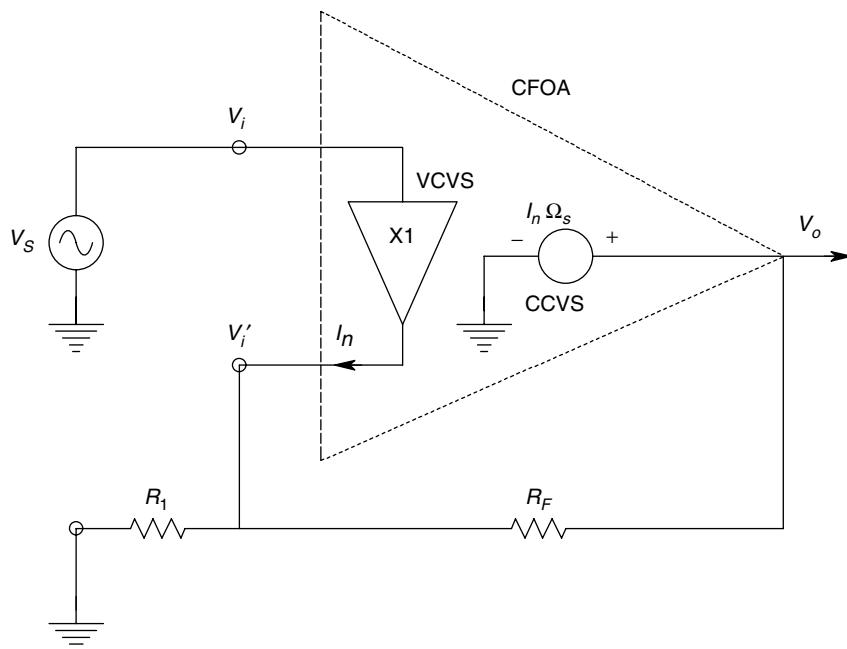


FIGURE 2.7

A non-inverting, current feedback, op-amp amplifier circuit.

The control current, I_n , may be found from Kirchoff's current law applied to the V'_i node (here, the CCVS output resistance is ignored).

$$I_n = V_s G_1 + (V_s - V_o) G_F \quad (2.48)$$

By substituting equation 2.48 for I_n into equation 2.47, the transfer function for the non-inverting CFOA amplifier can be found:

$$\frac{V_o}{V_s} = \frac{(G_1 + G_F)\Omega_o / (1 + G_F\Omega_o)}{s[\tau / (1 + G_F\Omega_o)] + 1} \quad (2.49)$$

if we assume that $G_F\Omega_o \gg 1$, then the transfer function reduces to:

$$\frac{V_o}{V_s} = \frac{1 + R_F/R_1}{s(\tau R_F/\Omega_o) + 1} \quad (2.50)$$

The circuit's dc gain, A_{VO} , is:

$$A_{VO} = (R_1 + R_F)/R_1 \quad (2.51)$$

The circuit's GBWP is:

$$\text{GBWP} = \frac{\Omega_o(R_F + R_1)}{2\pi\tau R_F R_1} = \Omega_o A_{VO} / 2\pi\tau R_F \text{ Hz} \quad (2.52)$$

The -3 dB frequency of the circuit is given by equation 2.53:

$$f_b = (\Omega_o + R_F)/2\pi\tau R_F \approx \Omega_o/2\pi\tau R_F \text{ Hz} \quad (2.53)$$

Note that the break frequency, f_b , is dependent on R_F alone. Thus the dc gain can be set with R_1 , holding the break frequency constant with R_F . Note that with a conventional op-amp non-inverting amplifier, the break frequency varies inversely with the circuit's dc gain.

Although CFOAs offer advantages in terms of high output slew rate and f_T , and additionally, the absence of the customary trade-off between gain and bandwidth found with conventional op-amps, they are not as flexible in terms of stability and the ability to synthesize transfer functions.

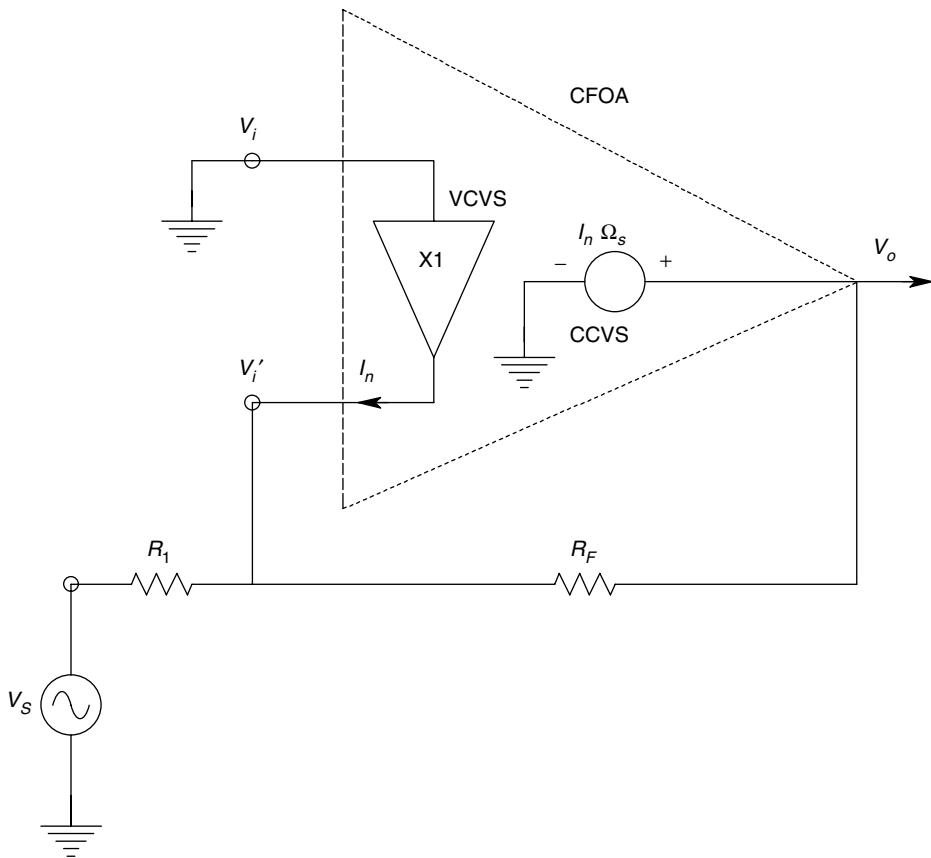
First, let us examine the transfer function of a CFOA used in the CFOA inverting configuration, shown in Figure 2.8. Here the positive voltage input (V_i) node is tied to ground, causing the output of the unity gain buffer to be zero. Neglecting R_o , $V'_i = 0$, I_n is given simply by:

$$I_n = -(V_s G_1 + V_o G_F) \quad (2.54)$$

If this expression for I_n is substituted into equation 2.47 for V_o , the transfer function of the inverting amplifier can be found. Assuming that $\Omega_o G_F \gg 1$:

$$\frac{V_o}{V_s} = \frac{-R_F/R_1}{s(\tau R_F/\Omega_o) + 1} \quad (2.55)$$

Note that for the inverter, the dc gain is simply $-R_F/R_1$, the -3 dB break frequency, f_b , is $\Omega_o/(2\pi\tau R_F)$ Hz, and the GBWP is $\Omega_o/(2\pi\tau R_1)$ Hz. In using a conventional op-amp, it is a

**FIGURE 2.8**

An inverting configuration of a current feedback op-amp.

simple extension of the gain relation to replace R_1 and/or R_F with capacitors, or series or parallel R - C circuits and the stability is generally preserved. In the case of CFOAs, however, the designer must avoid applying direct capacitive feedback between the V_o and V'_i nodes (some manufacturers provide an internal R_F to avoid this problem).

To examine what happens when a capacitor is connected in place of R_F in the inverting CFOA amplifier in Figure 2.8, we note that the current I_n can be written as:

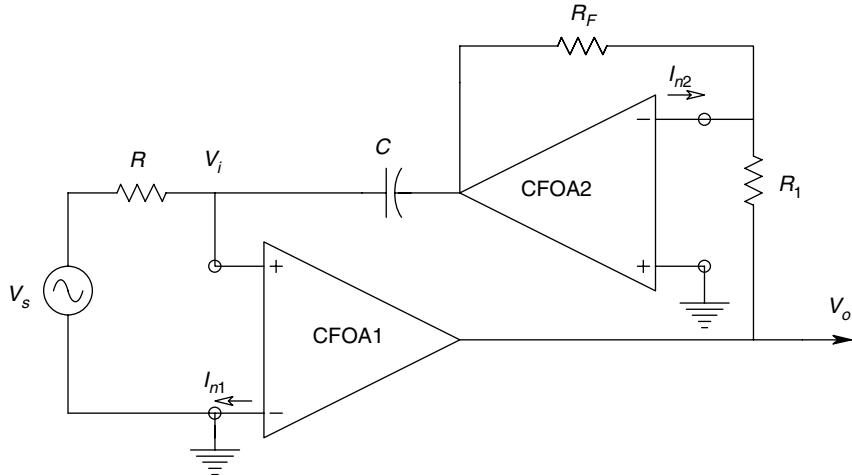
$$I_n = -(V_s G_1 + V_o s C) \quad (2.56)$$

When equation 2.56 for I_n is substituted into equation 2.47 for V_o , we can derive the transfer function, given in equation 2.57:

$$\frac{V_o}{V_s} = \frac{-\Omega_o G_1}{s(\tau + C\Omega_o) + 1} \quad (2.57)$$

If we set $R_1 = 1 \text{ M}\Omega$, $C = 1 \mu\text{F}$ and $\Omega_o = 10^8 \Omega$, the transfer function in equation 2.57 becomes:

$$\frac{V_o}{V_s} = -100/(s100 + 1) \quad (2.58)$$

**FIGURE 2.9**

A non-inverting integrator using two, current feedback op-amps.

This is certainly not an effective integrator; it is best characterized as an inverting low-pass filter with a break frequency of 10^{-2} rad/s.

An integrator which can be made from CFOAs is shown in Figure 2.9. This is a non-inverting integrator. If $\tau s \rightarrow 0$, in the CFOAs, the integrator gain may be shown to be approximately:

$$V_o/V_s \cong \frac{\Omega_{o1}/R_{o1}}{sRC(R_F\Omega_{o1}/R_{o1}R_1) + 1} \quad (2.59)$$

Note that a Thevenin output resistance, R_{o1} , must be assumed for the CFOA1 unity gain buffer to permit analysis in this case. At frequencies above the break frequency, f_b , the integrator gain becomes:

$$V_o/V_s = R_1/sRCR_F \quad (2.60)$$

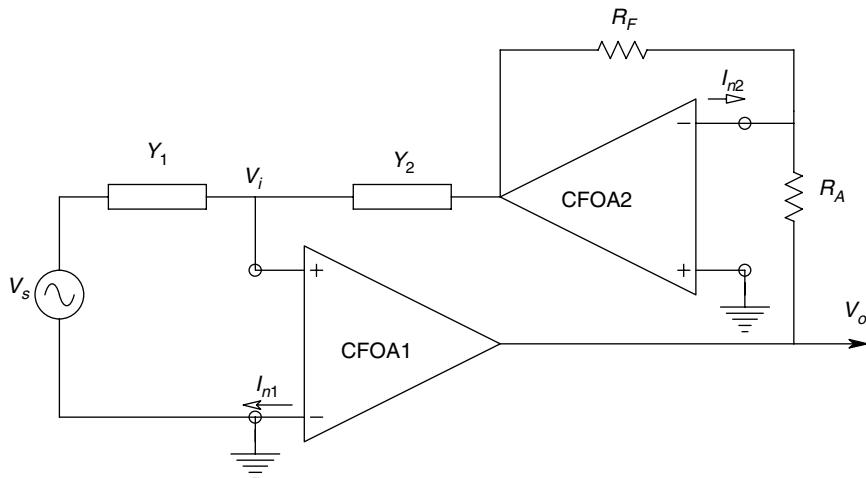
The break frequency for this integrator is quite low, about 5×10^{-7} r/s for $\Omega_o = 10^8 \Omega$, $C = 1 \mu F$, $R = 1 M\Omega$, $R_F = R_1 = 10^3 \Omega$ and $R_{o1} = 50 \Omega$.

As a means of synthesizing simple active filters using CFOAs, consider the general circuit shown in Figure 2.10. Here, we have replaced the R and C of the integrator in Figure 2.9 with admittances Y_1 and Y_2 , respectively. Using the same assumptions as those made above in the analysis of the integrator:

$$V_o/V_s = \frac{Y_1\Omega_1G_{o1}}{Y_1 + Y_2[(1 + \Omega_1\Omega_2G_{o1}G_A)/G_F\Omega_2]} \quad (2.61)$$

If $\Omega_1 = \Omega_2$, $R_F > R_A$ and assuming $\Omega_1G_{o1} \gg 1$, the gain expression in equation 2.61 reduces to:

$$V_o/V_s \cong Z_2R_A/Z_1R_F \quad (2.62)$$

**FIGURE 2.10**

A general, non-inverting, current feedback op-amp feedback circuit.

A more detailed analysis would include the real poles of the CFOAs in equation 2.61. Note that by using this circuit architecture, it is possible to synthesize a simple, two-pole, one-zero bandpass filter if

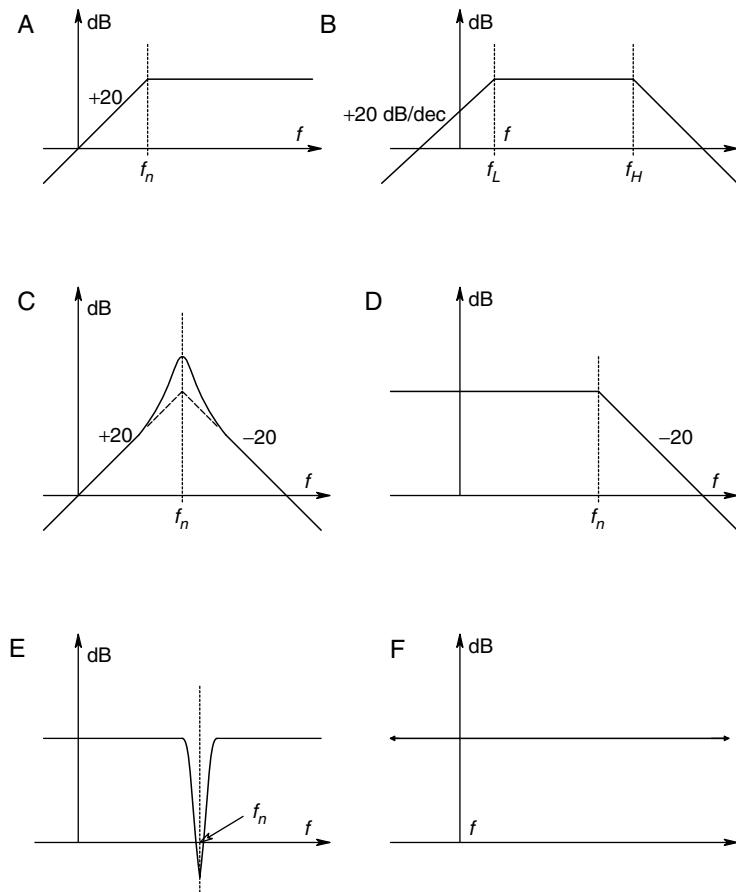
$$Z_1 = R_1 + 1/j\omega C_1 \quad \text{and} \quad Y_2 = G_2 + j\omega C_2$$

In summary, it can be seen that in simple inverting and non-inverting configurations, CFOAs offer high frequency response and slew rates. However, care must be exercised in using them in circuits having reactive elements. The basic inverting gain relationship of conventional op-amps, $V_o/V_s = -Z_F/Z_{IN}$, does not apply to CFOAs. Special circuit architecture must be used in many cases to realize even simple transfer functions, such as integration or bandpass filtering.

2.4 Analog Active Filter Applications Using Conventional Op-amps

2.4.1 Introduction

In the previous sections of this chapter, the use of op-amps to make single-ended, broadband amplifiers was discussed. There are many instances in instrumentation systems where it is also necessary to filter a time-varying signal by passing it through a linear, frequency dependent gain device. Such filtering is often used to improve the SNR by attenuating those frequencies containing noise, but which do not contain significant signal information. Coherent interference may also be reduced or effectively eliminated by filtering. Finally, before analog to digital conversion (periodic sampling), analog signals must be low-pass filtered to prevent errors in subsequent digital processing caused by aliasing (Sampling and aliasing are discussed in detail in

**FIGURE 2.11**

Bode asymptotes of common types of filters: A) high-pass, B) broad bandpass, C) narrow (tuned) bandpass, D) low-pass, E) band reject or notch, F) all-pass.

Chapter 9). Low-pass filtering is also used to smooth or average the outputs of nonlinear circuits, such as rectifiers, phase sensitive rectifiers, true RMS converters, etc.

Analog active filters fall into several broad categories, depending on the shape of the filter pass band. Figure 2.11 illustrates the major categories we will consider in the following sections. These include high-pass, broad bandpass, narrow bandpass (high Q), low-pass, notch (band-reject) and all-pass (used for phase compensation, the transfer function magnitude remains constant).

Our treatment of analog active filters given here is by no means complete—it is provided as a review of AF circuits which find common use in analog signal conditioning in instrumentation systems. We will not go into design details of high order AFs in this text (pole and zero placement). The reader should appreciate that the analysis and design of analog active filters is an extensive subject, and a large amount of literature is available on the subject. Many references on analog active filter design are included in the chapter bibliography.

2.4.2 Analog Active Filter Architectures

The reader should be aware that there are many active filter circuit architectures which can be used to realize a given transfer function. Some circuits offer the advantage of

few components or minimum sensitivity to op-amp gain changes, while others offer design flexibility in terms of being able to independently set one or more of the filter parameters (mid-band gain, break frequency, damping factor or Q) by adjusting a single unique R or C for each parameter. In this section, we stress on certain active filter architectures which allow easy design. To expedite analysis, we assume that all the op-amps are of conventional design and are ideal.

2.4.2.1 Controlled Source Active Filters

This type of filter can be used to realize bandpass, low-pass and high-pass quadratic transfer functions. A single op-amp is used as a low or unity gain VCVS with four or five resistive or capacitive feedback elements. Figure 2.12 illustrates the well-known *Sallen and Key low-pass filter*. To derive this filter's transfer function, we observe that the op-amp behaves as a VCVS with a gain given by

$$K_v = 1 + R_F/R_A$$

Thus

$$V_2 = \frac{V_1(1/sC_2)}{R_2 + 1/sC_2} = \frac{V_1}{1 + sR_2C_2} \quad (2.63)$$

Writing a node equation on the V_1 node:

$$V_1 \left(G_1 + \frac{1}{R_2 + 1/sC_2} + sC_1 \right) - sC_1 V_o = V_s G_1 \quad (2.64)$$

Equation 2.63 can be solved for V_1 and substituted into equation 2.64, such that V_o can be solved in terms of V_s . The transfer function of the Sallen and Key low-pass filter can thus be written as:

$$\frac{V_o}{V_s} = \frac{K_v}{s^2 C_1 C_2 R_1 R_2 + s[C_2(R_1 + R_2) + R_1 C_1(1 - K_v)] + 1} \quad (2.65)$$

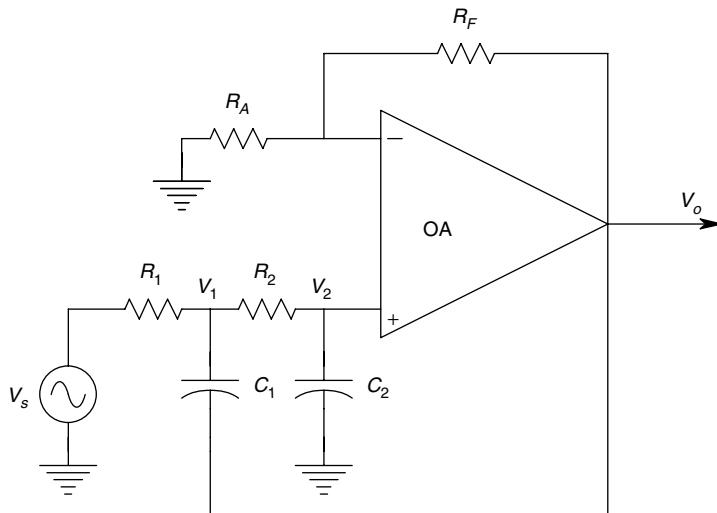


FIGURE 2.12

Sallen and Key quadratic low-pass filter.

If K_v is made unity, then the transfer function becomes

$$V_o/V_s = \frac{1}{s^2 C_1 C_2 R_1 R_2 + s C_2 (R_1 + R_2) + 1} \quad (2.66)$$

The denominator of the transfer function has the ‘standard’ quadratic form:

$$s^2/\omega_n^2 + s_2\zeta/\omega_n + 1$$

The undamped natural frequency, ω_n , for the Sallen and Key low-pass filter is:

$$\omega_n = \frac{1}{C_1 C_2 R_1 R_2} \text{ r/s} \quad (2.67)$$

The damping factor, ζ , can be written as:

$$\zeta = \frac{C_2(R_1 + R_2)}{2C_1 C_2 R_1 R_2} \quad (2.68)$$

If R_1 is made equal to R_2 , then the damping factor reduces to the simple expression:

$$\zeta = \sqrt{\frac{C_2}{C_1}} \quad (2.69)$$

Conveniently, damping factor of the low-pass filter is set by the square root of the ratio of the capacitor values. Simultaneous adjustment of $R_1=R_2$ with a ganged variable resistor can set ω_n independently from the damping factor.

Figure 2.13 illustrates a Sallen and Key high-pass filter. Analysis of this circuit is similar to the low-pass filter above. If K_v is set to unity, and $C_1=C_2=C$, then it is easy to show that the transfer function is:

$$V_o/V_s = \frac{s^2 C_2 R_1 R_2}{s^2 C_1 C_2 R_1 R_2 + s R_1 2C + 1} \quad (2.70)$$

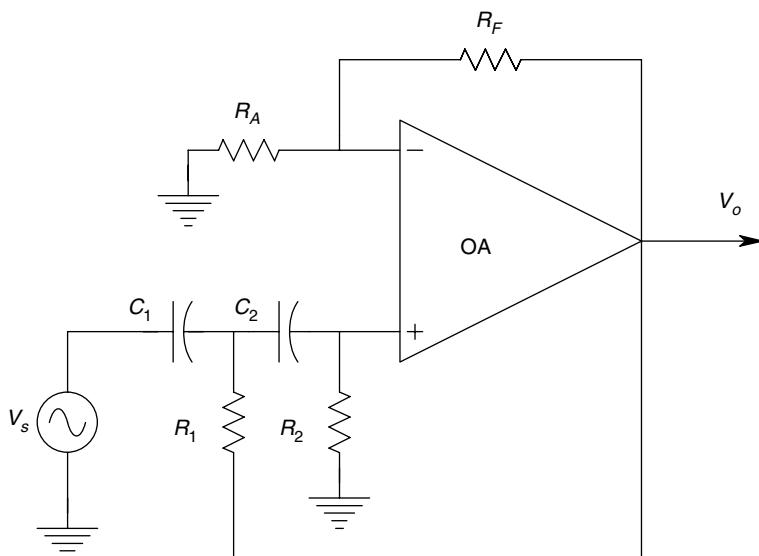
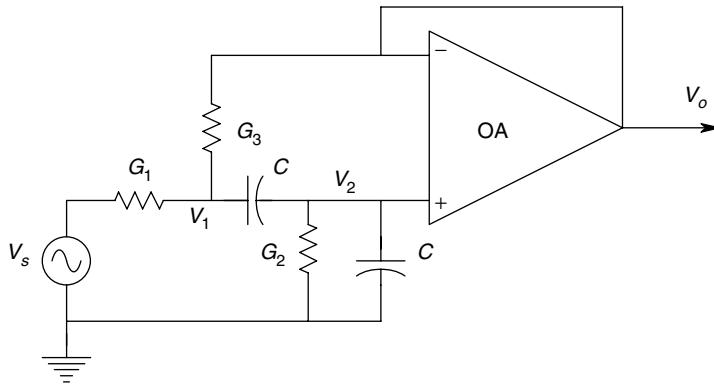


FIGURE 2.13
Sallen and Key quadratic high-pass filter.

**FIGURE 2.14**

A controlled source quadratic bandpass filter.

The undamped natural frequency, ω_n , of the high-pass filter too is given by equation 2.67. The damping factor is given by:

$$\zeta = \frac{R_1}{2} \sqrt{\frac{1}{C^2 R_1 R_2}} = \sqrt{\frac{R_1}{R_2}} \quad (2.71)$$

It is also possible to make a *controlled source bandpass filter*, as shown in Figure 2.14. Although a general analysis is interesting, it is more useful to realize this circuit with $K_v=1$ and $C_1=C_2=C$. The transfer of the bandpass filter is:

$$V_o/V_s = \frac{s[CG_1/(G_2G_1 + G_2G_3)]}{s^2[C^2/(G_2G_1 + G_2G_3)] + s[C(G_2 + 2G_1 + G_3)/(G_2G_1 + G_2G_3)] + 1} \quad (2.72)$$

Here

$$\omega_n = \frac{\sqrt{G_2(G_1 + G_3)}}{C} \text{ r/s}$$

The gain at resonance is:

$$A_{VPK} = \frac{G_1}{G_2 + 2G_1 + G_3} \quad (2.73)$$

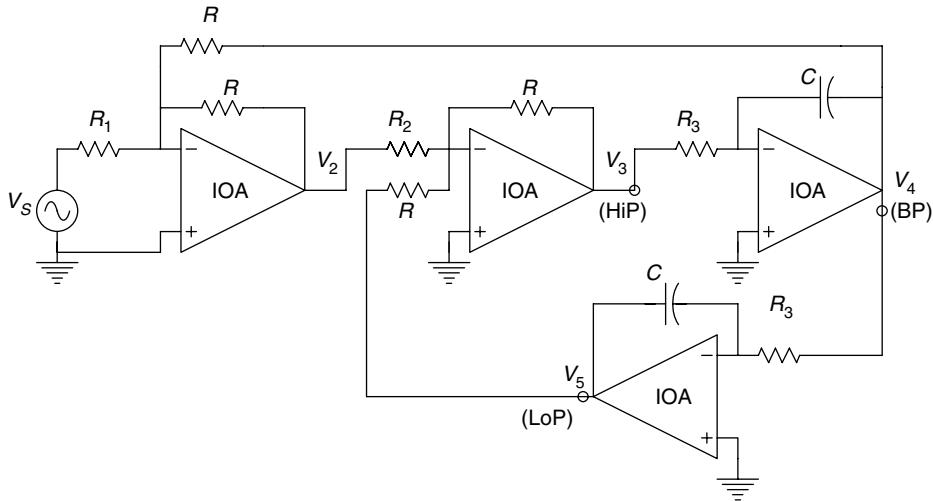
The Q of the tuned circuit is:

$$Q = \frac{\sqrt{G_2(G_1 + G_3)}}{G_2 + 2G_1 + G_3} \quad (2.74)$$

Note that the peak response, A_v , and Q are independent of C ; $C_1=C_2=C$ can be used to set ω_n at constant Q and A_v .

2.4.2.2 Biquad Active Filters

Another active filter architecture with which it is extremely easy to meet design criteria is the *biquad* AF. There are two versions of the biquad, but Figure 2.15 shows the most

**FIGURE 2.15**

The two-loop, biquad active filter.

versatile—the two-loop biquad. This circuit can act as a high-pass, sharply tuned bandpass or low-pass filter, depending on which output is selected. It also allows *independent adjustment* of the peak gain, break frequency and damping factor, or Q , of the filter. All-pass and band-reject (notch) filters can also be realized using the basic biquad filter, with the addition of another op-amp.

First, the low-pass filter transfer function of the biquad AF in Figure 2.15 is examined. Note that the subunits of this filter are two inverting summers and two integrators. The transfer function of the low-pass filter can be found by the application of Mason's Rule:

$$V_5/V_s = \frac{(-R/R_1)(-R/R_2)(-1/sCR_3)(-1/sCR_3)}{1 + [(-1)(-R/R_2)(-1/CR_3) + (-1)(-1/sCR_3)(-1/sCR_3)]} \quad (2.75)$$

This transfer can be simplified to standard quadratic (time-constant) form:

$$V_o/V_s = \frac{R^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1} \quad (2.76)$$

Inspection of equation 2.76 shows that the low frequency gain of this low-pass filter is:

$$A_{VLOW} = R^2/(R_1R_2) \quad (2.77)$$

The natural frequency is

$$\omega_n = 1/CR_3 \text{ r/s} \quad (2.78)$$

and the damping factor is

$$\zeta = R/2R_2 \quad (2.79)$$

The same biquad filter can be used as a *tuned bandpass filter*. In this case, the output is taken from the V_4 node. The transfer function can again be found by the application of Mason's Rule:

$$V_4/V_s = \frac{-sR_3CR^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1} \quad (2.80)$$

The bandpass filter has a peak response at the natural frequency given by equation 2.78. The peak gain is

$$A_{VPK} = -R/R_1 \quad (2.81)$$

The Q of the bandpass filter, a measure of the sharpness of its tuning, is given by:

$$Q = R_2/R \quad (2.82)$$

Finally, the biquad AF of Figure 2.15 can produce a *quadratic high-pass characteristic*. Here the output is taken from the V_3 node. The transfer function can be shown to be:

$$V_3/V_s = \frac{s^2R_3^2C^2R^2/(R_1R_2)}{s^2R_3^2C^2 + (sR_3CR/R_2) + 1} \quad (2.83)$$

The natural frequency and damping factor of the high-pass filter are the same as for the low-pass filter. Its high frequency gain is the same as the low frequency gain of the low-pass filter.

It is also possible, through the use of a fourth op-amp inverting summer, to generate a *notch filter*, which will sharply attenuate a selected frequency and pass all other frequencies. Such notch filters often find application in eliminating unwanted coherent interference, such as power line frequency hum. The basic operation performed by a notch filter is illustrated in Figure 2.16. Simply stated, the output of a bandpass filter is subtracted from its input. The bandpass filter's peak gain is made unity. As a result of this operation, the transfer function of the notch filter has a pair of conjugate zeros on the $j\omega$ axis in the s -plane at $\pm j\omega_n$ and the usual complex conjugate, quadratic pole pair at $s = \pm j\omega_n$. Its transfer function is given by equation 2.84.

$$V_6/V_s = \frac{(s^2/\omega_n^2) + 1}{(s^2/\omega_n^2) + s(1/Q\omega_n) + 1} \quad (2.84)$$

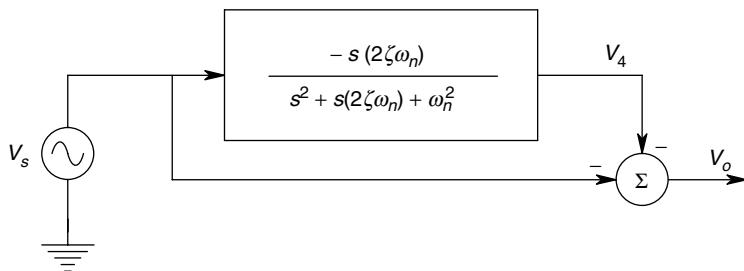
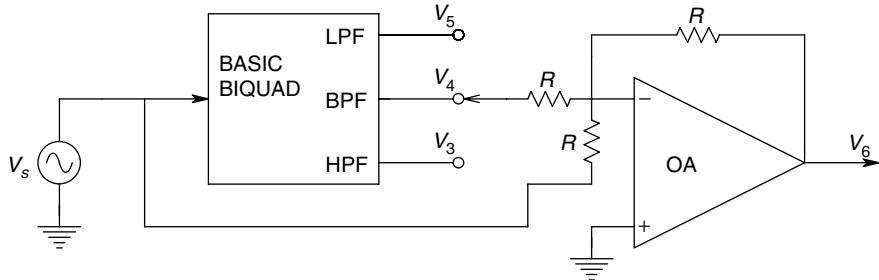
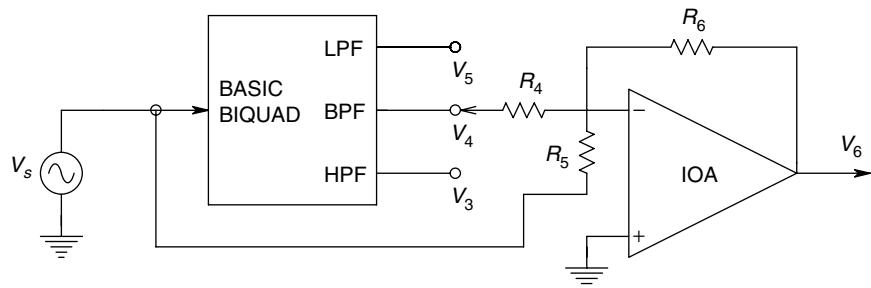


FIGURE 2.16

Block diagram for biquad notch filter. V_4 is the inverting bandpass output of the biquad filter of Figure 2.14.

**FIGURE 2.17**

Realization of biquad notch filter using the basic biquad of Figure 2.15.

**FIGURE 2.18**

Realization of biquad all-pass filter. See text for special relations for R_4 , R_5 and R_6 .

A biquad realization of a notch filter is shown in Figure 2.17. It is left as an exercise to the reader to derive its transfer function (note that R_1 should equal R to obtain the form of equation 2.84).

As a final example of the versatility of the biquad active filter architecture, we examine the *all-pass filter* design shown in Figure 2.18. All-pass filters have a flat frequency magnitude response; only the phase varies with frequency. All-pass filters are used to modify the phase response of electronic signal conditioning systems without introducing frequency dependent gain changes. Here, the bandpass output, V_4 , of the biquad of Figure 2.15 is summed with the filter input, V_s . The all-pass filter output, V_6 , is easily seen to be:

$$V_6 = -R_6[(V_s/R_5) + (V_4/R_4)] \quad (2.85)$$

When the relation for V_4/V_s , equation 2.80, is substituted into equation 2.85, it can be shown we can show that:

$$V_6/V_s = -\frac{(R_5/R_6)[s^2 R_3^2 C^2 + (s R_3 C R / R_2) + 1] - [(R_6/R_4)s R_3 C R^2 / R_1 R_2]}{s^2 R_3^2 C^2 + (s R_3 C R / R_2) + 1} \quad (2.86)$$

This relation may be reduced to the standard all-pass filter format given by equation 2.87 if we let $R_1=R$, $R_5=2R_4$ and $R_6=R_5$:

$$V_6/V_s = -\frac{s^2 R_3^2 C^2 - (s R_3 C R / R_2) + 1}{s^2 R_3^2 C^2 + (s R_3 C R / R_2) + 1} \quad (2.87)$$

In summary, note that the frequency response magnitude function for the all-pass filter is unity over all frequencies; only the phase varies. When $s=j\omega$ in equation 2.87, the phase of the frequency response function is given by:

$$\phi = -180 - 2 \tan^{-1} \left(\frac{1 - \omega^2 R_3^2 C^2}{\omega C R_3 R / R_2} \right) \quad (2.88)$$

2.4.2.3 Generalized Impedance Converter Active Filters

The Generalized Impedance Converter (GIC) circuits, using two op-amps and five two-terminal elements (either R_s or C_s), can be used to synthesize impedances to ground of the form, $Z(s) = sX$, or $Z(s) = 1/s^2D$. These impedances in turn, can be combined with other R_s , C_s and op-amps to generate various quadratic transfer functions, such as that discussed for the biquad AF architecture above. Figure 2.19 illustrates the basic GIC architecture. Analysis of the driving point impedance of the GIC is made easier if it is assumed that the op-amps are ideal. Under this condition, we can write the input impedance of the GIC to ground as:

$$Z_{11} = V_1/I_1 = V_1/[(V_1 - V_2)/Z_1] \quad (2.89)$$

Also, from the ideal op-amp assumption and Ohm's Law we have:

$$I_2 = (V_2 - V_1)/Z_2 = I_3 = (V_1 - V_3)/Z_3 \quad (2.90)$$

$$I_4 = (V_3 - V_1)/Z_4 = I_5 = V_1/Z_5 \quad (2.91)$$

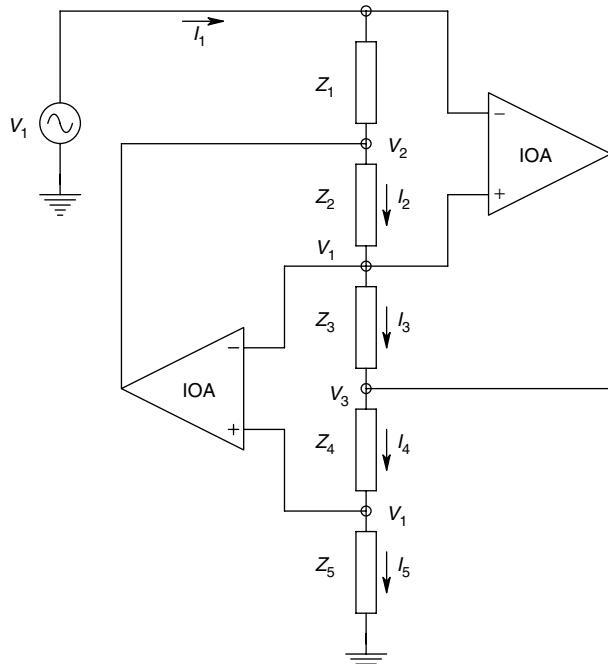


FIGURE 2.19

Basic Generalized Impedance Converter (GIC) architecture.

It can be shown from equations 2.89, 2.90 and 2.91 that:

$$V_3 = V_1[1 + (Z_4/Z_5)] \quad (2.92)$$

and

$$V_2 = V_1 \left(\frac{Z_3 Z_5 - Z_2 Z_4}{Z_3 Z_5} \right) \quad (2.93)$$

The GIC driving point impedance can be shown to be:

$$Z_{11}(s) = V_1/I_1 = Z_1 Z_3 Z_5 / Z_2 Z_4 \quad (2.94)$$

If we let $Z_2 = 1/sC_2$ (a capacitor), and the other Z s be resistors, then Z_{11} is in the form of a high- Q inductive reactance:

$$Z_{11} = s(C_2 R_1 R_3 R_5 / R_4) \quad (2.95)$$

where the equivalent inductance is given by

$$L_{eq} = [C_2 R_1 R_3 R_5 / R_4] \text{ H} \quad (2.96)$$

If both Z_1 and Z_5 are made capacitors, then the input impedance of the GIC becomes a frequency dependent negative resistance, or FDNR element:

$$Z_{11} = -R_3/s^2 C_1 C_5 R_2 R_4 = -1/s^2 D \quad (2.97)$$

Thus the D element of the GIC is given by:

$$D = C_1 C_5 R_2 R_4 / R_3 \quad (2.98)$$

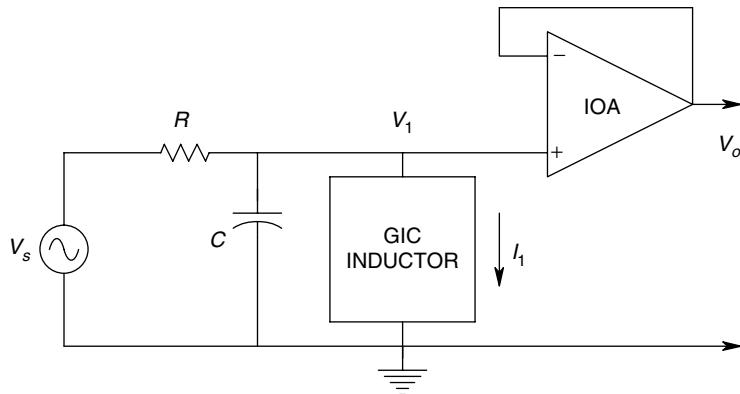
A number of active filter designs based on the inductive and 'D' Z_{11} forms have been devised. Bandpass, all-pass, notch, low-pass and high-pass designs are possible. Examples of these designs may be found in Chapter 4 of the text by Franco. We illustrate a two-pole bandpass filter in Figure 2.20 and a FDNR based two-pole low-pass filter in Figure 2.21.

The transfer function for the bandpass filter can be shown to be:

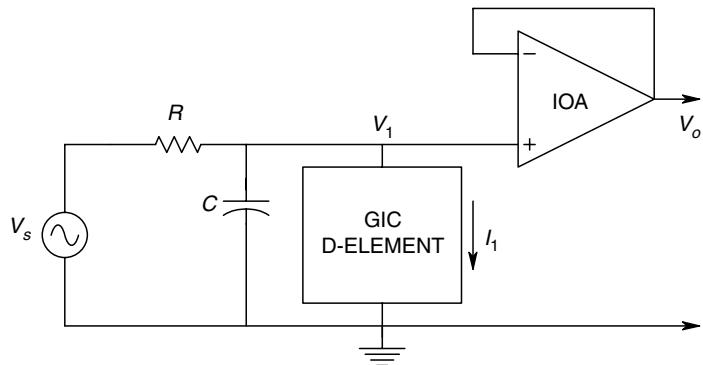
$$V_o/V_s = \frac{s L_{eq} / R}{s^2 C L_{eq} + (s L_{eq} / R) + 1} \quad (2.99)$$

where L_{eq} is given by equation 2.96. The transfer function of the low-pass filter is given by equation 2.100.

$$V_o/V_s = \frac{1}{s^2 R D + s R C + 1} \quad (2.100)$$

**FIGURE 2.20**

Tuned, quadratic bandpass filter using a GIC as a low loss inductor.

**FIGURE 2.21**

Quadratic low-pass filter realized with a GIC FDNR element.

where D is given by equation 2.98 above. Figure 2.22 shows a GIC derived all-pass filter and Figure 2.23 a GIC derived notch filter. Derivation of their transfer functions are left as exercises at the end of this chapter for the reader to attempt.

2.4.2.4 High-Order Active Filters

Often, the simple two-pole (quadratic) active filter designs which we have described above do not meet the signal conditioning requirements of an instrumentation system. In these cases, quadratic filters may be cascaded to create high order filters having four or more poles. Such high order filters are described according to how they perform in the time and frequency domains. We list below the more commonly used types of high-order active filters:

1. Butterworth Filters: The frequency response magnitude for this type of high order AF has no ripple in the passband. Its phase vs frequency response is nonlinear, so that the signal propagation time through the filter is not constant over frequency. There is considerable output overshoot in the time domain response of Butterworth filters, of order greater than $n=4$ to transient inputs. The slope of the stopband attenuation is $-20n$ dB/decade.
2. Chebychev Filters: In the frequency domain, the Chebychev filter has no ripple in its stopband, but has a specified passband ripple. Generally, a Chebychev

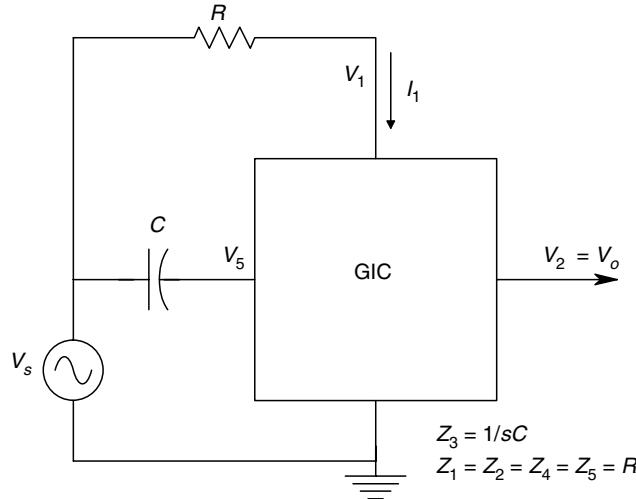


FIGURE 2.22
GIC derived all-pass filter.

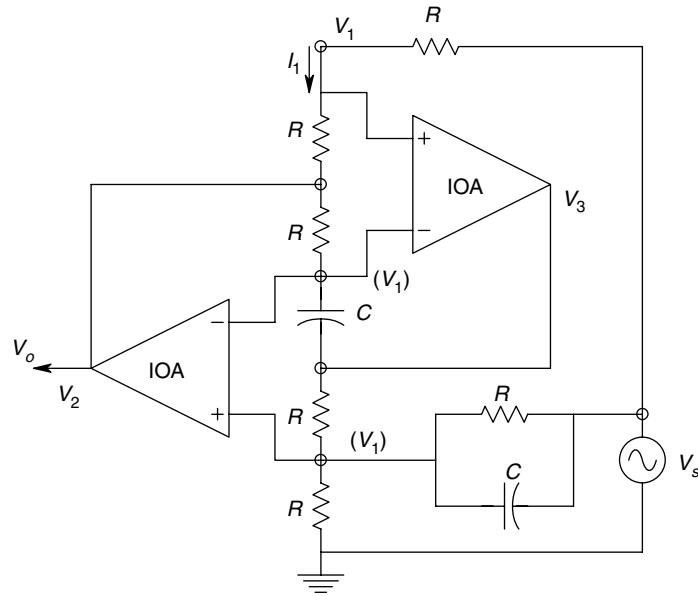


FIGURE 2.23
GIC derived notch filter.

low-pass filter, having the same attenuation slope in the stopband as a comparable Butterworth LPF, requires a lower order (fewer poles).

3. Elliptic or Cauer Filters: In order to attain a higher attenuation (for a given order) in the transition band of frequencies (between passband and stopband), the elliptic filter allows ripple in the frequency response magnitude in both the pass and stopbands. High order elliptic filters have oscillatory transient responses.
4. Bessel or Thompson Filters: In order to achieve a transient response, that is free of overshoot and ringing, Bessel filters trade off steep attenuation slope in the transition band for linear phase response. For any order n , where faithful time domain signal reproduction is required, Bessel filters make good LPFs.

2.4.3 Operational Amplifier Integrators and Differentiators

Often, in conditioning analog measurement records, we wish to either integrate or differentiate these signals. Simple *analog integration* of an input signal can present two practical problems. One is that if the signal contains a dc or average level, the integrator will eventually saturate (i.e. the output voltage of the op-amp is bounded at levels less than its power supply voltage). The second problem with a practical analog integrator is that even when its input is grounded, it will integrate its own dc bias current and offset voltage. Using Figure 2.24, and assuming an otherwise ideal op-amp, it is easy to show that the output of the integrator due to its bias current and offset voltage is given by:

$$V_o(t) = (-t I_B/C_F) + V_{OS} + (t V_{OS}/R C_F) \quad (2.101)$$

Here, we assume zero initial conditions until $t=0$. Clearly, $V_o(t)$ will linearly approach the positive or negative saturation voltage of the op-amp. Use of electrometer or autozero op-amps can minimize analog integrator drift, however, as we will see in Chapter 12, digital integration can eliminate most of the bias current and drift problems.

Analog differentiation also presents practical problems. In *op-amp differentiators*, bias current and drift present little problem. However, the finite bandwidth of a real op-amp can give high frequency transient problems (Figure 2.25). At the summing junction, we can write the node equation as:

$$V'_i(G + sC) - V_o G = V_i sC \quad (2.102)$$

The op-amp output is given by:

$$\frac{V_o}{V'_i} = \frac{-K_V}{\tau_A s + 1} \quad (2.103)$$

If we combine equations 2.101 and 2.102, we can show:

$$\frac{V_o}{V_i} = \frac{-[K_V/(K_V + 1)] sRC}{[s^2 \tau_A R C / (K_V + 1)] + [s(\tau_A + RC)/(K_V + 1)] + 1} \quad (2.104)$$

At low input frequencies,

$$\frac{V_o}{V_i} \approx -sRC \quad (2.105)$$

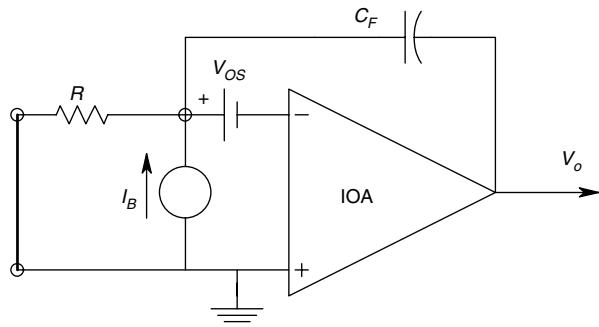
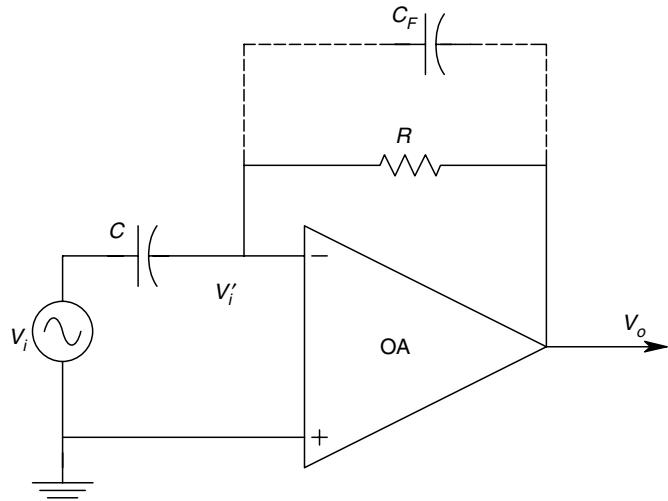


FIGURE 2.24

Op-amp integrator circuit showing sources of dc drift error. The signs and magnitudes of I_B and V_{OS} depend on the particular op-amp design and its temperature.

**FIGURE 2.25**

Op-amp differentiator. C_F is used to damp the high frequency response.

At high input frequencies, the quadratic denominator of the transfer function, equation 2.104, is significant. The break frequency of the transfer function is:

$$\omega_n = \sqrt{\frac{K_V + 1}{\tau_A R C}} \text{ r/s} \quad (2.106)$$

The damping factor of the analog differentiator can be shown to be:

$$\zeta = \frac{(\tau_A + RC)}{2(K_V + 1)} \sqrt{\frac{K_V + 1}{\tau_A R C}} \quad (2.107)$$

To see where the problem arises, let us substitute typical numerical values for the circuit:

$$R = 1 \text{ M}\Omega \quad C = 1 \mu\text{F}, \quad \tau_A = 0.001 \text{ s}, \quad K_V = 10^6$$

Now, we calculate $\omega_n = 3.16 \times 10^4 \text{ r/s}$ and the damping factor is $\zeta = 1.597 \times 10^{-2}$, which is quite low. Thus, any transients in V_i will excite a poorly damped, high frequency, sinusoidal transient at the output of the differentiator. This behavior can present a problem if the differentiator input is not bandwidth limited. If this is the case, then the bandwidth of the circuit must be limited by placing a small capacitor in parallel with R . The computation of the transfer function of the practical differentiator with a parallel RC feedback circuit is left as an exercise for the reader.

2.4.4 Summary

In the preceding subsections, we have reviewed the architectures of three commonly used quadratic, active filter ‘building blocks’ used in the design of analog signal conditioning systems. There are many other AF architectures which are also used by filter designers. However, lack of space prohibits their description here. The interested reader may consult the chapter bibliography for further information.

2.5 Instrumentation Amplifiers

As defined by current practice, instrumentation amplifiers (IAs) are generally described as direct coupled, low noise, differential amplifiers, having very high input impedances, high common-mode rejection ratio, user settable gain in the range 0–60 dB, and bandwidth from dc to hundreds of kHz, depending on design and operating gain. Manufacturers offer IAs as ICs. IAs are often used at the output of a transducer, such as a bridge circuit, as the first stage in an analog signal conditioning system. As head stages, IAs should have low noise characteristics in noise critical applications. Two IA circuits which can be made from low noise op-amps in order to obtain custom low noise IA designs are illustrated below.

Table 2.1 summarizes the systems characteristics of some commercially available instrumentation amplifiers. Note that IA bandwidths are normally not very broad and with some exceptions, cover from dc to the upper audio frequencies. Their slew rates are also modest. However, IAs are generally characterized by high CMRRs and high input impedances.

2.5.1 Instrumentation Amplifiers Made from Op-amps

Figure 2.26 illustrates the well-known, three-op-amp IA circuit. By adjusting the value of R_1 , the practical range of the differential gain of this circuit ranges from 1 to 1000. Note that op-amp 3 is set up as a DA, using matched resistors R_3 and R'_3 , and R_4 and R'_4 . Assuming OA3 is ideal, it is easy to show that its output is given by:

$$V_o = (V_4 - V_3)(R_4/R_3) \quad (2.108)$$

This design also assumes that the resistors R_4 and R'_4 , and R_3 and R'_3 are perfectly matched. The DA circuit of OA3, although having an apparent high CMRR and differential gain, is unsuitable as an instrumentation amplifier because of its unequal and relatively low input resistances. The V_3 sees an input impedance of R_3 to the summing junction of OA3, which is at virtual ground, while V_4 sees an input impedance of $R_3 + R_4$. Op-amps 1 and 2 form a symmetrical circuit which can be analyzed for the cases of pure common-mode input signals and pure difference-mode input signals (note that any combination of V_s and V'_s can be broken down into CM and DM input components to OA1 and OA2, and the responses to these components can be summed at the output by superposition).

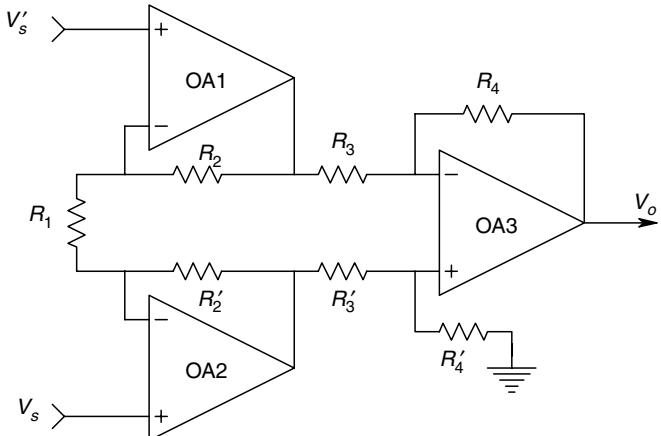
For purely common-mode (CM) inputs, $V_v = V'_s$, and symmetry considerations lead to the conclusion that the summing junction potentials of OA1 and OA2 are equal. Hence there is no current flowing in R_1 . It is assumed here that OA1 and OA2 are ideal op-amps, and that $R_2 = R'_2$. Since no current flows in R_1 under CM input excitation, we may remove it from the circuit (i.e. replace it with an open circuit). Thus OA1 and OA2 are seen to be unity gain followers under CM excitation. Thus, their outputs, V_3 and V_4 , are seen to be equal to each other as well as to the CM input, V_{sc} . The following DA circuit of OA3 subtracts V_3 from V_4 , ideally producing a $V_o = 0$ for CM inputs. Since the op-amps are not ideal and have finite CMRRs, gains, bandwidths, input and output impedances, and because resistor matching may not be perfect, there will be a small V_o for a pure common-mode input.

In the case of a purely difference-mode (DM) input, $V'_s = V_s$, and the summing junctions of OA1 and OA2 are at $-V_s$ and V_s respectively. Hence a current of $2V_s/R_1A$

TABLE 2.1
Selected specifications of selected instrumentation amplifiers

IA Mfr. & Model	CMRR @ Gain dB	R_{in}/C_{in} PF	Small-signal kHz BW @ Gain: V/ μ s	Slew Rate V/ μ s	e_{ina} @ 1 kHz @ Gain: nV/ $\text{RMS}/\sqrt{\text{Hz}}$	i_{ina} pA/ $\sqrt{\text{Hz}}$	V _{os} /Tempco $\mu\text{V}/^\circ\text{C}$	I_B , nA	I_B Tempco
PMI AMP-01	100 @ 1 120 @ 10 130 @ 10^2 - 10^3	20 G Ω , CM 1 G Ω , DM	570 @ 1 100 @ 10 82 @ 100 26 @ 1000	4.5 @ G = 10 0.5 @ G = 1 0.5 @ G = 10 0.25 @ G = 10 ²	540 @ 1 59 @ 10 10 @ 100 5 @ 1000	0.15 @ G = 10 ³	0.3 $\mu\text{V}/^\circ\text{C}$	3	40 pA/ $^\circ\text{C}$
Maxim MAX4462	120	2 G Ω , CM 2 G Ω , DM	2500 @ 1 250 @ 10 25 @ 100	0.5 @ G = 1 0.5 @ G = 10 0.25 @ G = 10 ²	38	—	—	—	—
AD8221	90 @ 1 110 @ 10 130 @ 100 140 @ 1000	100 G Ω /2, CM & DM	825 @ 1 562 @ 10 100 @ 100 14.7 @ 10 ³	2 @ G = 1 2.5 @ G = 10-100	8	40 fA/ $\sqrt{\text{Hz}}$	45 $\mu\text{V}/0.3 \mu\text{V}/^\circ\text{C}$	0.2	1 pA —
AD621B	110 @ 10 130 @ 100	10 G Ω /2	800 @ 10 200 @ 100	1.2	13 @ 10 9 @ 100	100 fA/ $\sqrt{\text{Hz}}$	50 $\mu\text{V}/0.6 \mu\text{V}/^\circ\text{C}$	0.5	1.5 pA/ $^\circ\text{C}$
MN2310	ca. 95 dB	600 M Ω /12 DM 150 M Ω /48 CM	1000	3.5	1.65	4 fA/ $\sqrt{\text{Hz}}$	—	4 pA	—
BB2128	86 @ 1 106 @ 10 125 @ 100 130 @ 1000	10^{10} $\Omega/2$, DM 10^{11} $\Omega/9$, CM	1300 @ 1 700 @ 10 200 @ 100 20 @ 1000	4	8 @ 1000	0.3	$\pm 10 \pm 100/\text{G} \mu\text{V}$	± 2	$\pm 30 \text{ pA}/^\circ\text{C}$
BB INA118	90 @ 1 110 @ 10 120 @ 100 125 @ 1000	10^{10} $\Omega/1$, DM	800 @ 1 500 @ 10 70 @ 100 7 @ 1000	0.9 @ G = 10 —	10 @ 1000	0.3	$\pm 10 \pm 50/\text{G} \mu\text{V}/$ $\pm 0.2 \pm 2/\text{G} \mu\text{V}/^\circ\text{C}$	± 1	$\pm 40 \text{ pA}/^\circ\text{C}$
BB PGA206/207	92 @ 1 96 @ 2 100 @ 4 or 5 100 @ 8 or 10	10^{13} $\Omega/1$, DM 10^{12} $\Omega/4$, CM	5000 @ 1 4000 @ 2 1300 @ 4.5 600 @ 8,10	25 @ G = 1-10 —	18 @ 8	1.5 fA/ $\sqrt{\text{Hz}}$ @ G = 8 ± 0.5 mV/± 2 $\mu\text{V}/^\circ\text{C}$	2 pA	0.4 pA/ $^\circ\text{C}$	

Note: BB = Burr-Brown, AD = Analog Devices, MN = Micro-Networks, PMI = Precision Monolithics

**FIGURE 2.26**

A three op-amp instrumentation amplifier circuit.

flows in R_1 . Furthermore, consideration of Kirchoff's voltage law leads to the conclusion that the center of R_1 (i.e. $R_1/2\Omega$ from either end) is at zero volts potential with respect to ground and thus, may be tied to ground without disturbing the circuit. Now, the DM OA1 and OA2 circuits are seen to consist of non-inverting amplifiers with feedback resistor R_2 and resistance $R_1/2$ to ground from the summing junction. Hence both the DM OA1 and OA2 circuits can easily be shown to have gains:

$$V_3/V'_s = V_4/V_s = \left(1 + \frac{2R_2}{R_1}\right) \quad (2.109)$$

Thus the output for DM inputs is just:

$$V_o = \left(1 + \frac{2R_2}{R_1}\right) (R_4/R_3) (V_s - V'_s) \quad (2.110)$$

The three op-amp IA configuration is used in some commercial IA designs. It provides the designer with some flexibility if special circuit behavior is desired, such as low noise or wide bandwidth. Resistor matching is critical for obtaining a high CMRR. Often a resistor, such as R_4 , can be made variable to tweak the CMRR to a maximum.

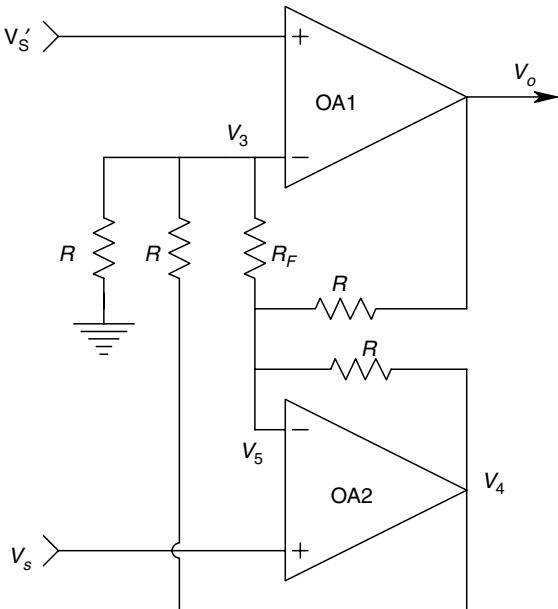
Figure 2.27 illustrates a two op-amp IA design. Again, two matched op-amps must be used, and the resistors, R , must be closely matched to obtain a high CMRR. This circuit may be analyzed using superposition and the ideal op-amp assumption. First it is assumed that $V'_s > 0$ and $V_s \rightarrow 0$. Thus, $V_3 = V'_s$ and $V_5 = 0$. Now, the node equations on the V_3 and the V_5 nodes can be written as:

$$V'_s(2G + G_F) - V_4G - V_5G_F = 0 \quad (2.111)$$

$$V_5(2G + G_F) - V_3G_F - V_oG - V_4G = 0 \quad (2.112)$$

Simultaneous solution of equations 2.111 and 2.112 gives:

$$V_o = -V'_s 2 \left(1 + \frac{R}{R_F}\right) \quad (2.113)$$

**FIGURE 2.27**

A two op-amp instrumentation amplifier circuit.

In a similar manner, we set $V_s > 0$ and $V'_s \rightarrow 0$, and note that $V_5 = V_s$ and $V_3 = 0$. Node equations are written on the V_3 and V_5 nodes:

$$V_3(2G + G_F) - V_4G - V_5G_F = 0 \quad (2.114)$$

$$V_5(2G + G_F) - V_3G_F - V_oG - V_4G = 0 \quad (2.115)$$

Solution of these equations gives us:

$$V_o = V_s 2 \left(1 + \frac{R}{R_F} \right) \quad (2.116)$$

Hence by superposition, the outputs of the two op-amp IAs can be written as:

$$V_o = (V_s - V'_s) 2 \left(1 + \frac{R}{R_F} \right) \quad (2.117)$$

Note that for both the above IA designs, the inputs are made directly to the non-inverting inputs of op-amps. Hence the input impedance of these IAs will essentially be that of the op-amp. If FET input op-amps are used, then input resistances in excess of $10^{12} \Omega$ can be attained with very low bias currents.

2.5.2 Isolation Amplifiers

Isolation amplifiers are a special class of IAs which find extensive application in biomedical instrumentation, specifically, the recording of low level bioelectric signals from human and animal subjects. Their use is generally dictated by electrical safety considerations. They are also used when the input signal is found in the presence of a high common-mode voltage, as well as in a situation where ground loop currents can introduce errors in the signals under measurement. Isolation amplifiers have an extremely low conductance between the reference ground of the output signal and the

ground and input terminals of the differential headstage. The dc conductive isolation between the headstage ground and the output stage ground is generally accomplished by a magnetic transformer coupling or by opto-coupling techniques.

In the Analog Devices family of isolation amplifiers, the output of the input DA is coupled to the output stage of the isolation amplifier by an isolation transformer. The operating power for the input DA is also coupled with the transformer from a high frequency power converter oscillator, which derives its power from the non-isolated or normal supply for the isolation amplifier. The block diagram of an Analog Devices, AD295 precision hybrid IA, is shown in Figure 2.28. Note that in this design, both the input DA and the output circuit have isolated (internal) power supplies.

To illustrate the differences between ordinary instrumentation amplifiers and isolation amplifiers, some of the key specifications for the Analog Devices Model AD294A Medical Isolation Amplifier are reviewed below.

The overall gain is the product of G_{IN} and G_{OUT} .

$$G_{IN} = (1 + 10^5/R_G), \quad R_G \geq 1\text{k}\Omega$$

$$G_{INmax} = 100, \quad G_{OUT} = \left(1 + \frac{R_A}{R_B}\right), \quad 1 < G_{OUT} < 10$$

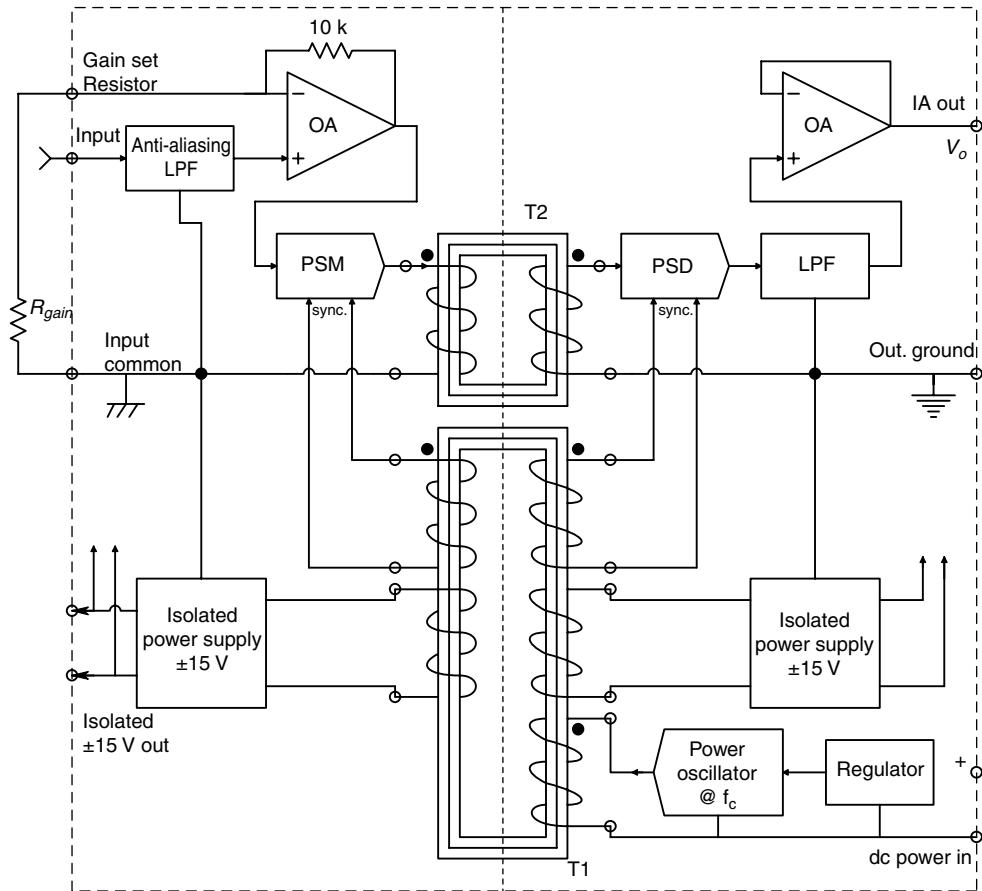


FIGURE 2.28

Simplified block diagram of an Analog Devices AD295 precision isolation amplifier. (Source: Figure used with permission of Analog Devices)

Hence the overall gain may be 1000. The CMRR of the AD294A is 100 dB, not exceptionally high for an IA. The maximum continuous common-mode voltage (CMV) is ± 3500 Vpk. The amplifier is rated to stand a ± 8000 Vpk, 10 ms pulse at the rate of 1 pulse every 10 seconds. The input impedance is $10^8 \Omega || 150$ pF. The input (dc) bias current I_B , is 2 nA. The small signal bandwidth is 2.5 kHz (gain 1–100 V/V). The slew rate is 9.1 V/ μ s. The maximum leakage current between headstage and output stage is 2 μ A rms, given that 240 V rms 60 Hz is applied to the input. The AD 294A is typically used to record electrocardiographic or electromyographic signals using body surface electrodes.

Another medical isolation amplifier suitable not only for ECG recording, but also EEG signal conditioning, is the Intronics Model IA296. This isolation amplifier also uses the two-transformer architecture to couple power into the differential headstage and signals out. Its specifications are summarized below.

The gain of IA296 is fixed at 10. The CMRR is 160 dB with a 5 k Ω source resistance imbalance. The maximum safe dc common-mode input voltage is ± 5000 V. The CM input impedance is $10^{11} \Omega || 10$ pF and the DM input impedance is $3 \times 10^8 \Omega || 2.2$ pF. The input bias current is ± 200 pA and the I_B tempco is ± 5 pA/ $^{\circ}$ C. The –3 dB bandwidth is 1 kHz. Input noise with a single-ended input with $R_s = 5$ k Ω is 0.3 μ V in a 10 to 1 kHz bandwidth (about 9.5 nV/ \sqrt Hz); the current noise is about 0.13 pA/ \sqrt Hz in a 0.05 Hz to 1 kHz bandwidth. Maximum leakage current due to component failure is 10 μ A. Due to its low noise, the Intronics IA296 is better suited for measurements of low level biological signals such as fetal ECGs, EEGs and evoked brain potentials.

Finally, as an example of a photo-optically coupled isolation amplifier, we examine the Burr-Brown Model 3652JG. Since this amplifier does not use a built-in transformer to couple power to the input DA, an external, isolating dc/dc converter must be used, such as the Burr-Brown Model 722. The BB722 supply uses a 900 kHz oscillator to drive a transformer which is coupled to two independent rectifiers and filters. It can supply ± 5 to ± 16 V dc, at about 16 mA to each of the four outputs. It has isolation test ratings of ± 8000 Vpk for 5 seconds, and ± 3500 Vpk continuous between inputs and outputs. The leakage current maximum for this power supply is 1 μ A for 240 V rms at 60 Hz, insuring medical safety. The Model 3652 isolation amplifier has the following specifications when used with the BB722 dc/dc power supply.

Gain is settable in the range 1–1000. The CMRR is 80 dB. The BB3652 can block a CM voltage of ± 2000 V dc and a pulse of ± 5000 V for 10 ms. The input resistance is given as $10^{11} \Omega$, CM and DM. The isolation impedance between input and output is $10^{12} \Omega || 1.8$ pF. The input bias current is 10 pA. The offset voltage tempco is 25 μ V/ $^{\circ}$ C. The –3 dB bandwidth is 15 kHz at all gains. The short circuit voltage noise is 5 μ V rms, measured over a 10 Hz to 10 kHz bandwidth. No figure is given for the input current noise. Maximum leakage current is 0.35 μ A.

2.5.3 AutoZero Amplifiers

Yet another type of IA merits attention. The commutating autozero (CAZ) amplifier employs a rather unique design to obtain an almost negligible dc input offset voltage drift. The Intersil ICL7605/7606 CAZ amplifiers make use of two, novel, switched capacitor networks to:

1. Convert a true differential input to a single-ended signal, which is the actual input to the CAZ amplifier

2. The second pair of switched capacitors is used in conjunction with two matched op-amps to cancel out the offset voltage of each op-amp.

This process is illustrated in Figures 2.29A and B. During the cycle shown in Figure 2.29A, C_2 is forced to charge up to $-V_{os2}$, while C_1 , which is charged to $-V_{os1}$, is switched in series with the input, V_1 and V_{os1} , to cancel out V_{os1} . In the cycle shown in 'B', C_1 is charged up to $-V_{os1}$, and C_2 which is now charged to $-V_{os2}$ is put in series with V_{os2} and V_1 to cancel V_{os2} , etc.

The maximum CMRR is 104 dB, and the amplifier bandwidth is 10 Hz. The offset voltage tempco is an amazing 10 nV/°C. The commutation clock frequency of the Intersil CAZ amplifier can range 160–2560 Hz. Overall gain can be set between 1 and 1000. The ICL7605/7606 CAZ amplifier is intended for use with dc measurement applications such as strain gauge bridges used in electronic scales.

National Semiconductor offers the LMC669 auto-zero module, which can be used with any op-amp configured as a single-ended inverting summer, or as a non-inverting amplifier. Figures 2.30A and B illustrate two applications of the LMC669 auto-zero circuit. Note that the auto-zero module does not limit the dynamic performance of the op-amp circuit to which it is attached; the amplifier retains the same dc gain, small signal bandwidth and slew rate. The effective offset voltage drift of any op-amp using the auto-zero is 100 nV/°C. The maximum offset voltage using the auto-zero is $\pm 5 \mu\text{V}$. The bias currents of the auto-zero are 5 pA (these must be added to the op-amp's IB); its clock frequency can be set from 100 Hz to 100 kHz. The NS LMC669 auto-zero is very useful to compensate for dc drift in circuits using special purpose op-amps, such as electrometer op-amps, which normally have large offset voltages (200 μV) and large V_{os} tempcos (5 $\mu\text{V}/^\circ\text{C}$). The auto-zero does contribute to an increase in the equivalent input voltage noise of the circuit. However, choice of sampling rate, step size and the use of low-pass filtering in the feedback path can minimize this effect.

2.5.4 Absolute Isolation

One scheme that can produce *absolute isolation* between the transducer, headstage, system under test environment and the signal conditioning, digital interface and computer, is the fiber optic cable (FOC). The transducer and headstage are powered by batteries. The headstage analog output modulates the intensity of an LED that projects light down the FOC. Several modulation schemes may be used, including pulse frequency modulation, sigma-delta modulation, or double-sideband, suppressed carrier modulation. A photodiode responds to the modulated light signal on the near end of the FOC. Its output is directed to a suitable demodulator and the signal is recovered. The FOC essentially provides zero conductance between the system under test and the signal conditioner.

2.5.5 Summary

In this section, we have examined some examples of instrumentation amplifiers. IAs are defined as differential amplifiers with high input impedances, high CMRR and programmable gain, usually set by a choice of one or two resistors. Their bandwidths generally depend on the gain setting, and roughly follow the constant gain-bandwidth rule. Typical unity gain bandwidths are in the hundreds of kHz, although some are now available in the 3 MHz range. Special isolation IAs and auto-zeroing amplifiers are also available.

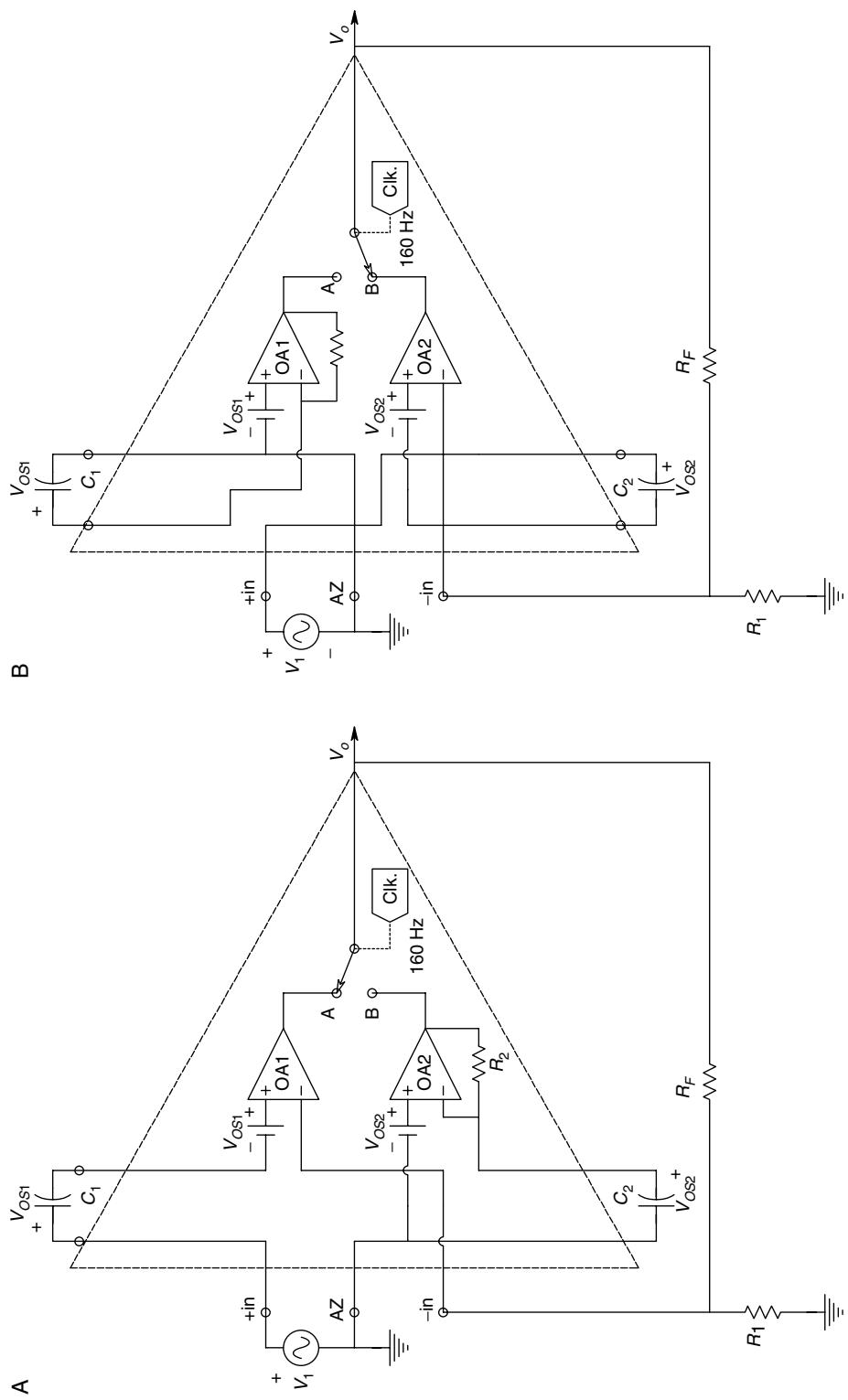
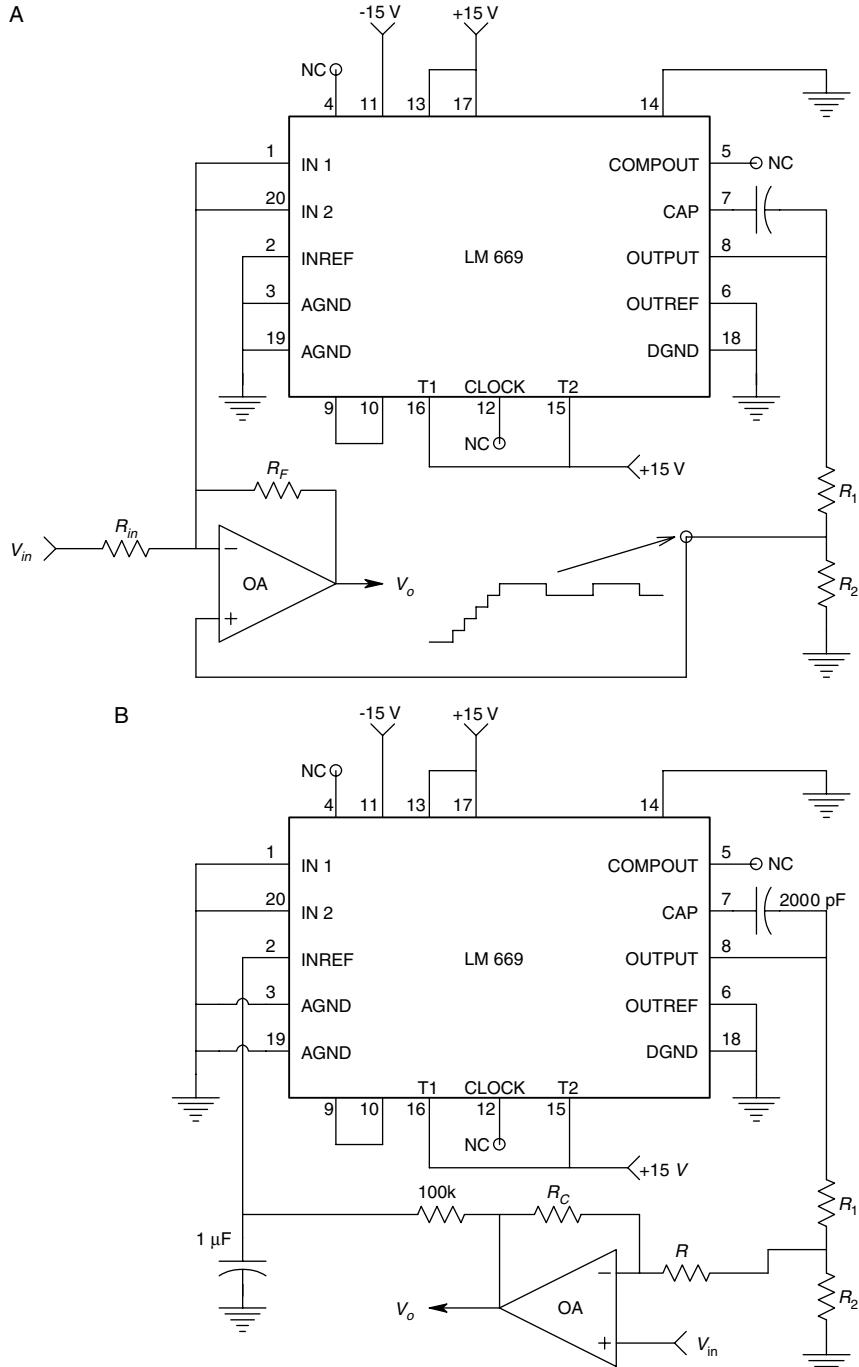


FIGURE 2.29
A. Schematic of the innards of an Intersil ICL7605/7606 commutating auto-zero amplifier. In this switching cycle, OA1 is connected to the output. OA2 is connected such that its dc offset voltage charges up C_2 . C_1 is switched in series with the input node of OA1 so that the voltage on C_1 cancels V_{oSI} . B. In the following switching cycle, OA2 is connected in series with C_2 , so $V_{oSI}(k)$ of OA2 subtracts from $V_{oSI}(k-1)$ on C_2 .

**FIGURE 2.30**

A. The LM669 auto-zero IC used to cancel out offset drift in an inverting op-amp application. The LM669 continuously samples the dc voltage at the op-amp's summing junction, and then generates a compensatory dc voltage at the non-inverting input. This action nulls the op-amp's V_{OS} to effectively $\pm 5 \mu\text{V}$. This offset voltage correction compensates for the drift of V_{OS} with temperature and power supply voltage changes. B. The LM669 acts as a dc servo integrating feedback loop around an op-amp used as a non-inverting amplifier for ac signals. The dc output error of the op-amp is reduced to about $5 \mu\text{V}$, the V_{OS} of the LM669. The LM669 replaces the ground reference for the resistor, ' R' .

2.6 Nonlinear Analog Signal Processing by Op-amps and by Special Function Modules

2.6.1 Introduction

Perhaps the best way to define what we mean by nonlinear signal processing is to first describe the properties of a *linear system*. Linear system (LS) outputs can be found, in general, by the operation of *real convolution*. Real convolution can be derived from the general property of linear systems, which is *superposition*. The operation of convolution can be expressed by the well-known definite integral:

$$y_1(t) = x_1 \otimes h = \int_{-\infty}^{\infty} x_1(\sigma)h(t - \sigma) d\sigma \quad (2.118)$$

where $y_1(t)$ = the system output

$x_1(t)$ = input to an LS

$h(t)$ = impulse response

\otimes = operation of real convolution.

The property of *superposition* is illustrated by equation 2.119.

We note that:

$$y_3 = (a_1 y_1 + a_2 y_2) \quad (2.119)$$

given input $x_3 = (a_1 x_1 + a_2 x_2)$. Linear systems outputs can be *scaled* (i.e. if $x_2 = ax_1$), then:

$$y_2 = a(x_1 \otimes h) \quad (2.120)$$

Finally, linear systems obey the property of *shift invariance*. If the input x is delayed by t_o seconds, then the output is also shifted as:

$$y(t - t_o) = x(t - t_o) \otimes h \quad (2.121)$$

Nonlinear systems do not obey the properties of linear systems discussed above. Nonlinear systems can be dynamic, static, or a combination of both. A static nonlinear system is one in which the output, y , is a nonlinear function of the input, x . For example, $y = \sin x$, $y = \text{sgn } x$, $y = a + bx + cx^2$ and $y = \text{abs } x$ describe nonlinear, static relations between input x and output y . $y = mx + b$ is a *linear* static function of x .

A dynamic nonlinearity is one described by a set of first order, nonlinear ordinary differential equations (ODEs). An example of such a system is the well-known Volterra equations describing predator-prey relations in a simple ecosystem. These are:

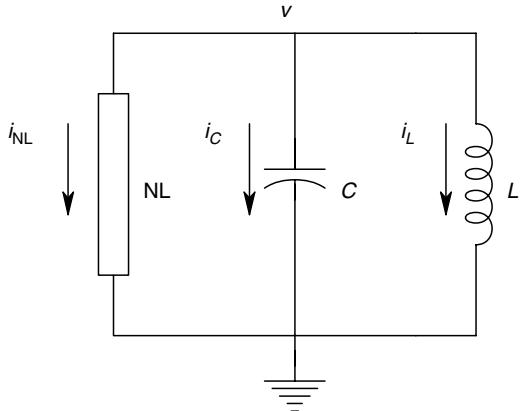
$$\dot{x}_1 = k_1 x_1 - k_3 x_1 x_2 \quad (2.122)$$

$$\dot{x}_2 = k_2 x_2 - k_4 x_1 x_2 \quad (2.123)$$

The Volterra system responds to initial conditions on x_1 and x_2 .

Another example of a dynamic nonlinear system is illustrated in Figure 2.31. A linear, parallel LC circuit is connected to a negative resistance element, whose current is determined by the relation:

$$i_{nl} = -av + bv^3 \quad (2.124)$$

**FIGURE 2.31**

A simple, nonlinear RLC circuit which can be used to illustrate the van der Pol equation for nonlinear oscillations.

Hence by Kirchoff's current law,

$$0 = \ddot{v} - \dot{v}(1 - 3bv^2/a)(a/C) + v/LC \quad (2.125)$$

This relation can be written as two first order ODEs:

$$\dot{x}_1 = -x_2/LC + x_1(1 - 3bx_2^2/a)(a/C) \quad (2.126)$$

$$\dot{x}_2 = x_1 \quad (2.127)$$

where obviously, $x_2 = v$. Equation 2.125 is the well-known *van der Pol equation*.

Often, nonlinear systems can be modeled by separating the linear dynamics from a no-memory nonlinearity, such as $y = \tanh x$, which can either precede or follow the linear block, or lie between two linear blocks, depending on circumstances.

There are several important nonlinear analog signal processing operations which are used in instrumentation. These include, but are not limited to, Precision Full-Wave Rectification or Absolute Value circuits, True RMS to dc Conversion circuits, Peak Detection circuits, Sample and Hold or Track and Hold circuits, Square Root circuits, Special Function Modules which give an analog output proportional to $x(y/z)^m$, Log Ratio Modules and Trigonometric Function Modules.

We will review the applications of each of the nonlinear signal processing circuits before discussing the architecture and design details. Certain nonlinear operations may easily be realized with op-amp circuits, while for others, it is more expedient to use off-the-shelf, special purpose microcircuits. We cite examples of these microcircuits, but for practical reasons, it is not possible to list all those available for a given type.

2.6.2 Precision Absolute Value (Absval) Circuits

As the name implies, this type of nonlinear circuit takes the absolute value (absval) of the input waveform. Thus the output approximates to:

$$y(t) = k \text{abs}[x(t)] = k|x(t)| \quad (2.128)$$

where k is a gain or scaling constant.

Precision absval circuits can be used for *transfer standards*. That is, they perform nearly ideal, full-wave rectification of symmetrical sinusoidal signals, which after

low-pass filtering, appear as a dc voltage proportional to the peak or RMS value of the input sine wave. This dc voltage can generally be measured with greater accuracy than the corresponding ac voltage, especially when the peak value is less than one volt. Precision absval circuits have been used in adaptive analog active filters, such as the Dolby BTM audio noise suppression system. Here, the absval/low-pass filter system output is proportional to the signal power in a certain range of high frequencies, and is used to adjust the high frequency response of the filter. Precision absval circuits can also be used to precede square law circuits used to square time variable signals. When this is done, only half of a parabolic nonlinearity needs to be used because:

$$y = [\text{abs}(x)]^2 = x^2 \quad (2.129)$$

Many precision absval circuits using op-amps exist (Sections 5.2–5.4, Graeme, 1977). Several of the more useful designs are examined here. Figure 2.32 illustrates the circuit of a commonly used absval operator which uses two op-amps. Its input resistance is just $R/2$, and its output is ideally given by:

$$V_o = +(R_2/R)|V_1| \quad (2.130)$$

However, at low levels of V_1 , some tweaking of the $R/2$ resistor may be required to obtain equal and opposite slopes of the absval curve.

Another absval circuit having high input resistance is shown in Figure 2.33. In this circuit, correct resistor values are important. The parameter n is the rectifier gain, and must be >1 . A small capacitor, C , is used to improve high frequency response of the circuit. R_1 and R_3 can be made variable in order to adjust the absval operation to be symmetrical (i.e. have slopes of $+n$ or $-n$).

Finally, Figure 2.34 illustrates an absval circuit having a differential input. An instrumentation amplifier with a high CMRR and differential gain K_D is used to condition the floating absval voltage across R_L . OA1 is used as a follower, while OA2 forces a unidirectional current through R_L . The output of this circuit is easily shown to be:

$$V_o = K_D(R_L/R)|V_1 - V'_1| \quad (2.131)$$

In describing the absval circuits above, second order errors caused by non-zero offset voltages and bias currents have been neglected. Also neglected are the effects of finite op-amp dynamic response and diode capacitance. If an absval circuit is to be used

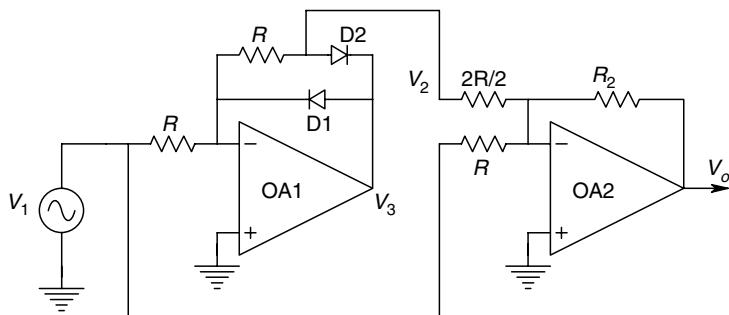
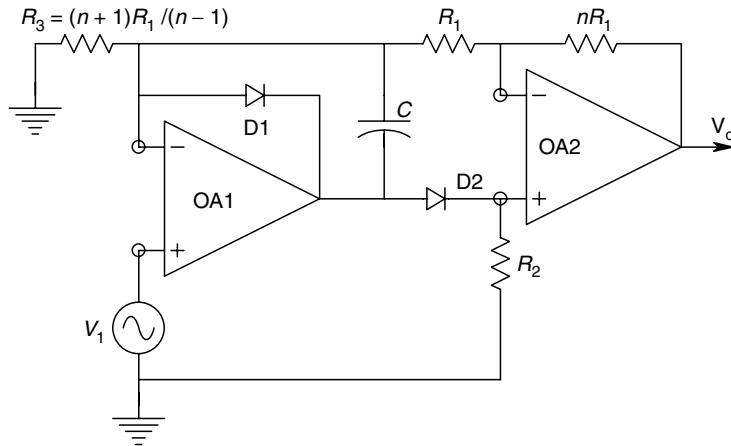
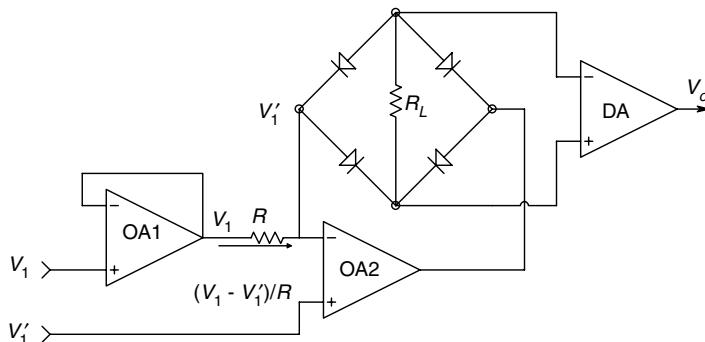


FIGURE 2.32

A precision, operational, full wave rectifier or absval circuit.

**FIGURE 2.33**

A precision absval circuit having high input impedance.

**FIGURE 2.34**

A precision absval circuit having a high impedance, differential input.

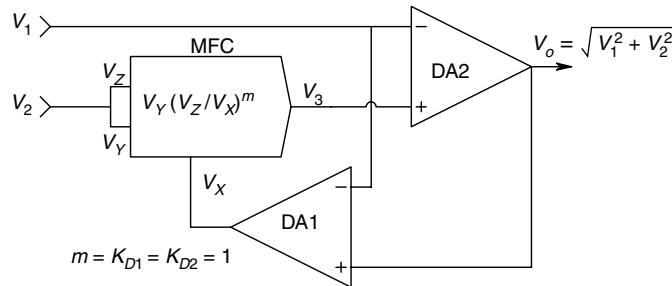
at high audio frequencies and above, then the designer must choose the op-amps with care and use high frequency switching diodes in the design. The designer should also simulate the high frequency behavior of the absval circuit in the time domain using an electronic circuit analysis program, such as a version of SPICE or MicroCapTM, in order to verify that it will meet specifications.

2.6.3 Multifunction Converters

Multifunction converters are LSI analog integrated circuits which perform the operation:

$$V_o = KV_Y(V_Z/V_X)^m \quad (2.132)$$

The input voltages can range from 0 to 10.0 V, and from dc to 400 or 500 kHz. The exponent, m , can range from 0.2 to 5.0. Commercial examples of multifunction converters are the Burr-Brown Model 4302 and the Analog Devices AD538 Analog Computational Unit.

**FIGURE 2.35**

A multifunction converter used in a real time, Pythagorean system. $V_o = \sqrt{V_1^2 + V_2^2}$, when $m = K_{D1} = K_{D2} = 1$.

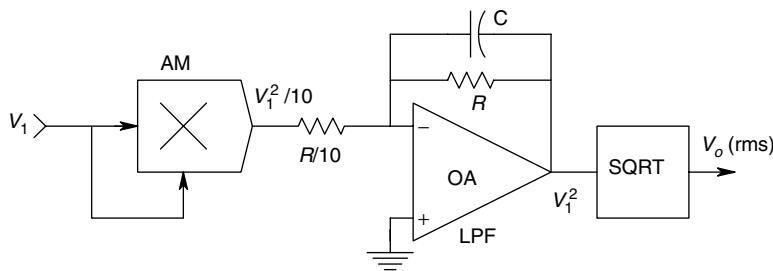
The Multifunction Converter is a useful IC which can be used as a one-quadrant multiplier, squarer, square rooter, or exponentiator, depending on the value of m and the accessory op-amp circuits used. The multifunction converter can also be used to produce an output voltage proportional to the sine, cosine, or arctangent of an input voltage using an external op-amp. In addition, using implicit analog computational techniques involving feedback, the true RMS of an input signal may be computed (discussed in the following section), as well as real time vector magnitude, as shown in Figure 2.35, where:

$$V_o = \sqrt{V_1^2 + V_2^2} \quad (2.133)$$

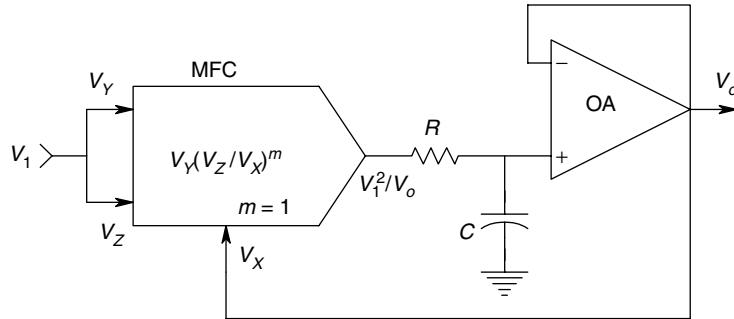
2.6.4 True RMS to dc Converters

True RMS converters are widely used in the quantitative amplitude measurement of non-sinusoidal signals and noise. Electrical noise may arise from resistors, devices (e.g. transistors), capacitive or magnetic pickup from environmental noise sources, such as sparking motor commutators, or slip-ring contacts. It is also important to measure acoustical noise (e.g. noise on a factory floor, mechanical vibrations of bearings), etc. Often, it is useful to measure the RMS value of noise-like bioelectric waveforms, such as muscle action potentials (EMGs), in order to diagnose problems with the neuromuscular system of a patient. EMGs can also be used to actuate prosthetic limbs and robotic manipulators. Yet another use for true RMS converters is in the design of analog adaptive filters, where the RMS voltage measured in some frequency band is used to adjust a feature of the overall frequency response of a filter.

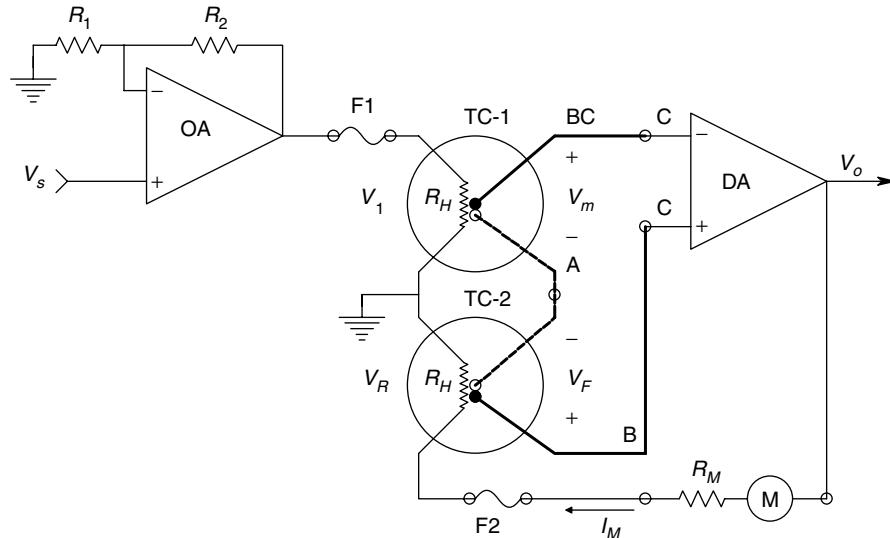
The operations on a signal done by an RMS converter are, in order of operation, squaring, smoothing or averaging by low-pass filtering and then square rooting. The direct method of computing the analog true RMS of a time varying signal is shown in Figure 2.36. Another method of finding the RMS value of a time varying signal is to

**FIGURE 2.36**

Circuit for the direct generation of the RMS value of an ac signal. Circuit output is a dc voltage.

**FIGURE 2.37**

R and C form a low-pass filter which performs the 'mean' operation in this feedback, true RMS Converter, which uses a multifunction converter; $m \equiv 1$.

**FIGURE 2.38**

Basic schematic diagram of a true RMS voltmeter using two matched vacuum thermocouples in a feedback system.

use the implicit method illustrated in Figure 2.37. Here, a multifunction module is used with feedback of the dc output voltage to force square rooting to occur; $m \equiv 1$ in this application. Yet another implicit method of computing the true RMS value of a signal is the feedback thermocouple method shown in Figure 2.38. The thermocouple method is also used in the design of true RMS voltmeters. In the feedback thermocouple circuit, we may write:

$$V_{m(\text{dc})} = K_T(\overline{V_1^2}) \quad (2.134 \text{ A})$$

$$V_{F(\text{dc})} = K_T(\overline{V_R^2}) = K_T I_m^2 R_H^2 \quad (2.134 \text{ B})$$

Now the dc output current of the DA is

$$I_m = K_D(V_m - V_F)/R = K_D K_T(V_1^2 - I_m^2 R_H^2)/R \quad (2.135)$$

where R = sum of the thermocouple heater resistance,

$R_H R_m$ = microammeter resistance

R_o = DA output resistance

Solving the above quadratic equation in I_m yields:

$$I_m = -R/(2K_D K_T R_H^2) \pm (1/2) \sqrt{\frac{R_2}{K_D^2 K_T^2 R_H^4} + \frac{4(V_1^2)}{R_H^2}} \quad (2.136)$$

The exact solution for I_m reduces to,

$$I_m = V_{1(\text{RMS})}/R_H \quad (2.137)$$

for the condition,

$$V_1^2 \gg \frac{R_2}{4K_D^2 K_T^2 R_H^4} \quad (2.138)$$

which is easily met. If an analog microammeter is not used to read the system output, then it is easy to see that the DA output voltage is given by:

$$V_o = V_{1(\text{RMS})} R / R_H \quad (2.139)$$

Several specialized true RMS to dc converter ICs are available. The Burr-Brown Model 4341 operates by first taking the absval of the input signal, then forming a voltage proportional to twice the log of the absval. An antilog transistor is used in a feedback circuit. The dc output voltage of the converter, V_o , is fed back to the base of the anti-logging transistor, so that its collector current is given by:

$$i_c(t) = k \log^{-1}(\log V_1^2 - \log V_o) = k(V_1^2/V_o) \quad (2.140)$$

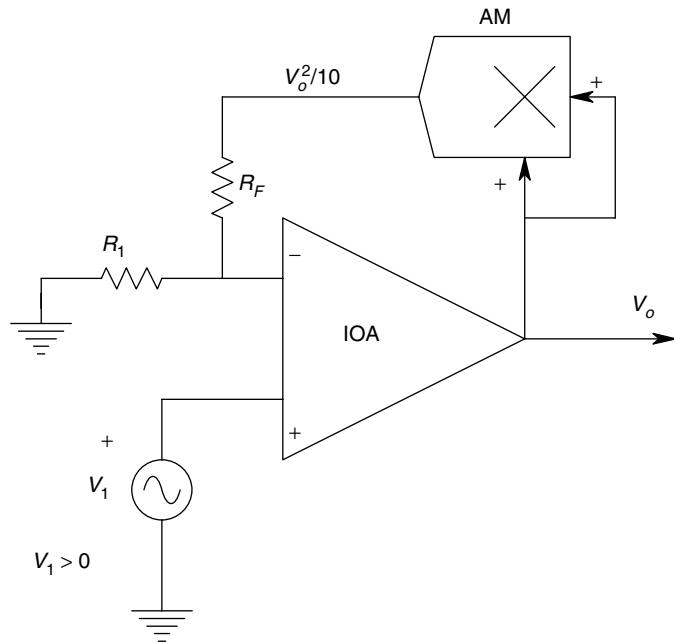
The $i_c(t)$ is averaged by an op-amp current to voltage low-pass filter to form V_o . It is easy to see that V_o is proportional to the true RMS value of V_1 . Of course, the BB 4341 true RMS converter works with dc inputs and is down 1% in frequency response, at 80 kHz and -3 dB at 450 kHz. To be useful at low frequencies, the output must not contain ripple. Hence, the capacitor in the low-pass (averaging) filter must be made very large. For example, for negligible output ripple, given a 1 V, 10 Hz input sine wave, the filter capacitor should be over 100 μF .

Analog Devices also offers several IC, true RMS to dc converters, having basically the same internal circuit architecture as the BB 4341. For example, the frequency response of the high accuracy AD637 extends to 8 MHz. Its maximum output is 7 V RMS, and its accuracy is $5 \text{ mV} \pm 0.2\%$ of reading, with 0.02% nonlinearity for 0–2 V RMS input.

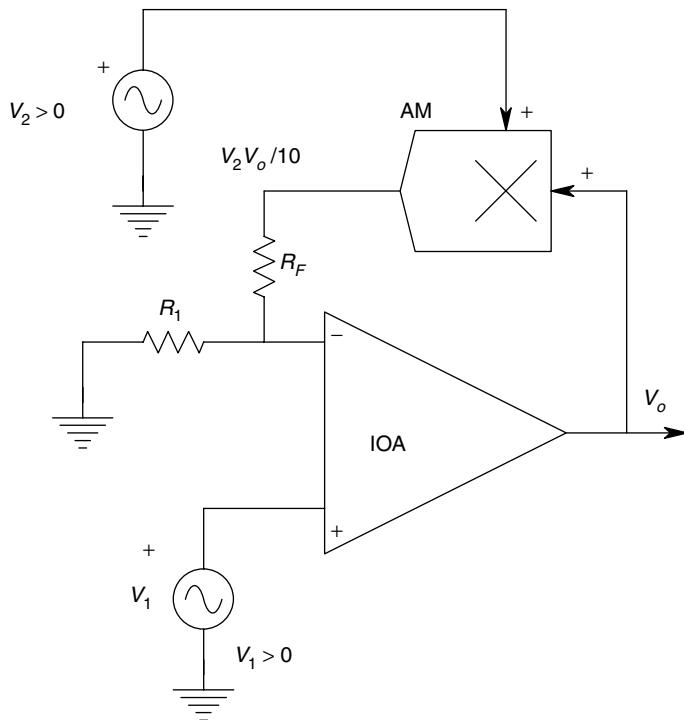
2.6.5 Square Root Circuits and Dividers

The square root of a voltage signal may be taken in several ways. Obviously, a multifunction converter with $m=0.5$ can be used. Another implicit method of square rooting is shown in Figure 2.39. Here, an analog multiplier is used to square the output and feed it back to the summing junction. Assuming that the summing junction is at $V_1 \geq 0$, the output voltage can be found from the node equation below:

$$V_1/R_1 = V_o^2/10R_F \quad (2.141)$$

**FIGURE 2.39**

Feedback square rooting circuit; V_1 must be non-negative.

**FIGURE 2.40**

The circuit of Figure 2.39 is modified to make an analog divider. Here V_1 must be non-negative, V_2 must be positive and greater than $V_{2\min} = (V_{1\min}/V_{osat})(10R_F/R_1)$ to prevent op-amp output saturation, and less than the supply voltage of the op-amp (e.g. +15 V) to avoid damaging the op-amp's input transistors.

This leads to the desired result:

$$V_o = \sqrt{V_1(10R_F/R_1)} \quad (2.142)$$

Analog division may also be accomplished by an implicit feedback circuit, an example of which is shown in Figure 2.40. Here, V_1 and V_2 must be >0 . Assuming an ideal op-amp, the node equation at the summing junction is:

$$V_1/R_1 = V_2 V_o / 10 R_F \quad (2.143)$$

Hence

$$V_o = (V_1/V_2)(10R_F/R_1) \quad (2.144)$$

Of course, a multifunction module with $m=1$ will also give the same result.

2.6.6 Peak Detectors and Track and Hold (T&H) Circuits

These circuits are closely related, and will be treated together. Peak detectors follow an irregular waveform until a maximum (or minimum, if a negative peak detector is used) is reached and then, that value is held at its output until an even greater maximum comes along, then that value is held and so on. Peak detectors are used in peak-reading voltmeters and in applications where measurement of the peak value of a waveform has significance (as opposed to the RMS value or rectified average value). An example of a waveform where the maximum and minimum values are important is arterial blood pressure, recorded during surgery. The peak value is the systolic pressure, the minimum value is the diastolic pressure. Figure 2.41 illustrates a two op-amp peak detector of conventional design. Note that OA2 is used as a voltage follower. It should be an FET input type, with a high input resistance and low bias current. Before a peak is reached in V_1 , D_1 is reverse biased by the forward drop on D_2 and $V_3 = V_C = V_o = V_1$. After the peak in V_1 is reached, $V_1(t)$ decreases, D_2 is reverse biased and D_1 is forward biased, so that OA1 does not saturate. The output voltage remains at V_{1pk} , assuming zero capacitor leakage current, diode D_2 reverse current and op-amp bias current. The output

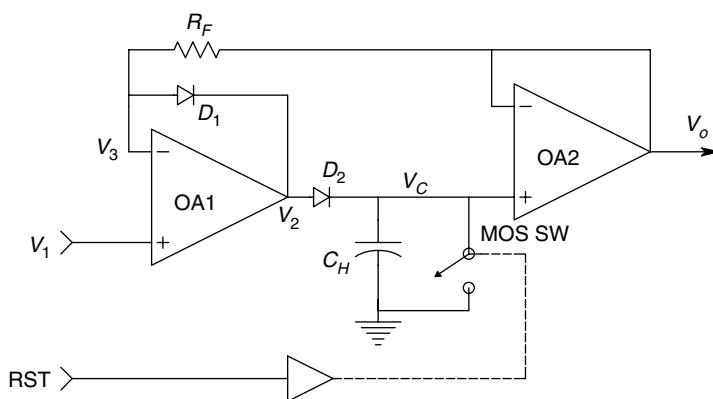
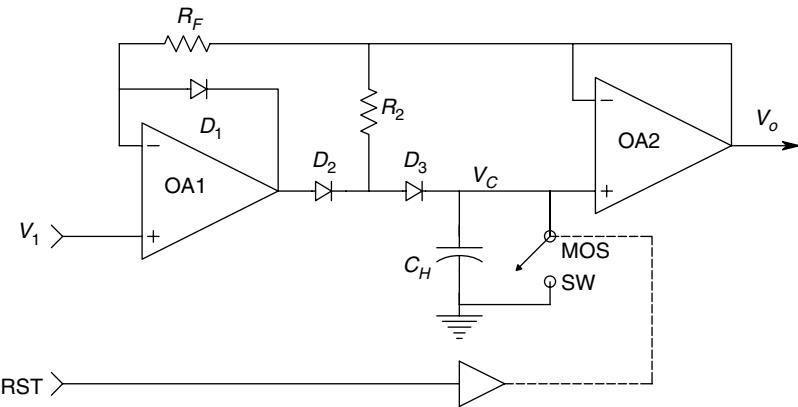


FIGURE 2.41

A MOSFET switch, SW1, is used to reset the two op-amp peak detector.

**FIGURE 2.42**

An improved peak detector design.

voltage remains at $V_{1\text{pk}}$. In practice, these currents will cause V_C to drift slowly. The reset switch, SW1, is generally a MOSFET switch, which is used to reset V_o to zero when required.

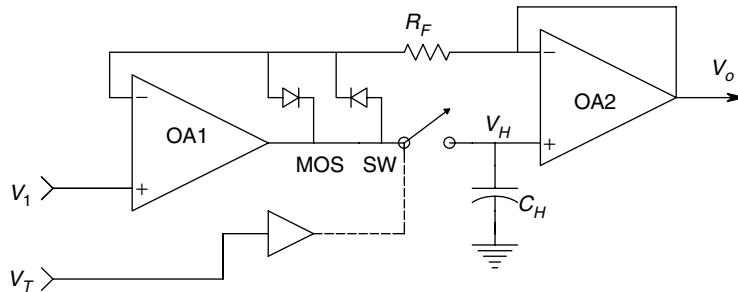
The addition of a third diode and a resistor R_2 can improve the performance of the peak detector by eliminating the effective leakage current through D_2 . Figure 2.42 shows that D_3 is held at zero bias by R_2 which pulls the anode of D_3 to $V_o = V_C$ after the peak is reached. Therefore, the overall rate of change in the peak detector output voltage is, practically:

$$dV_o/dt = I_B/C_H \text{ V/sec.} \quad (2.145)$$

There is a trade-off in determining the proper size of C_H . A large, low leakage C_H will minimize the drift, but is expensive. Also, too large a C_H will not allow V_C to follow V_1 below the peak due to current saturation in OA1. Typical C_H values are in the range 1 nF–0.1 μ F. A good discussion on the design of peak detectors can be found in Franco, 1988.

Track and hold (T&H) (also called sample and hold) circuits are important components of many analog to digital conversion systems. In operation, the output of a T&H circuit follows the conditioned analog input signal, until the hold command is given. After a short delay, the T&H output assumes a constant level, during which the A/D conversion takes place. Constancy is required because certain ADC designs compare the analog input signal to an analog signal from a digital to analog converter in the ADC, whose output is determined from the output digital ‘word’. Any change in the analog input of the ADC during the conversion process will result in error. Ideally, the digital output of an ADC is the result of an impulse modulation, where the analog input signal is multiplied by a periodic train of unit impulses. The output of the impulse modulator is a periodic sequence of numbers whose values represent the values of the analog input signal only at sampling instants. A T&H circuit allows a closer approximation of true impulse modulation to take place.

Figure 2.43 illustrates a basic T&H circuit. When the FET sampling switch is closed, the system is in the tracking mode. OA2 acts as a unity gain buffer, so $V_o = V_C = V_1$, as long as the op-amp slew rates are not exceeded, and the input signal frequencies lie well below the GBWPs s of the op-amps. The C_H provides a capacitive load for OA1 and thus, must be able to provide enough output current to satisfy $[C_H(dV_1/dt)_{\text{max}}]$. When

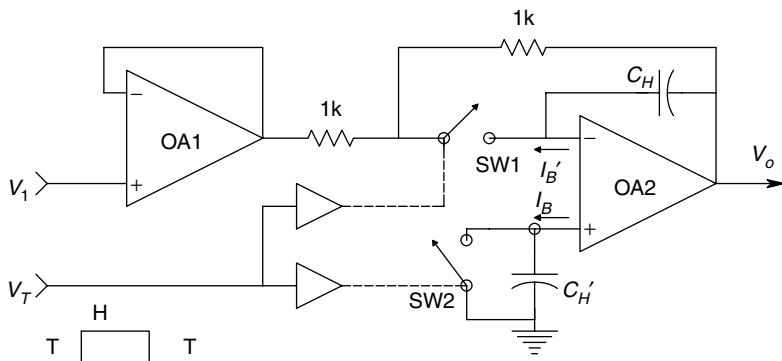
**FIGURE 2.43**

A basic, unity gain, track and hold circuit.

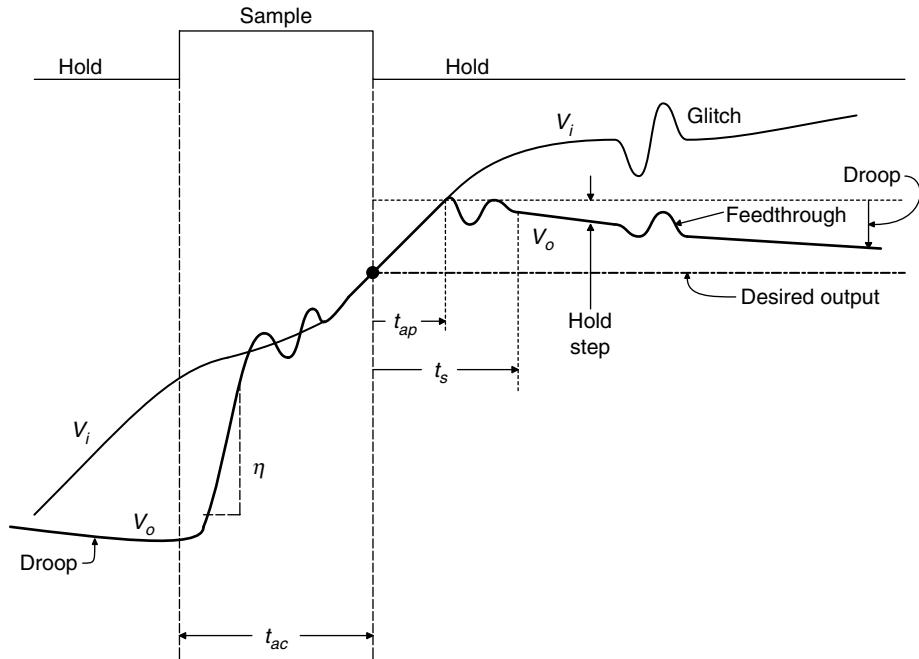
the FET switch opens, the charge on C_H is trapped there so that V_o is held at V_C . However, things are not that simple; V_C changes slowly due to charge leaking from C_H through its leakage resistance, and charge transfer to or from OA2 due to its dc bias current, I_B . This slow drift in V_C is called *droop* and its slope can be either positive or negative, depending on the sign of I_B . In addition to the droop, there is also a feed through error, caused by the fact that the FET switch does not have an infinite off resistance. When the FET switch is switched off, there is also an unwanted transfer of charge from the FET gate to C_H through the gate to drain capacitance of the FET. This charge can produce an unwanted *pedestal voltage* which is approximately given by:

$$\delta V_o = (C_{gd}/C_H)(V_{OB} - V_o) \quad (2.146)$$

where V_{OB} is the off level bias voltage on the FET gate and V_o is, of course, the value of V_o at the time the hold command is given. Thus the pedestal voltage depends on the input voltage and as numerical calculations reveal, can prove to be quite an objectionable source of T&H circuit error (Franco, Section 7.11 gives an example). A number of schemes have been devised to minimize pedestal error. One is to inject an equal and opposite switching charge into C_H by means of an auxiliary active circuit. Another is illustrated in the circuit architecture of the Burr-Brown SHC803BM T&H, shown in Figure 2.44. In this T&H system, a capacitor $C'_H = C_H$ is used to compensate for pedestal and droop at the same time, using the differential gain

**FIGURE 2.44**

Basic architecture of the Burr-Brown SHC803BM track and hold circuit. S2 and capacitor C'_H are used for cancellation of pedestal voltage.

**FIGURE 2.45**

Waveforms seen in the operation of a typical track and hold circuit. (Source: Franco, 1988)

of OA2. Capacitor leakage, the dc bias currents at OA2 and FET switch charge injection on switching cause equal voltage errors on both capacitors, which are subtracted out. The C_H is operated in the integrator mode in the Burr-Brown design. In the hold mode, both SW1 and SW2 are off, hence the charge on C_H is trapped, and V_o remains constant. In the tracking mode, it is assumed the on resistance of SW1 is $0\ \Omega$. Thus V_o follows V_1 with the transfer function:

$$V_o/V_1 = -1/(1000C_H s + 1) \quad (2.147)$$

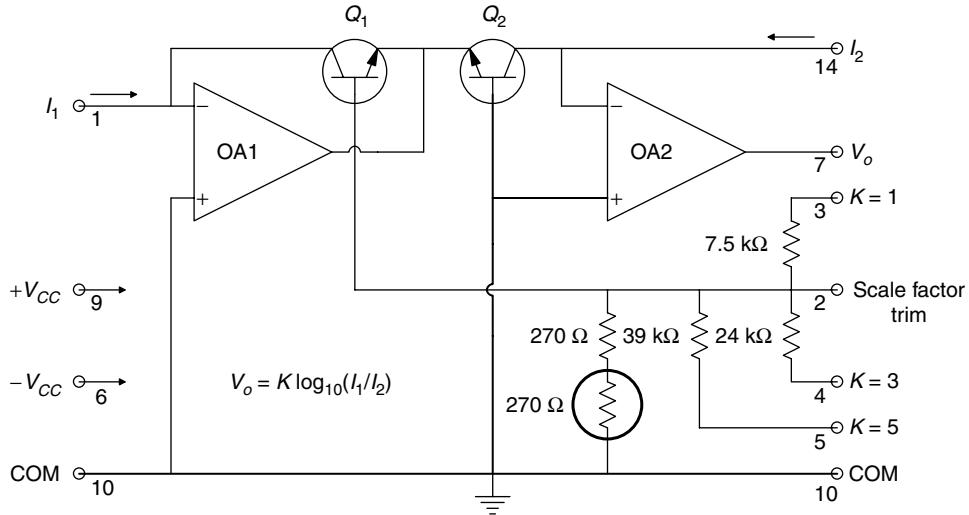
If C_H is 100 pF , then the -3 dB frequency is 1.6 MHz . In general, because C_H appears in the feedback loop of OA2, the tracking mode frequency response of the integrator type of T&H is lower than that attainable from the grounded C_H configuration shown in Figure 2.43. Figure 2.45 illustrates the T&H output errors discussed above. The droop, signal feed through due to finite FET switch off resistance, pedestal, and several important operating times specified by the manufacturers are illustrated.

2.6.7 Log Ratio and Trigonometric ICs

Another class of nonlinear analog signal processing circuit which should be considered is the log ratio IC. The analog output voltage of this IC is given by:

$$V_o = K \log_{10}(I_1/I_2) \quad (2.148)$$

Typically, K is set to some convenient integer value such as 1, 2, 3, or 5. Note that the inputs are currents which may vary over a six decade range (e.g. 1 nA – 1 mA). This nonlinear circuit module finds applications in spectrophotometry (analytical chemistry),

**FIGURE 2.46**

Schematic of the Burr-Brown LOG100 log ratio converter. (Source: Courtesy of Burr-Brown)

where parameters such as absorbance or optical density need to be calculated and displayed. Also, if I_1 and I_2 are made to be dc currents proportional to the RMS value of the sinusoidal output and input of a linear system, respectively, the log ratio converter output is proportional to the frequency response magnitude (Bode log magnitude) in dB. Of course, if I_2 is fixed, then it is a simple log converter.

The Burr-Brown LOG100 is a commercially available log ratio converter and Figure 2.46 shows the internal circuit of this IC. Analog Devices offers a similar unit, the AD757, which will accept either current or voltage inputs. Note that the log ratio modules are normally used with dc inputs. Their frequency responses are a function of the input signal amplitude, ranging from about 100 Hz for nA peak input currents to over 40 kHz for mA peak input currents.

Often, the nonlinear signal processing requirements of the system demand a broadband logarithmic nonlinearity, which can operate on video frequency signals. The Analog Devices AD640 is an example of an IC which can produce an output of the form,

$$V_o = V_y \log(V_1/V_{sc}) \quad (2.149)$$

over a 46 dB conversion range and an 80 MHz bandwidth. Here, V_{sc} is the intercept scaling voltage and V_y is the slope voltage, both constants. Two cascaded AD640s are claimed to have a 95 dB dynamic range in a 20–100 kHz bandwidth; a 70 dB range is claimed in a 50–150 MHz bandpass. However, the IC is directly coupled and will work in the range dc through the entire audio spectrum.

The Analog Devices AD639 trigonometric converter is another specialized, nonlinear analog processing module which has applications such as generation of a sine wave from a triangle wave at frequencies up to 1.5 MHz, coordinate conversion and vector resolution, design of quadrature and variable phase oscillators, and imaging and scanning linearization circuits. The nonlinear transfer function of the AD639 is:

$$V_o = U \frac{\sin(V_{x1} - V_{x2})}{\sin(V_{y1} - V_{y2})} \quad (2.150)$$

The scaling factor for the differential voltage inputs is $50^\circ/\text{volt}$. The IC has on board a 1.8 V dc bandgap voltage reference, which can be used to preset either angle argument to 90° . The output of the AD639 is scaled by the constant, K , which can be set between zero and +15 V, by either external or internal means. A wide variety of trigonometric functions can be modeled, such as $\sin x$, $\cos x$, $\tan x$, $\operatorname{cosec} x$, $\sec x$ and $\cotan x$, as well as certain useful inverse trigonometric functions. For the arctangent:

$$V_o = K\theta = K \tan^{-1} \frac{(V_{z2} - V_{z1})}{(V_{u1} - V_{u2})} \quad (2.151)$$

where $K = 20 \text{ mV}/^\circ$. With minor changes in the external circuit architecture, the AD639 can also calculate the arcsine and arctangent functions with the same ratio of voltage differences argument as the above arctangent function. With the exception of sine wave generation, it would appear that this versatile IC is best used with inputs which are dc, or at best, in the low audio frequency range.

2.6.8 Summary

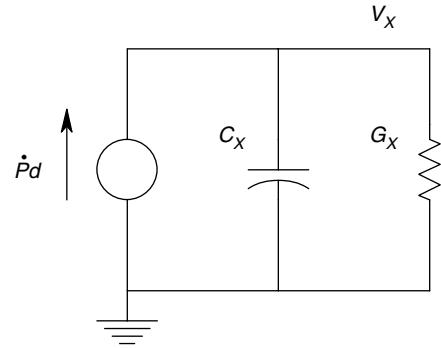
We have begun this section by discussing some of the properties of linear and nonlinear analog systems. There are many uses for nonlinear analog signal processing ICs. The present trend is to sample and digitize the input signal, use a computer to calculate the nonlinear operation, then realize the result in analog form by digital to analog conversion. However, there are instances where the instrumentation system will be less expensive if an all-analog approach is taken and a specialized analog function module is used. Computers, while versatile, can be expensive. Obviously, there is a trade-off between accuracy, cost and convenience in choosing between an all-analog *vs* an A/D to computer to D/A type of system for nonlinear signal processing in an instrumentation system. The designer should be aware that all-analog signal processing ICs are available which may compete with the digital approach in terms of function, accuracy and cost.

2.7 Charge Amplifiers

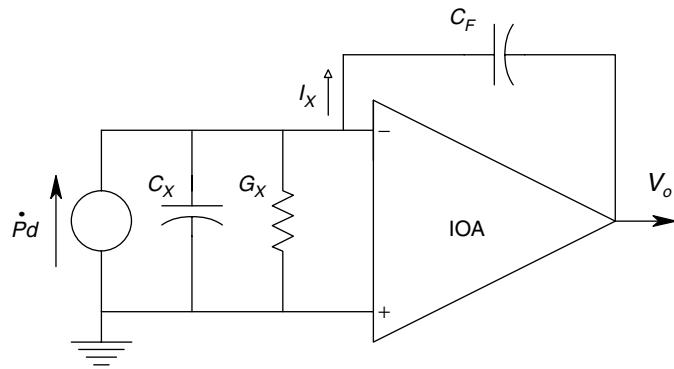
Charge amplifiers make use of electrometer (FET input) op-amps, which have ultra-low dc bias currents at their inputs and ultra-high input resistances. They find primary application in isolating and conditioning the outputs of piezoelectric transducers used to measure pressure and force transients. They are also used to measure the coulombic charge stored in capacitors and indirectly, the electric field strength.

2.7.1 Charge Amplifiers Used with Piezoelectric Transducers

In Chapter 6, we show that the equivalent circuit of a piezoelectric pressure transducer operated well below its mechanical resonant frequency is given by the simple parallel current source, conductance and capacitance, shown in Figure 2.47. Note that d is a constant which depends on the piezoelectric material and how it is cut relative to the crystal axes to form the transducer and that in general, current is the rate of transfer of charge through the plane through which it is passing. By U.S. convention, current is conventionally taken as positive in the direction of motion of positive charges, such

**FIGURE 2.47**

Equivalent circuit of a piezoelectric crystal responding to pressure on its active surface at frequencies well below its mechanical resonance frequency.

**FIGURE 2.48**

A charge amplifier circuit used to condition the output of a piezoelectric crystal sensor.

as holes in semiconductors, or positive ions in solutions or plasma discharges. Hence in a conductor, such as copper, where charge is carried by electrons, the current direction is opposite to the direction of electron motion.

Figure 2.48 illustrates an ideal op-amp used as a charge amplifier. The op-amp is assumed to have infinite gain, frequency response and input resistance, and zero bias current, offset voltage, noise and output resistance. Under these ideal op-amp assumptions, it is easy to see that because the summing junction is at zero volts, all of the current, \dot{P}_d , flows into the feedback capacitor, C_F . Hence V_o is proportional to $\dot{P}(t)$, and is given by:

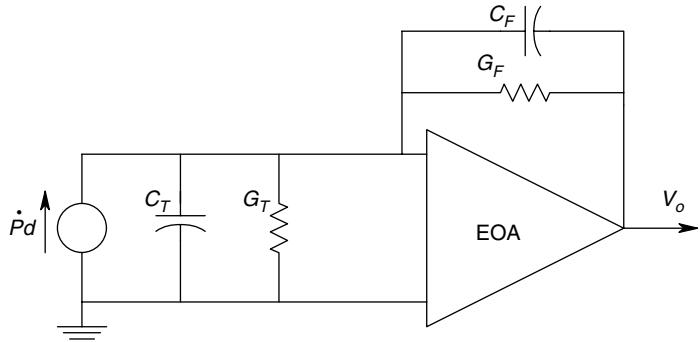
$$V_o(t) = -\dot{P}(t) d / C_F \quad (2.152)$$

In actual practice, this result of the ideal case is seen to become less simple and it should be noted that the electrometer op-amp has a non-infinite gain approximated by:

$$K_V = \frac{K_{V_o}}{(\tau s + 1)} (V_i - V'_i) \quad (2.153)$$

Also, we assume that there is a small but finite conductance, G_F , in parallel with C_F . The G_F is of the order of $10^{-10} S$, or smaller. Now, if we write the node equation at the summing junction and substitute the gain expression above to eliminate V'_i , we can show that the charge amplifier transfer function is given by (approximately):

$$V_o/P = \frac{-s(d/G_F)}{s^2[\tau(C_T + C_F)/K_{V_o}G_F] + s[\tau(G_T + G_F) + K_{V_o}C_F]/K_{V_o}G_F + 1} \quad (2.154)$$

**FIGURE 2.49**

Circuit of a practical charge amplifier and crystal sensor.

This is the transfer function of a bandpass system. The low frequency pole is approximately at:

$$f_{\text{low}} \approx 1/(2\pi C_F R_F) \text{ Hz} \quad (2.155)$$

The mid-band gain is:

$$A_{V\text{mid}} = -K_{V_0} d / [\tau(G_T + G_F) + K_{V_0} C_F] \cong -d/C_F \quad (2.156)$$

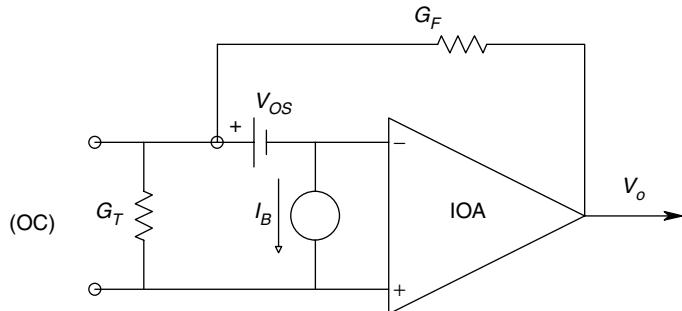
The high frequency pole is at:

$$f_{\text{hi}} = K_{V_0} C_F / [2\pi\tau(C_T + C_F)] \text{ Hz} \quad (2.157)$$

Note that $K_{V_0}/(2\pi\tau)$ is simply the small signal GBWP of the electrometer op-amp. Figure 2.49 shows the circuit of the non-ideal charge amplifier. Here, C_T is the total capacitance to ground which includes the capacitances of the piezoelectric crystal, the cable connecting the transducer to the charge amp, and the charge amp input. Typically, C_T might be about 400 pF. The G_T is the total shunt conductance to ground at the op-amp summing junction. The G_T includes the leakage conductances of the crystal, the cable, and the op-amp input. Typically, G_T is about 10^{-13} S. If, in addition, $C_F = 2 \times 10^{-9}$ F, $G_F = 10^{-10}$ S, $K_{V_0} = 5 \times 10^4$, $d = 2 \times 10^{-12}$ amp.sec/psi for quartz, and the op-amp GBWP = 350 kHz, the system ends up having a mid-band gain of 1 mV/psi, a high frequency pole at 1.5×10^5 Hz, and a low frequency pole at 8×10^{-3} Hz. Note that the low frequency pole is set by C_F/G_F , and is independent of the input circuit. Thus, the piezoelectric transducer charge amplifier is not a system which can respond to static (dc) pressures; it must be used to measure transient pressure changes. In addition to the low frequency limitations imposed on the transfer function by non-ideal electrometer op-amp behavior, there are also dc errors at the output due to the bias current and dc offset voltage of the op-amp. Figure 2.50 shows the dc equivalent circuit of the charge amplifier. Note that the capacitors drop out from the circuit at dc because they carry no current. It is easy to see that the dc error in the output is:

$$V_{o\text{DC}} = I_B R_F + V_{OS}(1 + G_T/G_F) \quad (2.158)$$

Note that I_B and V_{OS} can have either a plus or a minus sign. Typical values of V_{OS} are about ± 200 μ V, and I_B values are around 75 fA (75×10^{-15} A).

**FIGURE 2.50**

The dc equivalent circuit of a practical charge amplifier.

2.7.2 Charge Amplifier as Integrating Coulombmeter

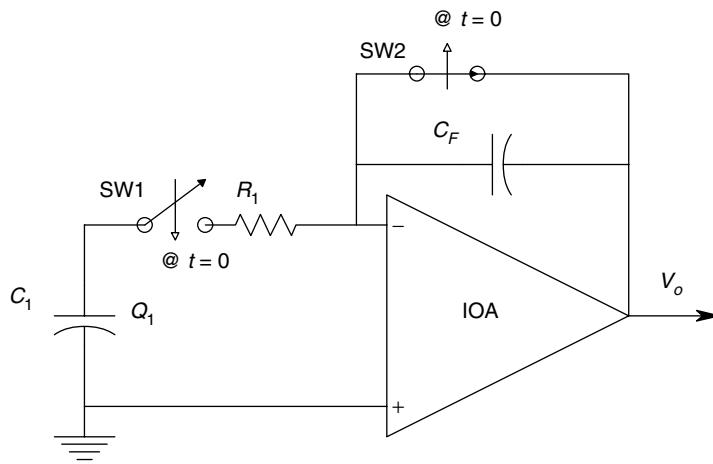
A charge of Q_1 coulombs is trapped on a capacitor, C_1 . The capacitor can be a real capacitor, or the capacitance of a charged object, such as the fuselage of a helicopter, or a human body. We wish to measure Q_1 . The voltage across C_1 is given by the well-known relation, $V_1 = Q_1/C_1$. In Figure 2.51, C_1 is connected to the summing junction of an electrometer op-amp. At the same time, the short is removed from the feedback capacitor, C_F , allowing it to charge from the op-amp output to a value which maintains the summing junction at 0 V. Assuming the op-amp to be ideal, the output voltage of the op-amp, after three or four R_1C_1 time constants, is:

$$V_{o(SS)} = -Q_1/C_F \text{ dc V} \quad (2.159)$$

Hence, the system can measure the charge on a small capacitor, provided the peak current, $i_{1PK} = Q_1/R_1C_1$, does not exceed the current sourcing or sinking capability of the op-amp.

In practice, this circuit is subject to dc errors caused by the bias current and offset voltage of a practical op-amp. The output error caused by I_B and V_{OS} can be found by superposition and is given by:

$$\delta V_o = T(I_B/C_F) + V_{OS}(1 + C_1/C_F) \text{ V} \quad (2.160)$$

**FIGURE 2.51**

An integrating coulombmeter circuit. An electrometer op-amp is generally used.

where T is the time elapsed between opening SW1 and closing SW2, and reading the steady state, V_o . In practice, δV_o can be made negligible by the judicious choice of circuit parameters.

2.7.3 Summary

In this section, we have presented two examples of applications of the op-amp charge amplifier configuration. The basic charge amplifier consists of an ideal op-amp, with an ideal feedback capacitor to integrate charge transferred to the summing junction. In practice, we use electrometer amplifiers which have ultra-high input resistances and finite gains, bandwidths, offset voltages and bias currents. Since a pure C_F will integrate the bias current of the summing junction, a large resistance, R_F , must be put in parallel with C_F to give a fixed dc error, which is proportional to I_B . The parallel combination of R_F and C_F sets the low frequency pole of a bandpass frequency response for the piezoelectric transducer. In the mid-band range of frequencies, the output voltage is proportional to the pressure on the transducer, rather than its derivative.

2.8 Phase Sensitive Rectifiers

Phase sensitive rectifiers (PSRs), also called *phase sensitive demodulators* or *synchronous rectifiers*, are an important class of analog circuit, widely used in instrumentation, control and communications systems. The PSR is used to convert the output of certain ac, carrier operated sensors to a low frequency signal proportional to the quantity under measurement (QUM). One example of the use of PSRs is to condition the output of a resistive Wheatstone bridge, using strain gauges to measure mechanical strain, pressure or force. The strain gauge bridge excitation is generally an audio frequency sine wave. The bridge output is a sine wave of the same frequency, with amplitude proportional to the strain, etc. The phase of the bridge output is 0° with respect to the excitation phase if $\Delta X > 0$ and is -180° with respect to the excitation if $\Delta X < 0$. The PSR uses the bridge excitation voltage to sense this phase change and to produce a rectified signal which has amplitude proportional to ΔX , and the sign of ΔX .

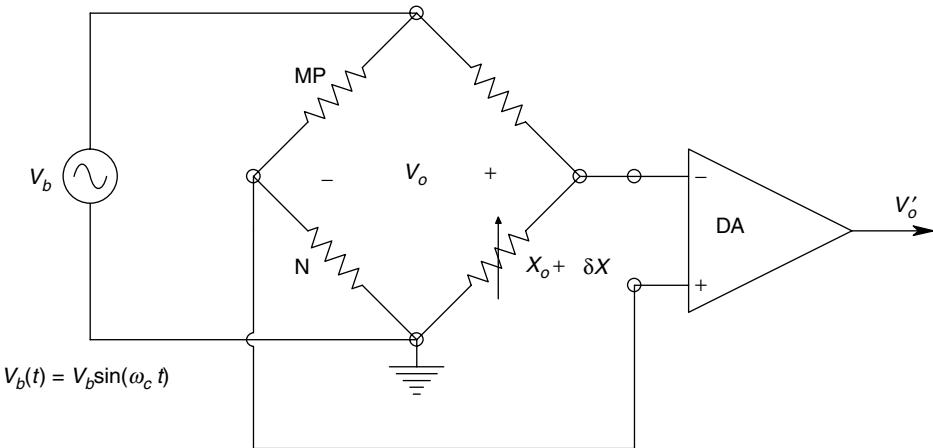
2.8.1 Double-Sideband, Suppressed Carrier Modulation

The outputs of many sensor systems have information encoded on them in what is called double-sideband, suppressed carrier modulation (DSBSCM). A DSBSCM signal is formed by multiplying the ac carrier by the modulating signal. To illustrate this process, we examine a Wheatstone bridge with one variable element and AC excitation, as shown in Figure 2.52. It is easy to show (Chapter 4) that the bridge output is given by:

$$V'_o(t) = K_D[V_b \sin(\omega_c t)] \left[\frac{X}{X + P} - \frac{N}{M + N} \right] \quad (2.161)$$

Now, the resistance X is allowed to vary with time:

$$X(t) = X_o + \Delta X(t) \quad (2.162)$$

**FIGURE 2.52**

A Wheatstone bridge circuit with ac excitation generates a double sideband, suppressed carrier modulated output.

Furthermore, the bridge is balanced to make it maximally sensitive. This requires that:

$$M = N = P = X_o \quad (2.163)$$

Thus, around null,

$$V'_o(t) = \Delta X(t) K_D V_b (1/4X_o) \sin(\omega_c t) \quad (2.164)$$

Note that we obtain the product of $\Delta X(t)$ and the excitation sinusoid at the bridge output. If $\Delta X(t)$ is sinusoidal, with $\omega_m \ll \omega_c$, using the well-known trigonometric identity, $[\sin(\alpha) \sin(\beta)] = [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2$ and

$$V'_o(t) = K_D V_b [\Delta X / 8X_o] [\cos((\omega_c - \omega_m)t) - \cos((\omega_c + \omega_m)t)] \quad (2.165)$$

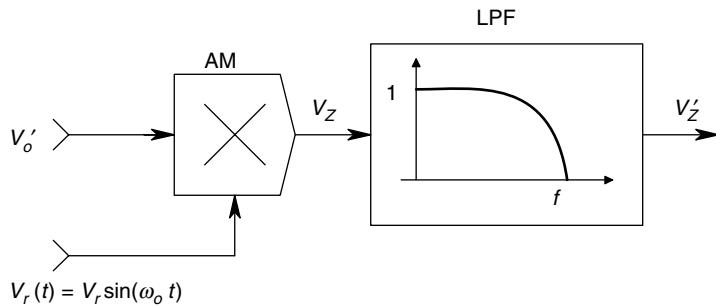
Thus, from equation 2.165, we see that the conditioned bridge output, $V'_o(t)$, consists of two sidebands and no carrier frequency term and hence the designation DSBSCM.

Another example of a transducer which produces a DSBSCM output is the linear variable differential transformer (LVDT), a device which is used to sense small, linear mechanical displacements. Some LVDTs come with a built-in PSR and provide a dc output directly proportional to core displacement, while others must be used with an external PSR (LVDTs are discussed in detail in Chapter 6).

2.8.2 Demodulation of DSBSCM Signals by Analog Multiplier

Figure 2.53 illustrates one of the simplest ways to demodulate a DSBSCM signal and recover $\Delta X(t)$. An analog multiplier IC is used to multiply a reference signal which is in phase with the transducer ac excitation by the transducer DSBSCM output. This product can be written as:

$$V_z(t) = \frac{\Delta X K_D V_b V_r}{10 \times 8X_o} \{ \sin(\omega_c t) \cos[(\omega_c - \omega_m)t] - \sin(\omega_c t) \cos[(\omega_c + \omega_m)t] \} \quad (2.166)$$

**FIGURE 2.53**

An analog multiplier and low-pass filter used to demodulate a DSBSC signal.

The bracketed term can again be expanded by a trigonometric identity as:

$$V_z(t) = \frac{\Delta X K_D V_b V_r}{(10) 16 X_o} [\sin((2\omega_c - \omega_m)t) + \sin(\omega_m t) - \sin((2\omega_c + \omega_m)t) - \sin(-\omega_m t)] \quad (2.167)$$

$V_z(t)$ is now passed through a unity gain low-pass filter which removes the two high frequency terms. The output of the LPF is the desired, demodulated signal, multiplied by a scale factor:

$$V'_z(t) = \frac{K_D V_b V_r}{80 X_o} [\Delta X \sin(\omega_m t)] \quad (2.168)$$

2.8.3 Other PSR Designs

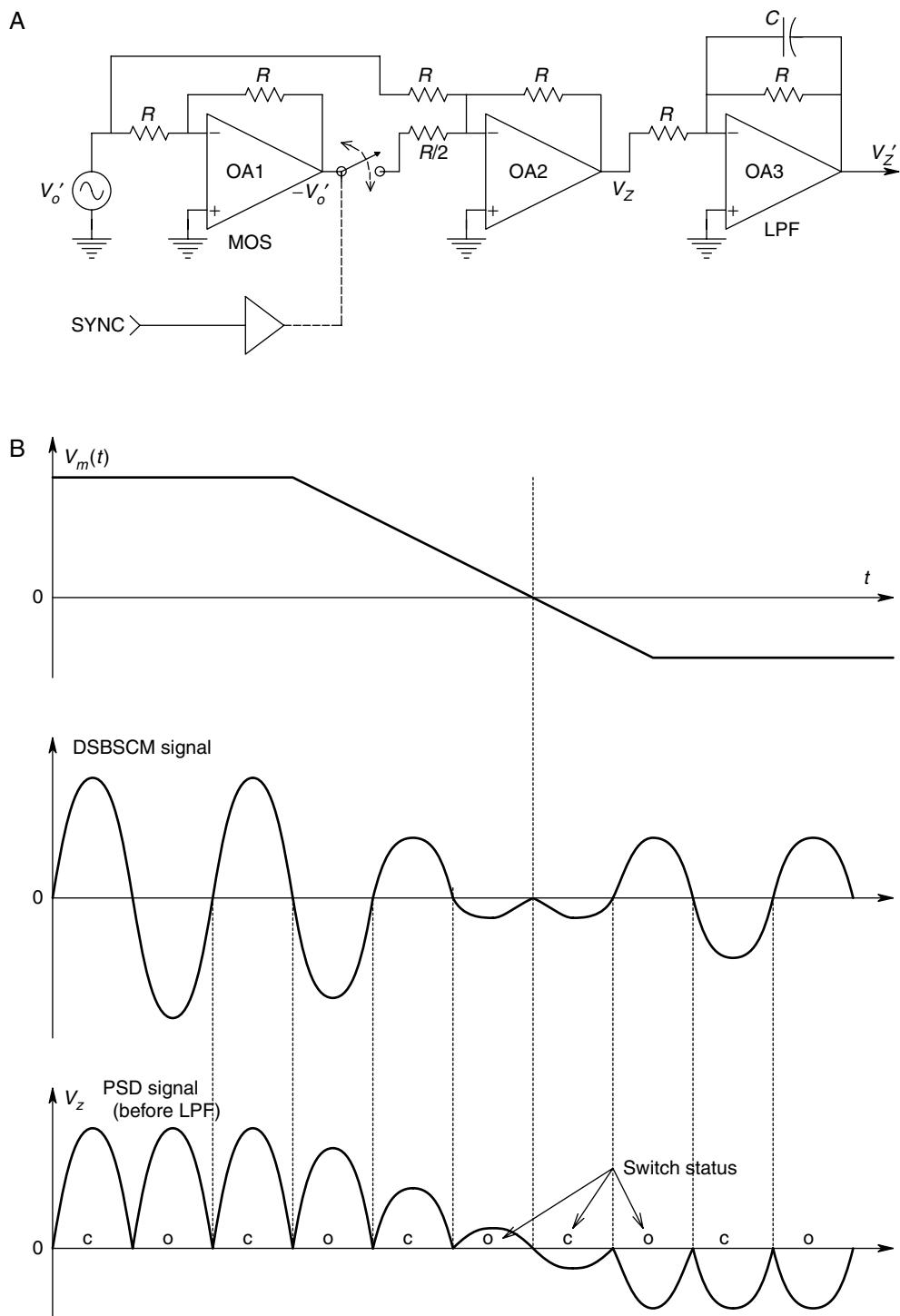
Figure 2.54A illustrates a simple PSR which can be made from two op-amps and a bidirectional MOS switch operated by a logical square wave. The wave is derived from the sinusoidal excitation source by first conditioning it with a comparator used as a zero-crossing detector and then, passing the TTL waveform through two one-shot multi-vibrators (e.g. a 74LS123) to adjust the reference signal phase. Since the MOS switch closes at the same time in the reference signal cycle as when the DSBSMC signal changes phase by 180° , the rectifier output changes sign, producing a negative output. A low-pass filter is used to smooth the full wave rectified output of the switching circuit. The relevant waveforms are shown in Figure 2.54B.

Phase sensitive rectifier ICs are available from a variety of manufacturers and include the Analog Devices AD630 balanced modulator/demodulator, the Precision Monolithics GAP-01 general purpose analog signal processing subsystem, the National Semiconductor LM1596 balanced modulator-demodulator and the Signetics NE/SE5520 LVDT signal conditioner, to mention a few.

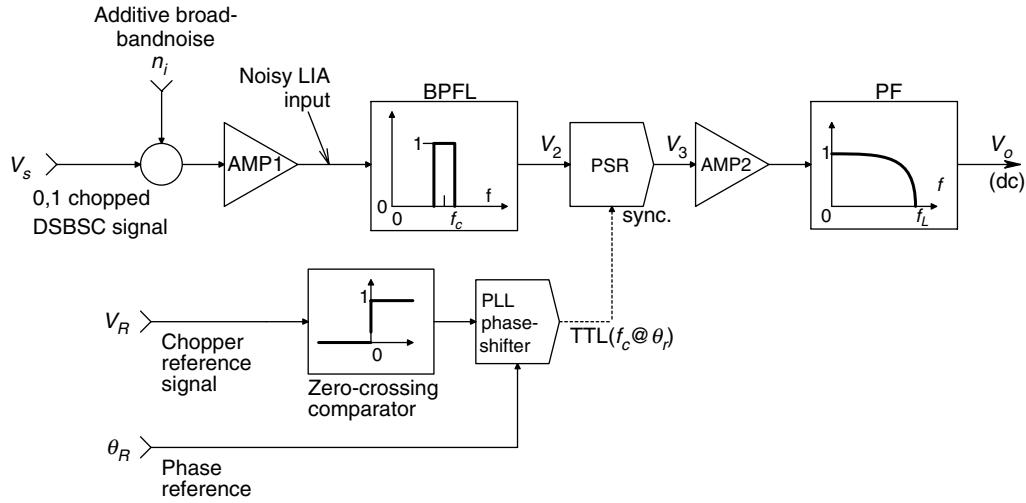
2.8.4 The Lock-In Amplifier

2.8.4.1 Introduction

The lock-in amplifier (LIA) can be used to condition and demodulate DSBSMC signals and recover the modulating signal under conditions of abominable SNRs, where the RMS noise voltage can exceed that of the modulated carrier by over 60 dB. Noise is reduced in the operation of the LIA by three mechanisms—bandpass filtering around the carrier frequency, rejection of coherent interference at frequencies other than that of

**FIGURE 2.54**

A. A simple phase sensitive rectifier and low-pass filter circuit using op-amps and a MOS switch. B. Waveforms in the PSR/LPF circuit.

**FIGURE 2.55**

Block diagram of a lock-in amplifier. Input is a 0,1 chopped, noisy signal, such as the output of a photosensor. Amplification and bandpass filtering at the chopper frequency converts the input signal into a noisy ac signal, which is then rectified by the PSR. The low frequency modulating signal is seen at the LIA output. The lowpass filter bandwidth is adjusted to pass the desired modulating signal with the least amount of noise.

the carrier and low-pass filtering at the PSR output. Carrier frequencies are best chosen to lie in the range 10–40 Hz, and 200–10⁵ Hz. Carrier frequencies lying close to the 60 Hz power line frequency and its first two harmonics are avoided, because they may ‘leak’ through the LPF following the PSR. Very low carrier frequencies are also avoided as they may place LIA operation in the range of 1/f noise from the analog electronic circuits.

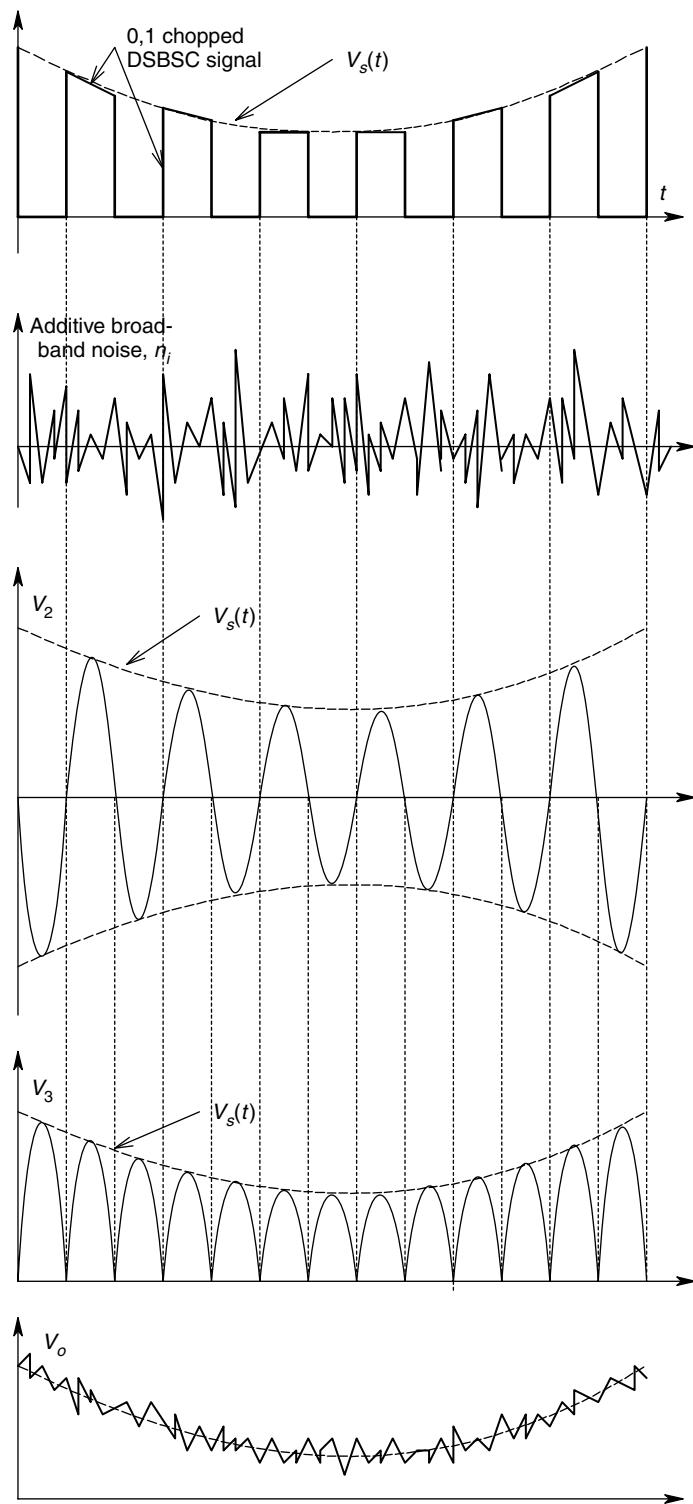
A block diagram of an LIA is shown in Figure 2.55, along with some pertinent LIA waveforms in Figure 2.56. We see that the LIA performs four basic operations on the DSB modulated carrier with additive, broadband noise:

1. The incoming DSBSCM signal plus noise is generally conditioned by a bandpass amplifier whose centre frequency is the same as that of the carrier.
2. The bandpass signal conditioning improves the SNR of the modulated carrier.
3. The signal next passes through a phase sensitive rectifier, usually of the switching type shown in Figure 2.54A.
4. The PSR output is next conditioned by a dc pass low-pass filter, which generally has an adjustable cut-off frequency and attenuation slope. Some dc post amplification by a low noise, low drift amplifier may precede the low-pass filter, if required.

A reference signal directly derived from the carrier is used to operate the PSR.

For optimum performance of the PSR, the phase of the reference signal must be adjusted to be exactly in phase with the carrier in the DSBSCM input. To illustrate this principle of LIA operation in the noise free case, we let the modulating signal be a dc voltage:

$$V_m(t) = V_m \quad (2.169)$$

**FIGURE 2.56**

Waveforms in the LIA of Figure 2.55.

The carrier and reference signal are sinusoids, respectively:

$$V_c(t) = V_c \sin(\omega_c t + \theta_c) \quad (2.170)$$

$$V_r(t) = V_r \sin(\omega_r t + \theta_r) \quad (2.171)$$

The DSBCSM signal is:

$$V_{mc}(t) = V_m V_c \sin(\omega_c t + \theta_c) \quad (2.172)$$

Demodulation, as we have shown above, requires the multiplication of the modulated carrier with the reference signal. A scale factor K_D is introduced. Thus the demodulated signal before low-pass filtering can be written as:

$$V_d(t) = K_D V_m V_c V_r \sin(\omega_c t + \theta_c) \sin(\omega_r t + \theta_r) \quad (2.173)$$

After the $\sin(\alpha)\sin(\beta)$ trig. identity is applied, we have:

$$V_d(t) = (K_D V_m V_c V_r / 2) [\cos(\theta_c - \theta_r) - \cos(2\omega_c t + \theta_c + \theta_r)] \quad (2.174)$$

After low-pass filtering, the dc output of the LIA can be written as:

$$V_o = V_m [K_D K_F V_c V_r / 2] \cos(\theta_c - \theta_r) \quad (2.175)$$

where K_F is the dc gain of the LPF. Note that the output is maximum when the phase of the reference signal and the carrier are equal (i.e. there is zero phase shift between the carrier and reference signal). Also, for the LIA to remain in calibration, K_D , K_F , V_c and V_r must remain constant and known. Most modern LIAs have adjustable phase shifters in the reference channel to permit output sensitivity maximization, while others perform this task automatically using phase lock circuitry [McDonald and Northrop, 1993].

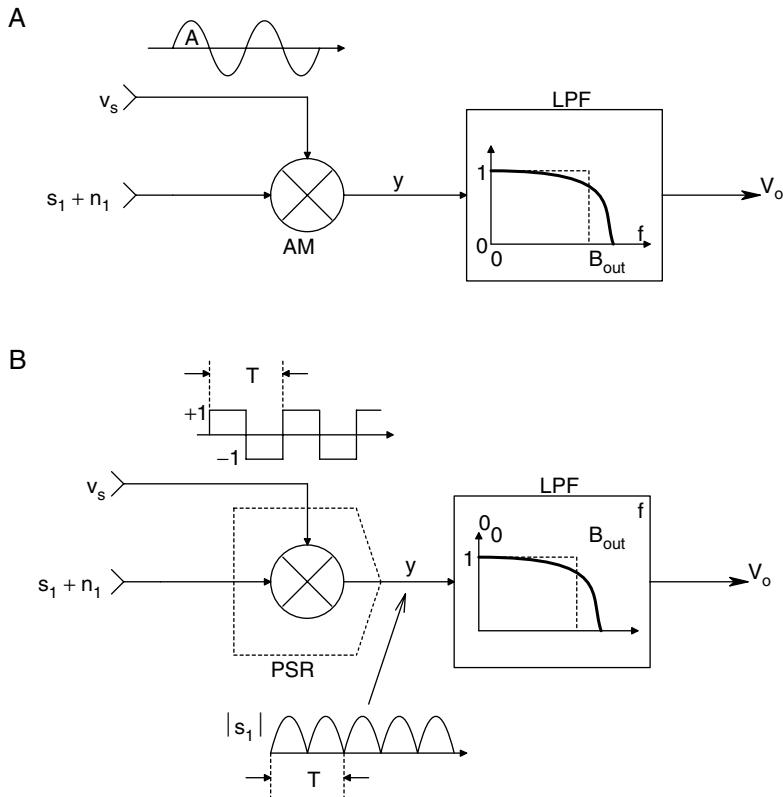
2.8.4.2 Calculation of the SNR Improvement Using a Lock-In Amplifier

The major application of an LIA is to measure coherent, DSBSC signals, generally at audio frequencies, that are 'buried in noise'. More precisely, the input SNR is extremely low, often as low as -60 dB. There are two basic architectures used for LIAs. One uses an analog multiplier for signal detection followed by a low-pass filter. This system is shown in Figure 2.57A. A second LIA architecture, shown in Figure 2.57B, uses a synchronous rectifier followed by a low-pass filter to measure the amplitude of a coherent signal of constant frequency which is present in an overwhelming amount of noise. Assuming that the input to the LIA is

$$x(t) = n_i(t) + V_s \cos(\omega_o t) \quad (2.176)$$

Assuming that the noise is zero mean, the Gaussian white noise with one-sided power density spectrum, $S_n(f) = \eta \text{ mV/Hz}$. The ms input SNR is:

$$\text{SNR}_{\text{in}} = \frac{V_s^2 / 2}{\eta B_{\text{in}}} \quad (2.177)$$

**FIGURE 2.57**

Two means of demodulation of DSBSC signals used in LIAs. A. Demodulation by analog multiplier and LPF. B. Demodulation by synchronous rectifier and LPF.

where B_{in} is the Hz bandwidth over which the input noise is measured.

In the analog multiplier form of the LIA, $x(t)$ is multiplied by an in phase sync signal, $v_s(t) = A \cos(\omega_o t)$. The multiplier output is thus:

$$y(t) = v_s(n_i + s_i) = A \cos(\omega_o t)[n_i(t) + V_s \cos(\omega_o t)] = An_i(t) \cos(\omega_o t) + AV_s \cos^2(\omega_o t) \quad (2.178)$$

$$y(t) = An_i(t) \cos(\omega_o t) + (AV_s/2)[1 + \cos(2\omega_o t)]$$

Note that the one-sided power spectrum of the first term in equation 2.178 is the noise PDS, shifted up to be centred at $\omega = \omega_o$. Since $S_n(f)$ is white, shifting the center of $S_n(f)$ from 0 to $f_o = \omega_o/2\pi$ does not change the noise output of the LIA output bandpass filter. Thus the ms output noise voltage from the LIA is $v_{on}^2 = A^2 \eta(1/4\tau) \text{ mV}^2$, where $(1/4\tau)$ can be shown to be the equivalent noise Hz bandwidth of the simple output R-C low-pass filter, and τ is the time constant of the output LPF. The output LPF removes the $2\omega_o$ frequency term in the RHS term of equation 2.178, leaving a dc output proportional to the peak value of the sinusoidal signal. Clearly, $v_{os} = AV_s/2$. Thus the ms SNR at the LIA output is:

$$\text{SNR}_{\text{out}} = \frac{A^2 V_s^2 / 4}{A^2 \eta(1/4\tau)} = V_s^2 \tau / n \quad (2.179)$$

The noise figure, F , is defined as the ratio of SNR_{in} to SNR_{out} . For an amplifier, it is desirable for F to be as small as possible. In a noise free amplifier (the limiting case), $F = 1$. It is easy to show that F for this kind of LIA is:

$$F_{\text{LIA}} = B_{\text{out}}/B_{\text{in}} \quad (2.180)$$

where $B_{\text{out}} = 1/4\tau$ is the Hz noise bandwidth of the LIA output LPF, and B_{in} is the Hz noise bandwidth over which the input noise is measured. Since the desired output from the LIA is a dc signal proportional to the peak sinusoidal input signal, B_{out} can be of the order of a single Hz, and F_{LIA} can be $\ll 1$.

In the second model for LIA noise, consider the synchronous rectifier LIA architecture of Figure 2.57B. Now, the sinusoidal signal is full wave rectified by passing it through an absval function which models the synchronous rectifier. The full wave rectified, sinusoidal input signal can be shown to have the Fourier series [Aseltine, 1958] as shown below:

$$y(t) = |V_s(t)| = V_s(2/\pi)[1 - (2/3)\cos(2\omega_0 t) - (2/15)\cos(4\omega_0 t) - \dots] \quad (2.181)$$

Clearly, the average or dc value of y is $V_s(2/\pi)$. Curiously, the zero-mean white noise is not conditioned by the absval function of the synchronous rectifier. This is because the synchronous rectifier is really created by effectively multiplying $x(t) = n_i t + V_s \cos(\omega_0 t)$ by an even ± 1 square wave, having the same frequency as and in phase with, $V_s \cos(\omega_0 t)$. This square wave can be represented by the Fourier series:

$$\text{SQWV}(t) = \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) \quad (2.182)$$

where $\omega_0 = 2\pi/T$

$$a_n = (2/T) \int_{-T/2}^{T/2} f(t) \cos(n\omega_0 t) dt = (4/n\pi) \sin(n\pi/2) \quad (\text{non-zero for } n \text{ odd}) \quad (2.183)$$

for a unit cosine (even) square wave which can be written in expanded form:

$$\text{SQWV}(t) = (4/\pi)[\cos(\omega_0 t) - (1/3)\cos(3\omega_0 t) + (1/5)\cos(5\omega_0 t) - (1/7)\cos(7\omega_0 t) + \dots] \quad (2.184)$$

The input Gaussian noise is multiplied by $\text{SQWV}(t)$:

$$y_n(t) = n(t) \times (4/\pi)[\cos(\omega_0 t) - (1/3)\cos(3\omega_0 t) + (1/5)\cos(5\omega_0 t) - (1/7)\cos(7\omega_0 t) + \dots] \quad (2.185)$$

The white, one-sided noise power density spectrum, before passing through the LIA output LPF, can be shown to be:

$$S_n(f) = \eta(4/\pi)^2[1 - (1/3) + (1/5) - (1/7) + (1/9) - (1/11) + (1/13) - \dots]^2 = \eta \text{ msV/Hz} \quad (2.186)$$

This white noise is conditioned by the output LPF, so the output ms SNR is:

$$\text{SNR}_{\text{out}} = \frac{V_s^2(2/\pi)^2}{\eta B_{\text{out}}} \quad (2.187)$$

The ms SNR at the input to the LIA is:

$$\text{SNR}_{\text{in}} = \frac{V_s^2/2}{\eta B_{\text{in}}} \quad (2.188)$$

Thus, F for this LIA architecture is found to be:

$$F = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} = \frac{\pi}{2^3} \frac{B_{\text{out}}}{B_{\text{in}}} = 1.23 \frac{B_{\text{out}}}{B_{\text{in}}} \quad (2.189)$$

which can be made quite small by adjusting the ratio of B_{out} to B_{in} .

As an example of signal recovery by an LIA, consider an LIA of the analog multiplier architecture. Let the signal to be measured be $V_s(t) = 20 \times 10^{-9} \cos(2\pi 50 \times 10^3 t)$. The Gaussian white noise present with the signal has a root power spectrum of $\sqrt{\eta} = 4n \text{ V RMS}/\sqrt{\text{Hz}}$. The input amplifier has a voltage gain of $K_v = 10^4$, with a 100 kHz noise and signal bandwidth. Thus the raw input ms SNR is:

$$\text{SNR}_{\text{in1}} = \frac{(20 \times 10^{-9})^2/2}{(4 \times 10^{-9})^2 \times 10^5} = \frac{2 \times 10^{-16} \text{ msV}}{1.6 \times 10^{-12} \text{ msV}} = 1.25 \times 10^{-4}, \text{ or } -39 \text{ dB} \quad (2.190)$$

if the sinusoidal signal is conditioned by an (ideal) unity gain LPF with a $Q = 50 = \text{center frequency}/\text{bandwidth}$, the noise bandwidth $B = 50 \times 10^3 / 50 = 1 \text{ kHz}$. The ms SNR at the output of the ideal BPF is still poor, given by:

$$\text{SNR}_{\text{filt}} = \frac{2 \times 10^{-16} \text{ msV}}{(4 \times 10^{-9})^2 \times 10^3} = 1.25 \times 10^{-2}, \text{ or } -19 \text{ dB} \quad (2.191)$$

Now, the signal plus noise are amplified and passed directly into the LIA and the LPF output of the LIA has a noise bandwidth of 0.125 Hz. The output dc component of the signal is $K_V V_s / 2 \text{ V}$.

Thus the ms dc output signal is

$$\overline{V_{so}^2} = K_V^2 V_s^2 / 4 = K_V^2 \times 1 \times 10^{-16} \text{ msV}$$

The ms noise output is

$$V_{no}^2 = K_V^2 (4 \times 10^{-9})^2 \times 0.125 \text{ msV}$$

Thus, the ms $\text{SNR}_{\text{out}} = 50$, or +17 dB.

The cost of using an LIA is that the ac signal is reduced to a proportional dc level and given a change in V_s , the output LPF is such that the LIA does not reach a new steady state output, until about 3 time constants of the LPF (about 3 seconds in the example above) have occurred. The benefit of using an LIA is that coherent signals, buried in up to 60 dB noise, can be measured.

Many applications of LIAs lie in the area of electro-optical or photonics instrumentation. In these systems, the DSBSCM signal is derived by optically chopping a light beam. The chopper wheel generates a square wave, 0,1 carrier, rather than a sinusoidal carrier. An LED and photodiode on the chopper can generate the reference signal, also a square wave. The PSR of the LIA demodulates the square wave DSBSCM signals by the same means as that used with a sinusoidal carrier.

Commercially available LIAs are multipurpose instruments and as such, are expensive. For a dedicated LIA application, one can assemble the basic LIA system components from low noise op-amps and a commercial PSR IC, such as the AD630 or LM1596. The reference signal is converted to a TTL wave by a zero-crossing comparator. The TTL reference phase can be adjusted independent of carrier frequency by using a phase lock loop as a phase shifter [Northrop, 1990]. Such a dedicated LIA can be built for less than about \$150 US, exclusive of power supplies.

2.8.5 Summary

The phase sensitive rectifier is a circuit that is widely used in instrumentation systems to demodulate double sideband, suppressed carrier signals. We have presented a description of the DSBSCM signal and how it is created and have shown two simple circuits that can be used to demodulate it. We have also mentioned a few of the commercially available ICs suitable for PSR use.

Probably the most important instrument based on the PSR is the lock-in amplifier, which can be used to recover the signals from DSBSCM carriers that are severely contaminated with noise, sometimes noise even in the signal passband. We algebraically illustrated the LIA demodulation process, as well as the subsystems of an LIA.

2.9 Chapter Summary

Chapter 2 dealt with the analysis and design of analog signal conditioning circuits. Our treatment was subdivided to include both linear and nonlinear signal processing systems. Integrated circuit designs were stressed and examples of commercially available ICs given. The Chapter was opened in Section 2.2 with a basic treatment of differential amplifiers which are widely used in the front ends of op-amps and instrumentation amplifiers to reduce common-mode interference.

Section 2.3 dealt with the ubiquitous operational amplifier which finds many applications in linear and nonlinear signal processing. Basic design types of op-amps were reviewed and examples of their applications in linear signal processing discussed. In Section 2.4, we reviewed the basic quadratic analog active filter designs using conventional op-amps. Although a number of active filter designs exist, we have focused on three of the more popular and easy to use designs, namely the controlled source active filters (including the well-known Sallen and Key designs), the biquad active filters and the GIC based active filter designs.

The design, characteristics and uses of instrumentation amplifiers, including isolation amplifiers and autozero amplifiers were described in Section 2.5.

Section 2.6 dealt with nonlinear analog signal processing. In it we cite examples of commercially available function modules, and nonlinear circuits which can be built with op-amps. Precision absval circuits, multifunction converters, true RMS converters,

square root and divider circuits (other than multifunction converters), peak detectors, track and hold circuits, log ratio circuits and a generalized trigonometric system were discussed.

Two examples of the use of the op-amp charge amplifier configuration were given in Section 2.7—conditioning the output of piezoelectric crystal pressure transducers and measuring charge on a capacitance by integration.

Section 2.8 on phase sensitive rectifiers included the demodulation of double-sideband, suppressed carrier signals generated by such systems as Wheatstone bridges with ac excitation and LVDTs. This section also introduced the lock-in amplifier and its uses.

Problems for the reader to solve, given at the end of this chapter illustrate the design of amplifiers, filters and signal conditioning systems. Use should be made of computer aided circuit analysis in reaching the solutions to these problems.

Problems

2.1 An electrometer charge amplifier, shown in Figure P2.1, is used to find the inductance of a coil. The dc resistance of the coil is measured to be $R_L = 10 \Omega$. A steady state dc current of 10 mA is passed through the coil. At $t=0$, the switch connects the coil to the op-amp summing junction. The C_F is initially uncharged so $V_o(0) = 0 \text{ V}$.

- (A) Sketch and dimension $i_L(t)$ for $t \geq 0$.
- (B) Sketch and dimension $V_o(t)$ for $t \geq 0$. Give an expression for L in terms of the steady state V_o .

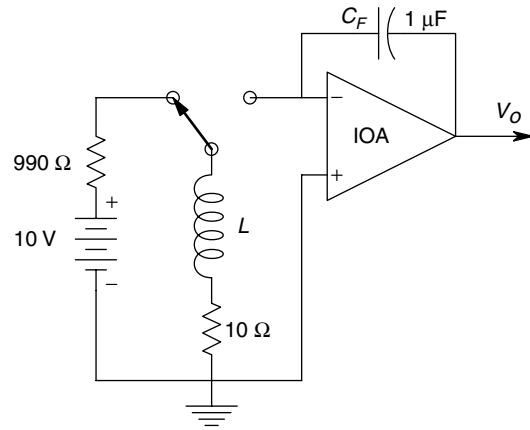


FIGURE P2.1

- 2.2 (A) Derive the transfer function of the controlled source active filter shown in Figure P2.2 in time-constant form.
- (B) Give expressions for the peak gain, the frequency of the peak gain, and its Q . Assume that the op-amp is ideal.
- 2.3 Design a 60 Hz notch filter using a two-loop biquad BPF and a fifth op-amp summer for the system shown in Figure P2.3. The passband gain of the notch filter is to be +20 dB, the BPF's Q is to be 50. The high frequency -3 dB frequency of the filter is to be above 100 kHz. Specify the op-amps and all R_s and C_s in your design. Verify

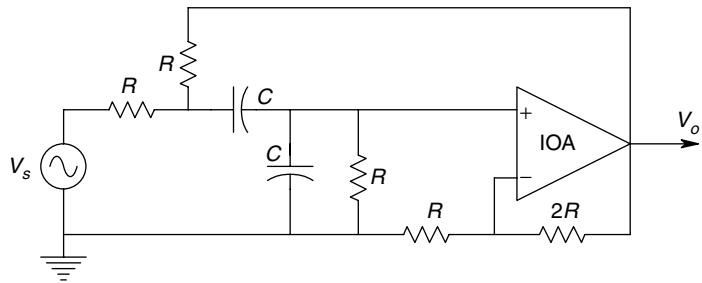


FIGURE P2.2

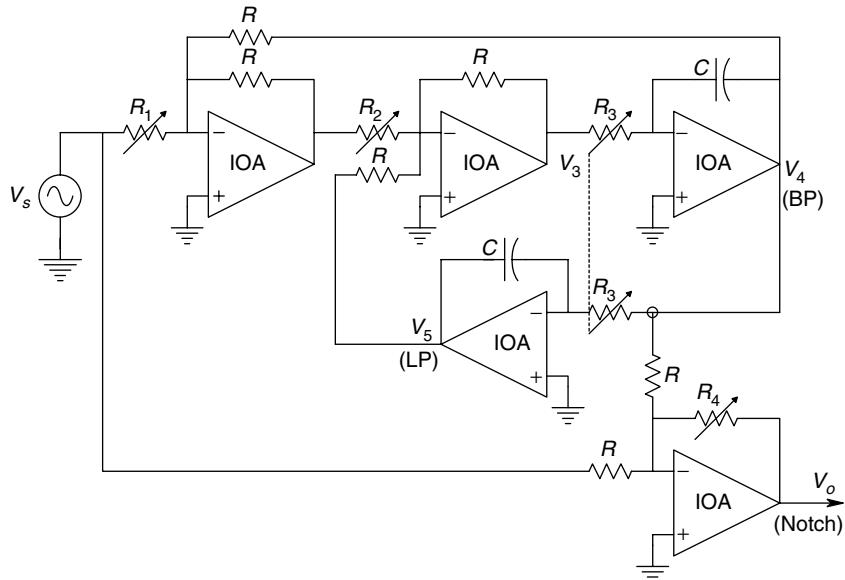


FIGURE P2.3

your design by a SPICE or MicroCap simulation. What is the attenuation of the simulated filter at 60Hz?

- 2.4 We wish to pass a constant current through a SnO_2 carbon monoxide sensor. The presence of CO causes a drop in the sensor's resistance, hence a drop in V_2 . A Howland current source is to be used, shown in Figure P2.4. Define $\alpha \equiv$

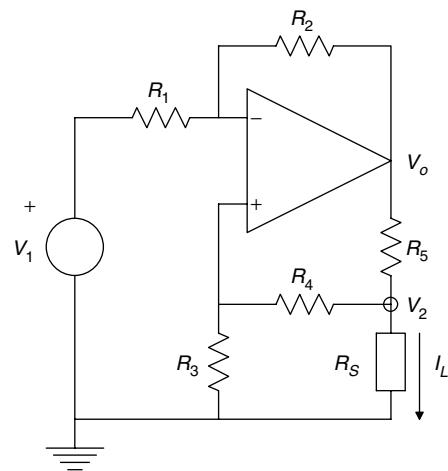


FIGURE P2.4

$R_2/(R_1 + R_2)$, $\beta \equiv R_1/(R_1 + R_2)$, $\gamma \equiv R_3/(R_3 + R_4)$. Assume an ideal op-amp. Using circuit analysis, find the conditions for α , β and γ and the various resistors that make I_L completely independent of the sensor resistance, R_s . Give an expression for the circuit's transconductance, $G_M = I_L / V_1$.

- 2.5 The circuit of Figure P2.5 is a Deboo non-inverting integrator. Assume that the op-amp is ideal and the resistors are exactly matched, $R_1 = R'_1$, $R_2 = R'_2$.

(A) Derive an expression for $\frac{V_o}{V_s}(s)$.

(B) Now let $V_s = 0$. Let dc bias current, I_B , enter the V_i node and a dc bias current, I'_B , enter the V'_i node. Find an expression for $V_o(t)$, assuming that the bias currents are applied as steps at $t = 0$, and $V_o(0) = 0$.

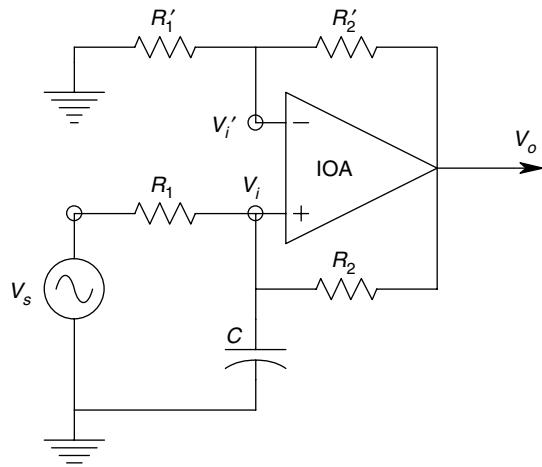


FIGURE P2.5

- 2.6 Two ideal op-amps and a single diode are used to perform a nonlinear operation on the input signal. Find an expression for $V_o = f(V_s, R, R')$. See Figure P2.6.

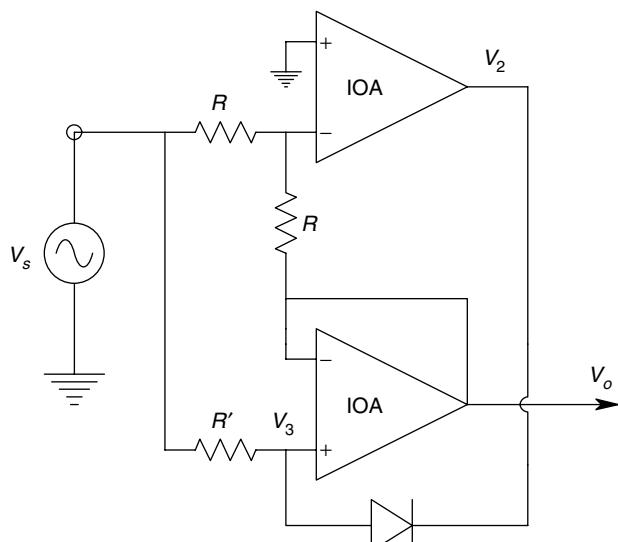


FIGURE P2.6

2.7 The circuit in Figure P2.7 is a type of notch filter design.

- (A) Design the filter to have its notch at 60 Hz. That is, find numerical values for R_1 , R_2 , R_3 and C . Give the dc gain of the filter, and the damping factor of its denominator.
- (B) Simulate the filter's frequency response. Use OP27 op-amp models in SPICE or MicroCap. From the Bode magnitude plot, find the notch's $Q = -3 \text{ dB}$ bandwidth/60.

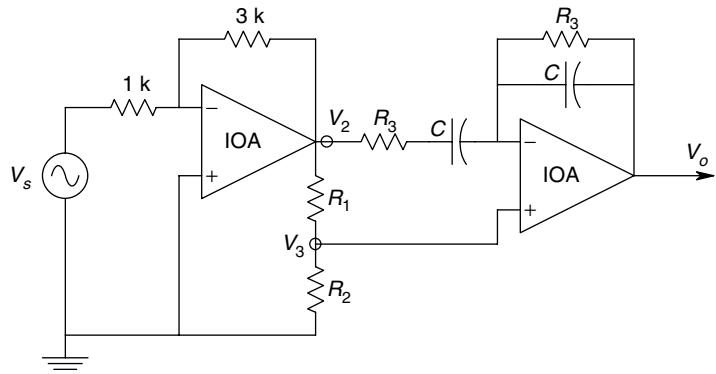


FIGURE P2.7

2.8 The circuit in Figure P2.8 is a single-loop biquad active filter. Assume that the op-amps are ideal.

- (A) Derive an expression for $\frac{V_o}{V_s}(s)$ in time-constant form. Sketch and dimension the Bode magnitude plot for this transfer function.
- (B) Give expressions for this filter's ω_n , Q , and gain at $\omega = \omega_n$. Does this version of the biquad AF have the design flexibility of the 4 op-amp, two-loop biquad AF?

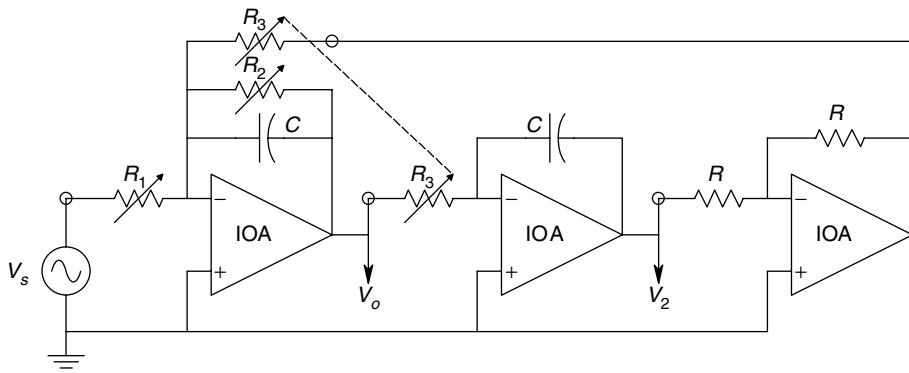


FIGURE P2.8

2.9 Find the expressions for the gains of the seven ideal op-amp circuits shown in Figure P2.9.

2.10 The circuit of Figure P2.10 uses two op-amps to linearize a Wheatstone bridge. Note that voltage V_3 and hence, V_2 are at virtual ground. Write node equations for V_2 and V_3 and use them to find $V_o = f(\Delta R, R, R_2, R_3, V_B)$.

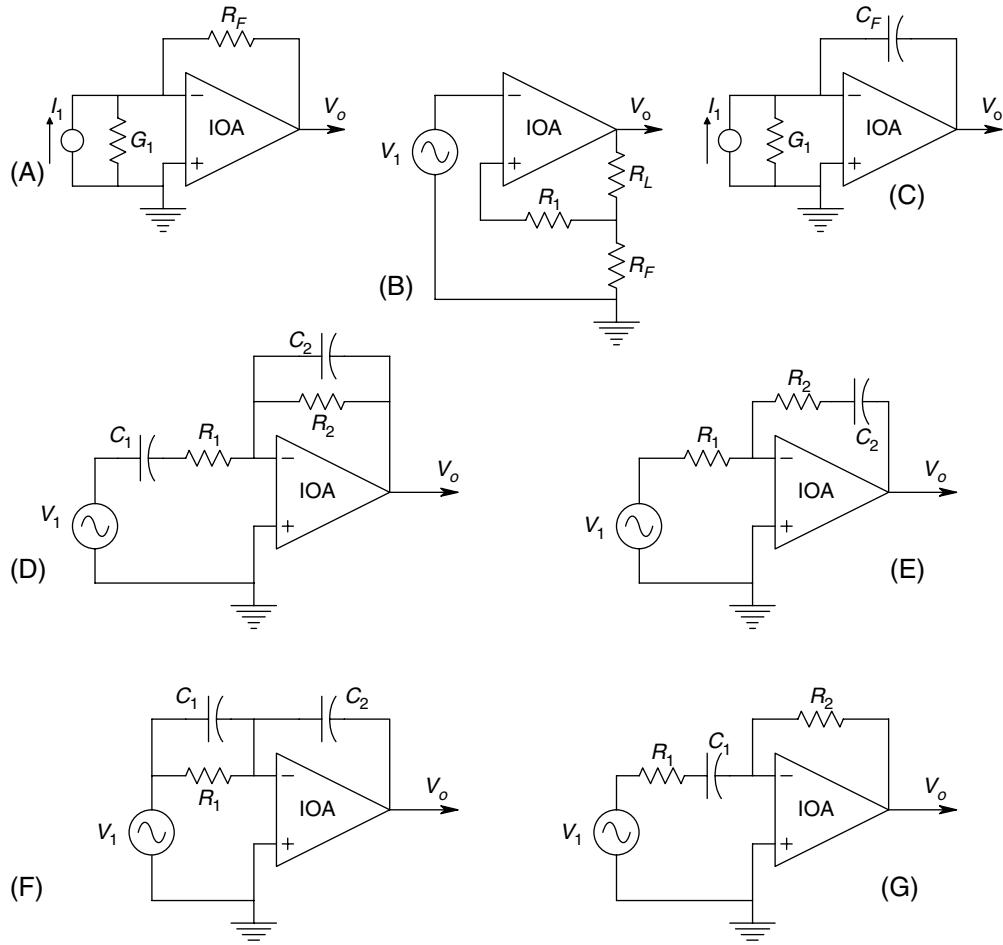


FIGURE P2.9

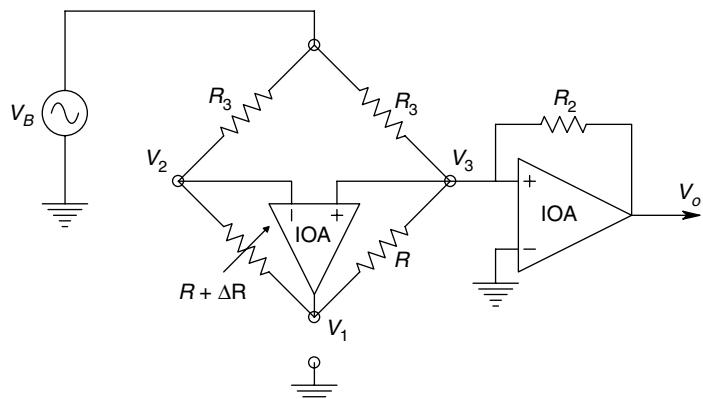


FIGURE P2.10

- 2.11 The CMRR of a certain DA is 120 dB at 60 Hz. When a source, $V_s(t) = 0.004 \sin(377t)$, is connected to the + input and the - input is grounded, the output voltage is 1.0 V RMS. Find numerical values for A_D and A_C in the gain equation: $V_o = A_D V_{1d} + A_C V_{1c}$.

- 2.12 The input voltages to a differential instrumentation amplifier are: $V_1(t) = 0.002 \sin(2\pi 400t)$, and $V'_1(t) = -0.002 \sin(2\pi 400t) + 2 \sin(377t)$. The amplifier's output is given by: $V_o = A_D V_{1d} + A_C V_{1c}$, where $A_D = 10^3$ and $A_C = 10^{-2}$.

- (A) What is the amplifier's CMRR in dB?
- (B) Write an expression for $V_o(t)$.
- 2.13 A Wien bridge is used to make a notch filter at 60 Hz. In Figure P2.13, DA1 has $A_D = K_D$, $A_C = 0$, and DA2 has $A_D = 1$, $A_C = 0$ (note that the DAs are not op-amps).

- (A) Find an expression for the frequency response of the filter, $\frac{V_o}{V_s}(j\omega)$, in time-constant form.
- (B) Pick R and C so the the notch frequency is exactly 60 Hz.
- (C) Give the filter's dc gain, A_{vo} , ω_n , and the Q of the denominator (note that the Q of an underdamped quadratic pole pair is $1/(2\xi)$).

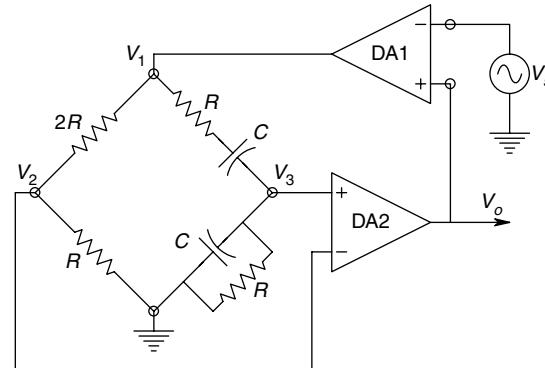


FIGURE P2.13

- 2.14 Design a quadratic, GIC bandpass filter for geophysical applications to condition a seismometer output having $f_n = 1$ Hz and $Q = 10$. Use low noise op-amps. Verify your design with a MicroCap or PSPICE simulation. Note all R s must be in the range 10^3 – 10^6 Ω, and capacitors must lie between 3 pF and 1 μF (cf P7.5, T7).

- 2.15 Figure P2.15 illustrates a voltage tuned active filter. An analog multiplier is used as a variable gain element.

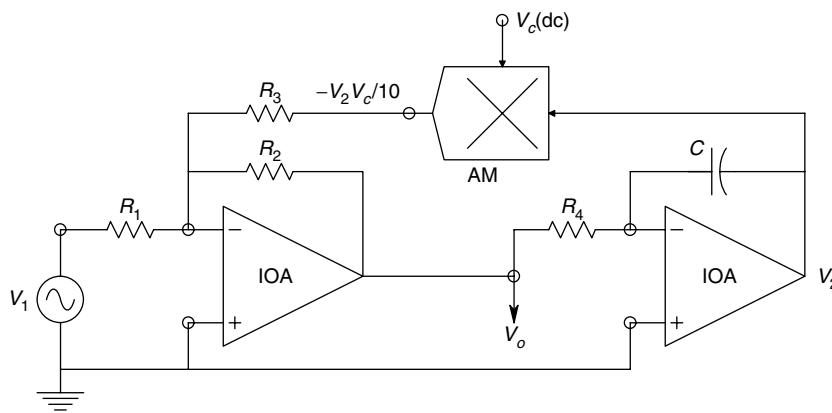


FIGURE P2.15

- (A) Derive an expression for $\frac{V_o}{V_1}(j\omega, V_c)$ in time-constant form.
- (B) Make a dimensioned Bode plot for the filter for $V_c = 0.1, 1.0$ and 10 V . Give expressions for the dc gain and break frequency (cf P7.9, T7)
- 2.16 Derive an expression for the frequency response function of the AF shown in Figure P2.16 in time-constant form. Give expressions for the filter's ω_n , dc gain and damping factors of the numerator and denominator. What kind of filter is this? (cf P7.3, T7)

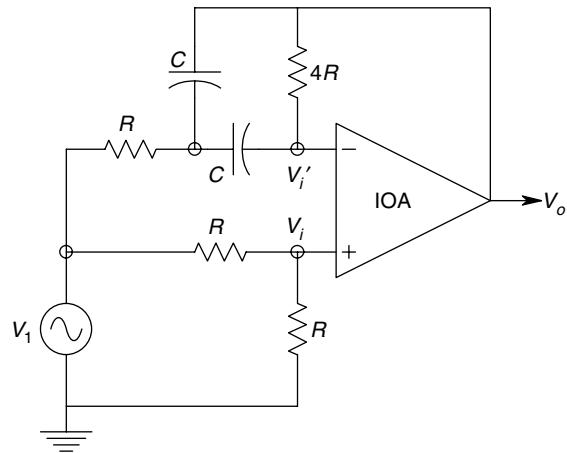


FIGURE P2.16

- 2.17 Find an algebraic expression for $V_o = f(V_1, V_2)$ in the nonlinear analog circuit, Figure P2.17. Assume ideal op-amps. Why is the inverter IOA2 required?

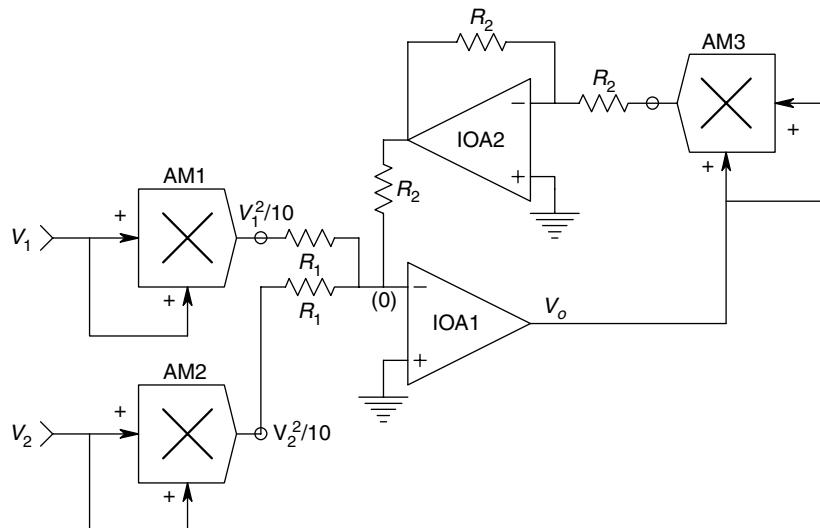


FIGURE P2.17

- 2.18 A comparator and an op-amp are used to create a symmetrical hysteresis voltage transfer function, as shown in Figure P2.18. An ideal comparator and an ideal op-amp are shown in the circuit. The ideal comparator's transfer characteristic is also shown. Assume $R_1 \gg R_3$.

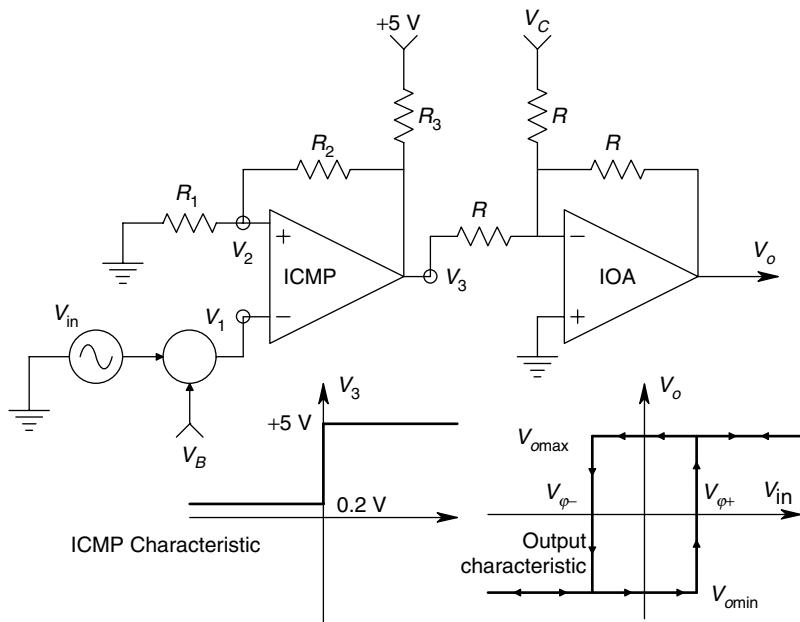


FIGURE P2.18

- (A) Assume $V_B = 0$. Plot and dimension $V_3 = f(V_{in}, R_1, R_2)$ for the ICMP. Note that it has hysteresis. Let $R_1 = R_2 = 100 \text{ k}\Omega$.
- (B) Plot and dimension $V_o = f(V_{in}, R_1, R_2)$. Let $V_B = V_C = 0$, $R = 10 \text{ k}\Omega$.
- (C) The symmetrical output hysteresis characteristic is shown in the figure. Find numerical values of V_B and V_C that are required to give symmetry. Give numerical values for $V_{\varphi+}$, $V_{\varphi-}$, $V_{o\max}$ and $V_{o\min}$.

3

Noise and Coherent Interference in Measurements

3.1 Introduction

Due to the differences in the sources and means of reduction of noise *vs* coherent interference, we will treat them separately in subsections of this chapter. Both noise and interference provide a major limitation to the precision of measurements and the detectability of the QUM. Thus, knowledge of how to design low noise instrumentation systems is an important skill that a practicing instrumentation engineer should learn.

First, let us examine the distinction between noise and coherent interference. Noise is considered to arise in a circuit or measurement system from completely random phenomena. Any physical quantity can be ‘noisy’, but we will restrict ourselves to the consideration of noise voltages and currents in this chapter. Such noise voltages and currents will be considered to have zero means, meaning that they have no additive dc components. The unwanted dc components are best considered to be drift or offset and their reductions are more appropriately treated in a text that deals with dc amplifier design or op-amp applications.

Coherent interference, as the name suggests, generally has its origins in periodic, man-made phenomena, such as power line frequency electric and magnetic fields, radio frequency sources, such as radio and television station broadcast antennas, certain poorly shielded computer equipment, spark discharge phenomena, such as automotive ignitions and motor brushes and commutators, and inductive switching transients, such as SCR motor speed controls, etc. As you will see, minimization of coherent interference is often ‘arty’ and may involve changing the grounding circuits for a system, shielding with magnetic and/or electric conducting materials, filtering, the use of isolation transformers, etc.

Minimizing the impact of incoherent noise in a measurement system often involves a prudent choice of low noise components, certain basic electronic design principles and filtering. The incoherent noise that will be considered is random noise from within the measurement system. Coherent noise usually enters a system from without.

3.2 Random Noise in Circuits

In the following discussions, we assume that the noise is *stationary*. That is, the physical phenomena giving rise to the noise are assumed not to change with time. When a noise is assumed to be stationary, averages over time are equivalent to ensemble averages. An example of a *non-stationary* noise source is a resistor which, at time $t=0$, begins to

dissipate power so that its temperature slowly rises, as does its resistance and thermal noise voltage.

Several statistical methods for describing random noise exist. These include, but are not limited to—the *probability density function*, the *cross* and *auto-correlation functions* and their Fourier transforms, the *cross* and *auto-power density spectra*, and the *root power density spectrum*, which are used to characterize amplifier equivalent input noise.

3.2.1 Probability Density Functions

The *probability density function* (PDF) considers only the amplitude statistics of a stationary noise waveform, $n(t)$, and not how $n(t)$ varies in time. The PDF of $n(t)$ is defined as:

$$p(x) = \frac{\text{Probability that } x < n \leq (x + dx)}{dx} \quad (3.1)$$

where x is a specific value of n taken at some time t and dx is a differential increment in x . The PDF is the mathematical basis for many formal derivations and proofs in statistics and probability theory. The PDF has the properties:

$$\int_{-\infty}^v p(x)dx = \text{Prob}[x \leq v] \quad (3.2)$$

$$\int_{v_2}^{v_1} p(x)dx = \text{Prob}[v_1 < n \leq v_2] \quad (3.3)$$

$$\int_{-\infty}^{\infty} p(x)dx = 1 = \text{Prob}[x \leq \infty] \quad (3.4)$$

Several PDFs are widely used to describe the amplitude characteristics of electrical and electronic circuit noise. These include:

$$p(x) = \left(\frac{1}{\sigma_x \sqrt{2\pi}} \right) \exp \left[\frac{(x - \bar{x})^2}{2\sigma_x^2} \right] \quad \text{Gaussian or normal PDF} \quad (3.5)$$

$$\begin{aligned} p(x) &= 1/2a, & \text{for } -a < x < a \\ p(x) &= 0, & \text{for } x > a \end{aligned} \quad \text{Rectangular PDF} \quad (3.6)$$

$$p(x) = \left(\frac{x}{\alpha^2} \right) \exp \left[\frac{-x^2}{2\alpha^2} \right] \quad \text{Rayleigh PDF} \quad (3.7)$$

Under most conditions, it is assumed that the random noise arising in electronic circuits is modeled by a Gaussian PDF. Many mathematical benefits follow this approximation, such as the output of a linear system is Gaussian with variance σ_y^2 , given that the input

is Gaussian with variance σx^2 . In general, if Gaussian noise passes through a nonlinear system, the output PDF will not be Gaussian.

In this chapter, our concern about noise in instrumentation systems will be focussed in the frequency domain, in particular, the power density spectrum and the root power spectrum, as treated below. We will generally assume that the noise PDFs are Gaussian.

3.2.2 The Power Density Spectrum

We will use a heuristic approach is used to illustrate the meaning of the power density spectrum (PDS) of noise. Formally, the two-sided PDS is the Fourier transform of the *auto-correlation function*, defined below by:

$$R_{nn}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{T} n(t)n(t + \tau)dt \quad (3.8)$$

The two-sided PDS can be written as the Fourier transform of the auto-correlation function:

$$\Phi_{nn}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{nn}(\tau)e^{-j\omega\tau}d\tau \quad (3.9)$$

Since $R_{nn}(\tau)$ is an even function, its Fourier transform, $\Phi_{nn}(\omega)$ is also an even function, which, stated mathematically, means:

$$\Phi_{nn}(\omega) = \Phi_{nn}(-\omega) \quad (3.10)$$

In the following discussion, the *one-sided PDS*, $S_n(f)$, is considered, which is related to the two-sided PDS by:

$$S_n(f) = 2\Phi_{nn}(2\pi f) \quad \text{for } f \geq 0 \text{ and} \quad (3.11)$$

$$S_n(f) = 0 \quad \text{for } f < 0 \quad (3.12)$$

To see how we may experimentally find the one-sided PDS, we examine the system shown in Figure 3.1. Here, a noise voltage is the input to an ideal low-pass filter with adjustable cut-off frequency, f_c . The output of the ideal low-pass filter is measured with a broadband, true RMS, AC voltmeter. We begin with $f_c = 0$ and systematically increase f_c , each time recording the square of the RMS meter reading, or mean squared output

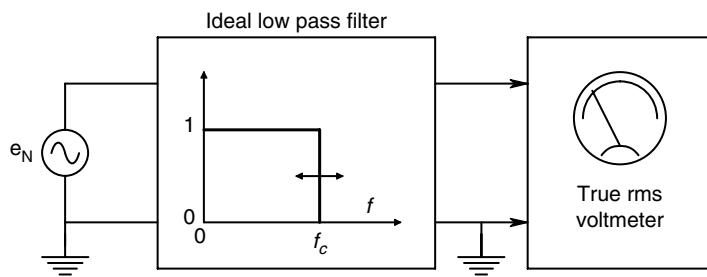
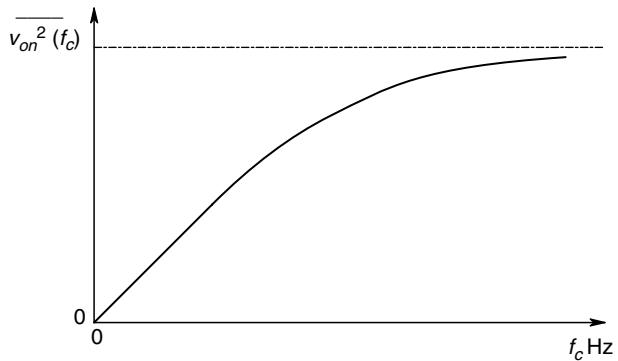
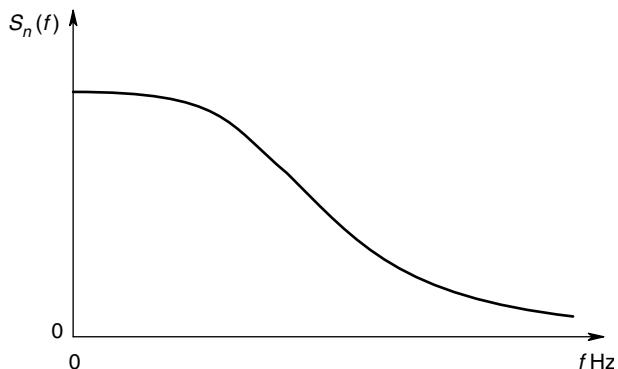


FIGURE 3.1

System relevant to the derivation of the one-sided power density spectrum.

**FIGURE 3.2**

Mean squared output noise of the system of Figure 3.1 as a function of ideal low-pass filter cut-off frequency.

**FIGURE 3.3**

A one-sided, noise power density spectrum; derived by differentiation of the plot of V_{on}^2 vs f_c with respect to f .

voltage, V_{on}^2 , of the filter. As f_c is increased, V_{on}^2 increases monotonically, as shown in Figure 3.2. Due to the finite bandwidth of the noise source, V_{on}^2 eventually reaches an upper limit, which is the total mean squared noise voltage of the noise source. The plot of $V_{on}^2(f_c)$ vs f_c is called the *cumulative mean squared noise characteristic* of the noise source. Its units are mean squared volts, in this example. A simple interpretation of the one-sided, noise power density spectrum is that it is the derivative, or slope, of the mean squared noise characteristic curve. Mathematically stated, this is:

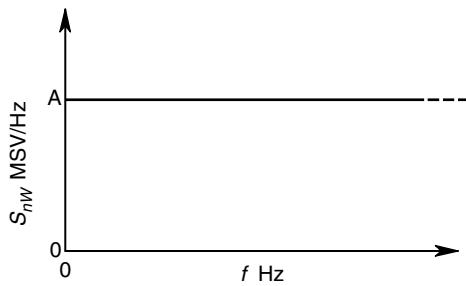
$$S_n(f) = \frac{dV_{on}^2(f)}{df} \text{ MS V/Hz} \quad (3.13)$$

A plot of a typical noise PDS is shown in Figure 3.3. Note that a practical PDS generally drops off to zero at high frequencies.

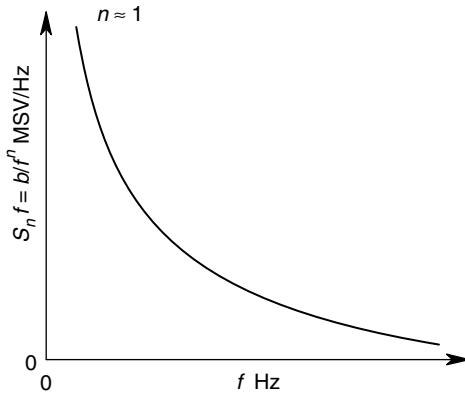
A question which is sometimes asked by those first encountering the PDS concept is why it is called a power density spectrum. The power concept has its origin in the consideration of noise in communication systems and has little meaning in the consideration of noise in instrumentation systems. One way to rationalize the power term is to consider a noise voltage source with a one ohm load. In this case, the average power dissipated in the resistor is simply the total mean squared noise voltage.

From the heuristic definition of the PDS above, we see that the total mean squared voltage in the noise voltage source can be written as:

$$\overline{V_{on}^2} = \int_{-\infty}^{\infty} S_n(f) df \quad (3.14)$$

**FIGURE 3.4**

A white noise power density spectrum.

**FIGURE 3.5**

A $1/f$ noise power density spectrum.

The mean squared voltage in the frequency interval, (f_1, f_2) , is found by:

$$\overline{v_{on}^2}(f_1, f_2) = \int_{f_1}^{f_2} S_n(f) df \quad (3.15)$$

Often noise is specified or described using root power density spectra, which are simply plots of the square root of $S_n(f)$ vs f , which have the units of RMS volts (or other units) per root Hz.

Special (ideal) PDSs are used to model or approximate portions of real PDSs. These include The white noise PDS, and the one-over- f PDS. A white noise PDS is shown in Figure 3.4. Note that this PDS is flat; this implies that

$$\int_0^{\infty} S_{nw}(f) df = \infty \quad (3.16)$$

which is clearly not realistic. A $1/f$ PDS is shown in Figure 3.5. The $1/f$ spectrum is often used to approximate the low frequency behavior of real PDSs. Physical processes which generate $1/f$ -like noise include surface phenomena associated with electrochemical electrodes, carbon composition resistors carrying direct current (metallic resistors

are substantially free of $1/f$ noise) and surface imperfections affecting emission and diffusion phenomena in semiconductor devices. The presence of $1/f$ noise can present problems in the electronic signal conditioning systems used for low level, low frequency and dc measurements, such as those found in biomedical and geophysical applications.

3.2.3 Sources of Noise in Signal Conditioning Systems

Sources of noise in signal conditioning systems can be separated into two major categories—noise from passive resistors and noise from active circuit elements such as bipolar junction transistors, field effect transistors and vacuum tubes. In most cases, the Gaussian assumption for noise amplitude PDFs is valid and the noise generated can generally be assumed to be white over a major portion of spectrums.

3.2.3.1 Noise from Resistors

From statistical mechanics, it can be shown that any pure resistance at some temperature T K will have a zero mean noise voltage associated with it. This noise voltage appears in series with the (noiseless) resistor to form a Thevenin equivalent circuit. From dc to radio frequencies, where the resistor's capacitance to ground and its lead inductance can no longer be neglected, the resistor's noise is well-modeled by a Gaussian white noise source. Noise from resistors is called *thermal* or *Johnson noise*. Its one-sided PDS is given by:

$$S_n(f) = 4kTR \text{ MS V/Hz}, \quad 0 \leq f \leq \infty \quad (3.17)$$

where k is Boltzmann's constant (1.38×10^{-23}), T is in kelvin and R is in ohms. In a given bandwidth, the mean squared noise from a resistor can be written as:

$$v_{on}^2(B) = \int_{f_1}^{f_2} S_n(f) df = 4kTR(f_2 - f_1) = 4KTRB \text{ MS V} \quad (3.18)$$

where B is the equivalent noise bandwidth, $(f_2 - f_1)$ in Hz.

A Norton equivalent of the Johnson noise source from a resistor can be formed by assuming a white noise current source with PDS:

$$S_{ni}(f) = 4kTG \text{ MS A/Hz} \quad (3.19)$$

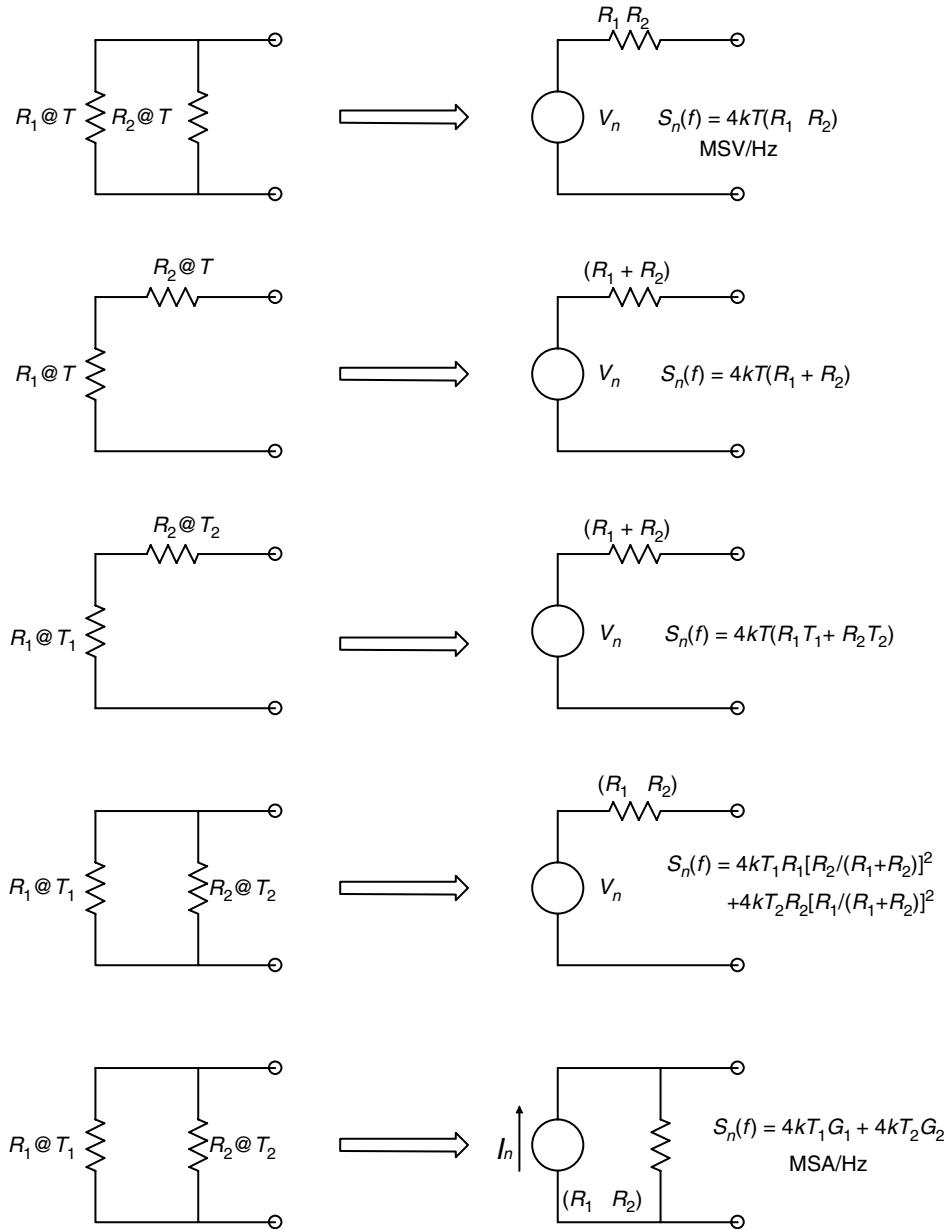
in parallel with a noiseless conductance, $G = 1/R$.

The Johnson noise from several resistors connected in a network may be combined into a single noise voltage source in series with a single, noiseless, equivalent resistor. Figure 3.6 illustrates some of these reductions for two terminal circuits.

It has been observed that when dc (or average) current is passed through a resistor, the basic Johnson noise PDS is modified by the addition of a $1/f$ spectrum.

$$S_n(f) = 4kTR + AI^2/f \text{ MS V/Hz} \quad (3.20)$$

where I is the average or dc component of current through the resistor, and A is a constant that depends on the material from which the resistor is constructed (e.g. carbon

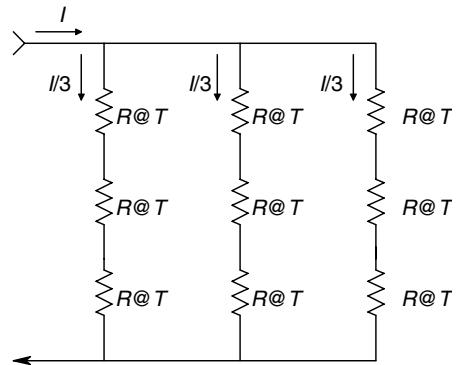
**FIGURE 3.6**

Means of representing the thermal or Johnson white noise from resistors using Thevenin and Norton equivalent circuits. Noises are always added as mean squared quantities.

composition, resistance wire, metal film, etc.). An important parameter for resistors carrying average current is the crossover frequency, f_c , where the $1/f$ PDS equals the PDS of the Johnson noise. This is:

$$f_c = AI^2 / 4kTR \text{ Hz} \quad (3.21)$$

It is possible to show that the f_c of a noisy resistor can be reduced by using a resistor of the same type, but having a higher wattage or power dissipation rating. As an example

**FIGURE 3.7**

Equivalent circuit for a nine-fold increase in resistor wattage rating, the resistance value remaining the same. Note that an N -fold increase in wattage does not affect the thermal noise from the resistor, but does give an N -fold reduction in the $1/f$ noise spectrum from the resistor.

of this principle, consider the circuit of Figure 3.7 in which a single resistor of $R \Omega$, carrying a dc current I is replaced by nine resistors of resistance R , connected in series-parallel to give a net resistance R which dissipates nine times the power of a single resistor R . The noise PDS in any one of the nine resistors is seen to be:

$$S'_n(f) = 4kTR + A(I/3)^2/f \text{ MSV/Hz} \quad (3.22)$$

Each of the nine PDSs given by equation 3.22 above contributes to the net PDS seen at the terminals of the composite 9W resistor. Each resistor's equivalent noise voltage source 'sees' a voltage divider formed by the other eight resistors in the composite resistor. The attenuation of each of the nine voltage dividers is given by:

$$\frac{3R/2}{3R/2 + 3R} = 1/3 \quad (3.23)$$

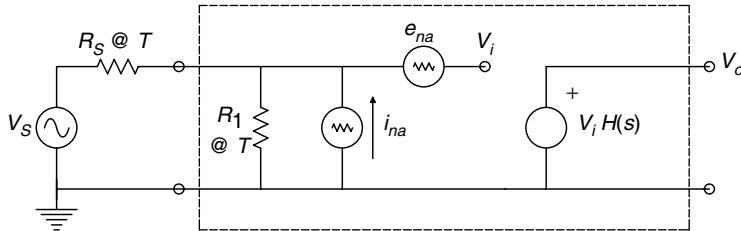
The net voltage PDS at the composite resistor's terminals may be found by superposition of PDSs as:

$$S_{n(9)}(f) = \sum_{j=1}^9 [4kTR + A(I/3)^2/f](1/3)^2 = 4kTR + AI^2/9f \text{ MS V/Hz} \quad (3.24)$$

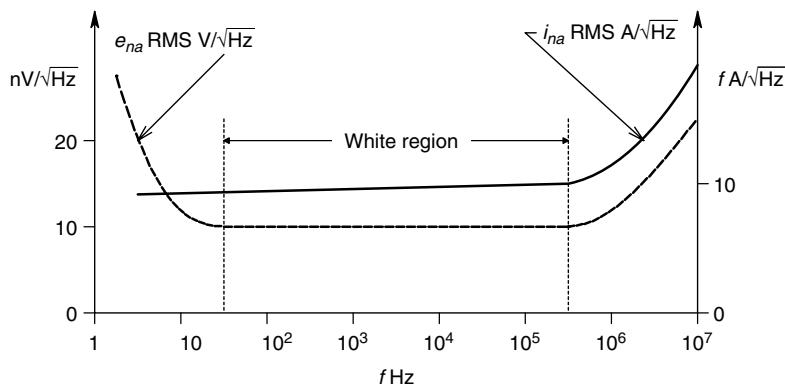
Thus the composite, 9 W resistor enjoys a nine-fold reduction in the $1/f$ spectral power because the dc current density through each element is reduced by $1/3$. However, the Johnson noise PDS remains the same. It is safe to generalize that the use of high wattage resistors of a given type and resistance will result in reduced $1/f$ noise generation when the resistor carries dc (average) current.

3.2.3.2 The Two Source Noise Model for Active Devices

Noise arising in JFETs, BJTs and other complex IC amplifiers is generally described by the *two source input model*. The total noise observed at the output of an amplifier, given that its input terminals are short circuited, is accounted for by defining an equivalent short circuited input noise voltage, e_{na} , which replaces all internal noise sources affecting the amplifier output under short circuited input conditions. The amplifier, shown in Figure 3.8, is now considered noiseless. The e_{na} for many low noise, discrete transistors and IC amplifiers is specified by manufacturers.

**FIGURE 3.8**

Equivalent two-source model for a noisy amplifier.

**FIGURE 3.9**

Plots of typical input noise root power spectrums for a JFET input amplifier.

The e_{na} is a root PDS (i.e. it is a function of frequency) and has the units of RMS V/\sqrt{Hz} . Figure 3.9 illustrates a plot of a typical $e_{na}(f)$ vs f for a low noise JFET. Also shown in Figure 3.9 is a plot of $i_{na}(f)$ vs f for the same device.

In addition to the equivalent short circuited input noise voltage, the modelling of the net noise characteristics of amplifiers requires the inclusion of a current noise source, i_{na} , as shown in Figure 3.8. The $i_{na}(f)$ is the root PDS of the input equivalent noise current. Its units are RMS A/\sqrt{Hz} . Note that both $e_{na}(f)$ and $i_{na}(f)$ have flat, mid-frequency portions which invite approximation by white noise sources. At high frequencies, both the equivalent noise root PDSs slope upward. The $e_{na}(f)$ for discrete JFETs, BJTs and IC amplifiers shows a $1/\sqrt{f}$ region at low frequencies.

3.2.3.3 Noise in JFETs

Certain select, discrete JFETs are used in the design of low noise headstages for signal conditioning in measurement systems. Some JFETs give exceptional low noise performance in the audio and sub-audio frequency regions of the spectrum, while others excel in the video and radio frequency end of the spectrum, making them suitable for applications for RF oscillators, mixers and tuned amplifiers.

van der Ziel (1974) has shown that the theoretical thermal noise generated in the conducting channel of a JFET can be approximated by an equivalent short circuited input noise with PDS given by:

$$e_{na}^2 = 4kT/g_m = 4kT/g_{m0} \sqrt{\frac{I_{DSS}}{I_{DQ}}} \text{ MS V/Hz} \quad (3.25)$$

where g_m is the FET's small signal transconductance, measured when $V_{GS}=0$ and $I_D=I_{DSS}$. I_{DSS} = the dc drain current measured for $V_{GS}=0$ and $V_{DS} > V_P$, and I_{DQ} = the quiescent dc drain current at the FET's operating point where $V_{GS}=V_{GSO}$. In reality, due to the presence of $1/f$ noise, the theoretical short circuited input voltage PDS expression can be modified to:

$$e_{na}^2(f) = (4kT/g_m)(1 + f_c/f^n) \text{ MS V/Hz} \quad (3.26)$$

The exponent n has the range $1 < n < 1.5$ and is determined on the basis of device and lot. This n is usually set equal to one for algebraic simplicity. The origins of the $1/f^n$ effect in JFETs are poorly understood. Note that e_{na} , given by equation 3.26 is temperature dependent. Heatsinking or actively cooling the JFET will reduce e_{na} . The parameter f_c , used in equation 3.26 is the corner frequency of the $1/f$ noise spectrum. Depending on the device, it can range from below 10 Hz to above 1 kHz. The f_c is generally quite high in RF and frequency JFETs because in this type of transistor, $e_{na}(f)$ dips to around 2 nV/ $\sqrt{\text{Hz}}$ in the 10^5 – 10^7 Hz region, which is quite good.

Reverse biased JFET gates have a dc leakage current, I_{GL} , which produces *shot noise*, which is superimposed on the leakage current. This noise component in I_{GL} is primarily due to the random occurrence of charge carriers that have enough energy to cross the reverse biased gate channel diode junction. The PDF of the gate current shot noise is Gaussian and its PDS is given by:

$$i_{na}^2 = 2qI_{GL} \text{ MS A/Hz} \quad (3.27)$$

where $q = 1.602 \times 10^{-19}$ C (electron charge) and I_{GL} is the dc gate leakage current in amperes. I_{GL} is typically about 2 pA, hence i_{na} is about 1.8 RMS fA/ $\sqrt{\text{Hz}}$ in the flat mid-range of $i_{na}(f)$. The $i_{na}(f)$, like $e_{na}(f)$, shows a $1/f$ characteristic at low frequencies. The PDS, $i_{na}^2(f)$, can be modeled by:

$$i_{na}^2(f) = 2qI_{GL}(1 + f/f_{ic}) \text{ MS A/Hz} \quad (3.28)$$

where f_{ic} is the current noise corner frequency.

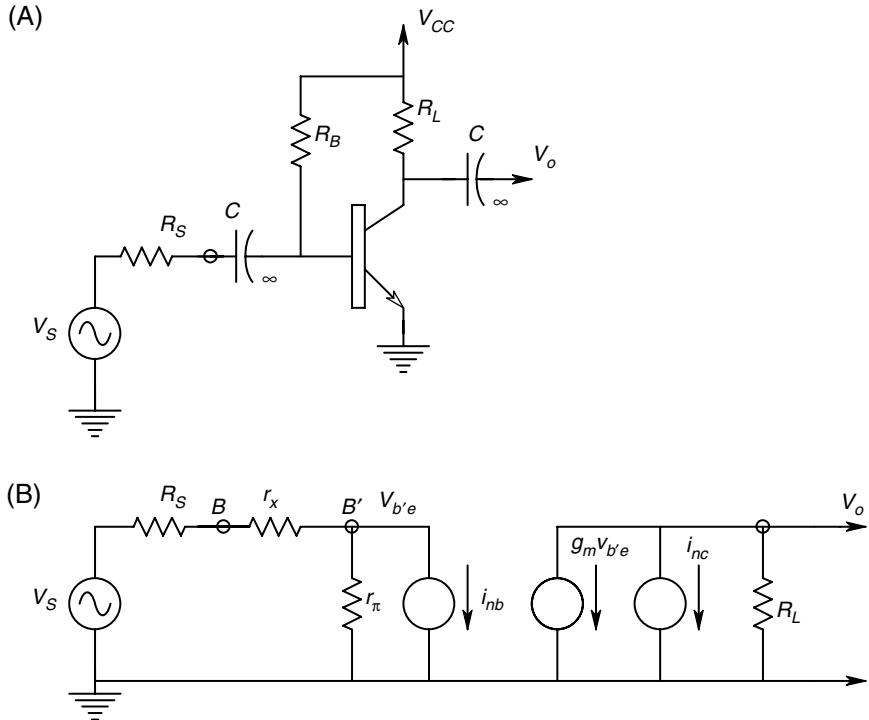
For some transistors, the measured $e_{na}(f)$ and $i_{na}(f)$ have been found to be greater than the predicted theoretical values, while in other cases, they have been found to be less. No doubt, the causes for these discrepancies lie in the oversimplifications used in their derivations.

3.2.3.4 Noise in BJTs

The values of e_{na} and i_{na} associated with bipolar junction transistor amplifiers depend strongly on the device operating (Q) point. This is because there are shot noise components superimposed on the quiescent base and collector currents. The small signal model of a grounded emitter BJT amplifier is shown in Figure 3.10A. In this circuit, we assume negligible noise from R_L , the voltage controlled current source, $g_m v_{be}$, and the small signal base input resistance, r_π . The shot noise sources are:

$$i_{nb}^2 = 2qI_{BQ} \text{ MS A/Hz} \quad (3.29)$$

$$i_{nc}^2 = 2q\beta I_{BQ} \text{ MS A/Hz} \quad (3.30)$$

**FIGURE 3.10**

A. A simple, noisy BJT, RC amplifier. B. Mid-frequency equivalent circuit of the BJT amplifier. It is assumed that $R_B \gg r_\pi$. r_π and R_L are assumed to be noiseless, and $R'_s = r_x + R_s$ makes thermal noise. i_{nb} and i_{nc} are white shot noise current sources.

In this example, it is algebraically simpler to not find the equivalent input e_{na} and i_{na} . Rather, we work directly with the two white shot noise sources in the mid-frequency, hybrid pi, small signal model (Figure 3.10B). It can be shown [Northrop, 1990] that the total output noise voltage PDS is given by:

$$S_{no} = 4kTRL + 2q(I_{BQ}/\beta)(\beta R_L)^2 + \frac{4kTR'_s(\beta R_L)^2 + 2qI_{BQ}R'_s(\beta R_L)^2}{(V_T/I_{BQ} + r_x)^2} \text{ MSV/Hz} \quad (3.31)$$

where it is clear that r_π is approximated by V_T/I_{BQ} , $R_s = r_x + R_s$ and we neglected the Johnson noise from R_L because it is numerically small compared to the other terms. It is easy to show that the MS output signal can be written as:

$$v_{os}^2 = v_s^2(\beta R_L)^2 / (V_T/I_{BQ} + r_x)^2 \text{ MSV} \quad (3.32)$$

Thus the MS SNR at the amplifier output can be written as:

$$\text{SNR}_o = \frac{v_s^2/B}{4kTR'_s + 2qI_{BQ}R'_s{}^2 + 2q(I_{BQ}/\beta)(V_T/I_{BQ} + R'_s)^2} \quad (3.33)$$

where B is the specified, equivalent (Hz) noise bandwidth for the system. The SNR_{o} , given by equation 3.33, has a maximum for some non-negative I_{BQMAX} . The I_{BQMAX} which will give this maximum can be found by differentiating the denominator of equation 3.33 and setting the derivative equal to zero. This gives:

$$I_{BQMAX} = V_T / (R'_s \sqrt{\beta + 1}) \text{ A} \quad (3.34)$$

What we should remember from the above exercise is that the best noise performance for BJT amplifiers is a function of quiescent biasing conditions (Q point). Often these conditions have to be found by trial and error when working at high frequencies. While individual BJT amplifiers may best be modelled for noise analysis with the two shot noise current sources, as we did above, it is customary when describing complex, BJT IC amplifier noise performance to use the more general and more easily applicable e_{na} and i_{na} two-source model.

3.3 Propagation of Gaussian Noise through Linear Filters

In a formal, rigorous treatment of noise in linear systems, it is possible to show that given a Gaussian input, the PDF of the output of a linear system is Gaussian. In addition, the PDS of the system's output noise is given by:

$$S_y(f) = S_x(f)|H(2\pi f j)|^2 \quad (3.35)$$

This is the scalar product of the positive real input PDS and the square of the magnitude of the system's frequency response function. This relation can be extended to include two or more cascaded systems, as shown in Figure 3.11.

$$S_y(f) = S_x(f)|H(2\pi f j)|^2|G(2\pi f j)|^2 \text{ MS V/Hz} \quad (3.36)$$

or

$$S_y(f) = S_x(f)|H(2\pi f j)G(2\pi f j)|^2 \text{ MS V/Hz} \quad (3.37)$$

If white noise with a PDS, $S_x(f) = \eta \text{ MS V/Hz}$, is the input to a linear system, then the output PDS is simply:

$$S_y(f) = \eta|H(2\pi f j)|^2 \text{ MS V/Hz} \quad (3.38)$$

The total MS output noise of the system is:

$$\overline{v_{on}^2} = \int_0^\infty S_y(f)df = \eta \int_0^\infty |H(2\pi f j)|^2 df \text{ MS V} \quad (3.39)$$

FIGURE 3.11

Propagation of noise through two cascaded linear systems.

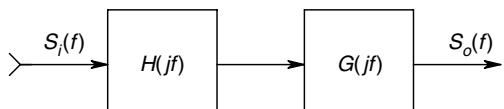


TABLE 3.1

Noise gain squared bandwidth products for some common system frequency response functions

Frequency Response Function, $H(j\omega)$	Gain Squared Bandwidth (Hz)
1. $\frac{K_V}{j\omega\tau + 1}$	$K_V^2[1/4\tau]$
2. $\frac{K_V}{(j\omega\tau_1 + 1)(j\omega\tau_2 + 1)}$	$K_V^2[1/(4(\tau_1 + \tau_2))]$
3. $\frac{K_V}{(1 - \omega^2/\omega_n^2) + j\omega(2\xi/\omega_n)}$	$K_V^2[\omega_n/8\xi]$
4. $\frac{K_V j\omega(2\xi/\omega_n)}{(1 - \omega^2/\omega_n^2) + j\omega(2\xi/\omega_n)}$	$K_V^2[\omega_n\xi] = K_V^2[\omega_n/2Q]$
5. $\frac{K_V j\omega\tau_1}{(j\omega\tau_1 + 1)(j\omega\tau_2 + 1)}$	$\left[K_V^2 \frac{\tau_1}{4\tau_2(\tau_1 + \tau_2)} \right]$

for transfer functions with one more finite pole than zeros, the right hand integral of equation 3.39 may be shown to be the product of the square of the transfer function's low frequency or mid-band gain and the filter's equivalent Hz noise bandwidth. Therefore, the filter's *gain squared bandwidth* product is:

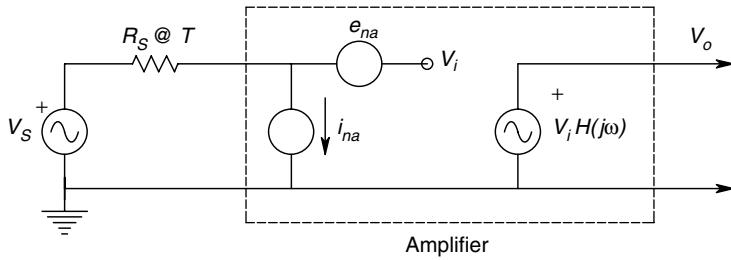
$$\text{GAIN}^2\text{BW} = \int_0^\infty |H(2\pi jf)|^2 df \quad (3.40)$$

Gain squared bandwidth integrals have been evaluated using complex variable theory for a number of transfer functions [James, Nichols and Philips, 1947]. Table 3.1, gives the gain squared bandwidth integrals for five common transfer functions. Note that the equivalent noise bandwidths (in brackets in each case) are in Hz, not r/s. Also note the absence of 2π factors in these expressions. Gain squared bandwidth integrals are used to estimate the total MS output noise from amplifiers having (approximate) white noise input sources and are useful in calculating output SNRs.

3.4 Broadband Noise Factor and Noise Figure of Amplifiers

The noise factor of an amplifier is defined as the ratio of the mean squared signal-to-noise ratio at the input of the amplifier to the signal-to-noise ratio at the output of the amplifier. Since a real amplifier is noisy and adds noise to the signal as well as amplifying it, the SNR_o is always less than the SNR_I , hence the *noise factor* (F) is always >1 for a noisy amplifier.

$$F \equiv \frac{\text{SNR}_I}{\text{SNR}_o} \quad (3.41)$$

**FIGURE 3.12**

Minimum noisy amplifier model relevant to the calculation of amplifier noise figure.

The *noise figure* is defined as

$$NF \equiv 10 \log_{10}(F) \quad (3.42)$$

when the SNRs are in terms of MS quantities. Figure 3.12 illustrates a minimum model for a noisy amplifier. Here, we assume that the spectrums of e_{na} and i_{na} are white and that $R_s \gg R_1$. The MS input signal is $S_i = v_s^2$, the MS output signal is $S_o = K_V^2 v_s^2$, where K_V^2 is the mid-band gain of the amplifier. The MS input noise is simply that which is associated with v_s (here zero) plus the Johnson noise from the source resistance, R_s , in a specified Hz noise bandwidth, B . Therefore it is written as:

$$N_i = 4kTR_s B \text{ MS V} \quad (3.43)$$

The MS noise at the amplifier's output, N_o , is composed of three components: one from the R_s Johnson noise and two from the equivalent noise sources. N_o can be written as:

$$\begin{aligned} N_o &= (4kTR_s + e_{na}^2 + i_{na}^2 R_s^2) \int_0^\infty |H(2\pi jf)|^2 df \\ &= (4kTR_s + e_{na}^2 + i_{na}^2 R_s^2) K_V^2 B \text{ MS V} \end{aligned} \quad (3.44)$$

Using the definition for F , we find that the noise factor for the simple noisy amplifier model can be written as:

$$F = 1 + \frac{e_{na}^2 + i_{na}^2 R_s^2}{4kTR_s} \quad (3.45)$$

Note that this expression for F contains no bandwidth terms since they cancel out. When the NF is given for an amplifier, R_s must be specified, as well as the Hz bandwidth B over which the noise is measured. The temperature should also be specified, although common practice usually sets T at 300 K.

For practical amplifiers, NF and F are functions of frequency because e_{na} and i_{na} are functions of frequency (Figure 3.9). For a given R_s , F tends to rise at low frequencies due to the $1/f$ components in the equivalent input noise sources. F also increases again at high frequencies due to the high frequency increases in e_{na} and i_{na} . Often, we are interested in the noise performance of an amplifier in either low or high frequencies, where NF and F are not minimum. To examine the detailed noise performance of an amplifier at low and high frequencies, we use the *spot noise figure*, discussed below.

3.5 Spot Noise Factor and Figure

Spot noise measurements are made through a narrow bandpass filter in order to evaluate an amplifier's noise performance in a certain frequency range, particularly where $e_{na}(f)$ and $i_{na}(f)$ are not constant, such as in the $1/f$ range. Figure 3.13A illustrates a set of spot noise figure contours for a commercial, low noise pre-amplifier, having applications at audio frequencies. Figure 3.13B shows the spot NF contours when an ideal transformer with $n = 100$ is used to couple the Thevenin source to the amplifier.

A system for determining the spot noise of an amplifier is shown in Figure 3.14. An adjustable white noise source is used at the input for the amplifier. Note that the output resistance of the white noise source plus some external resistance must add up to R_s , the specified Thevenin equivalent input resistance. The system is used is described below. First, the bandpass filter (BPF) is set to the desired center frequency and the white noise generator output is set to $e_N = 0$. We assume that the total MS noise at the system output under these conditions can be written as:

$$N_o(f_c) = [4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2]K_V^2 B_F \text{ MS V} \quad (3.46)$$

where $e_{na}(f_c)$ = value of e_{na} at the center frequency, f_c ,

B_F = equivalent noise bandwidth of the BPF

K_V = combined gain of the amplifier under measurement at f_c , the BPF at f_c and the post amplifier.

K_V can be written as:

$$K_V = |H(2\pi f_c)| K_F K_A \quad (3.47)$$

Second, the white noise source is made non-zero and adjusted so that the true RMS meter reads $\sqrt{2}$ higher than in the first case with $e_N = 0$. The MS output voltage can now be written as:

$$\begin{aligned} N'_o(f_c) &= 2N_o(f_c) = 2[4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2]K_V^2 B_F \\ &= [e_N^2 + 4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2]K_V^2 B_F \end{aligned} \quad (3.48)$$

Under this condition, it is evident that

$$e_N^2 = 4kTR_s + e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2 \quad (3.49)$$

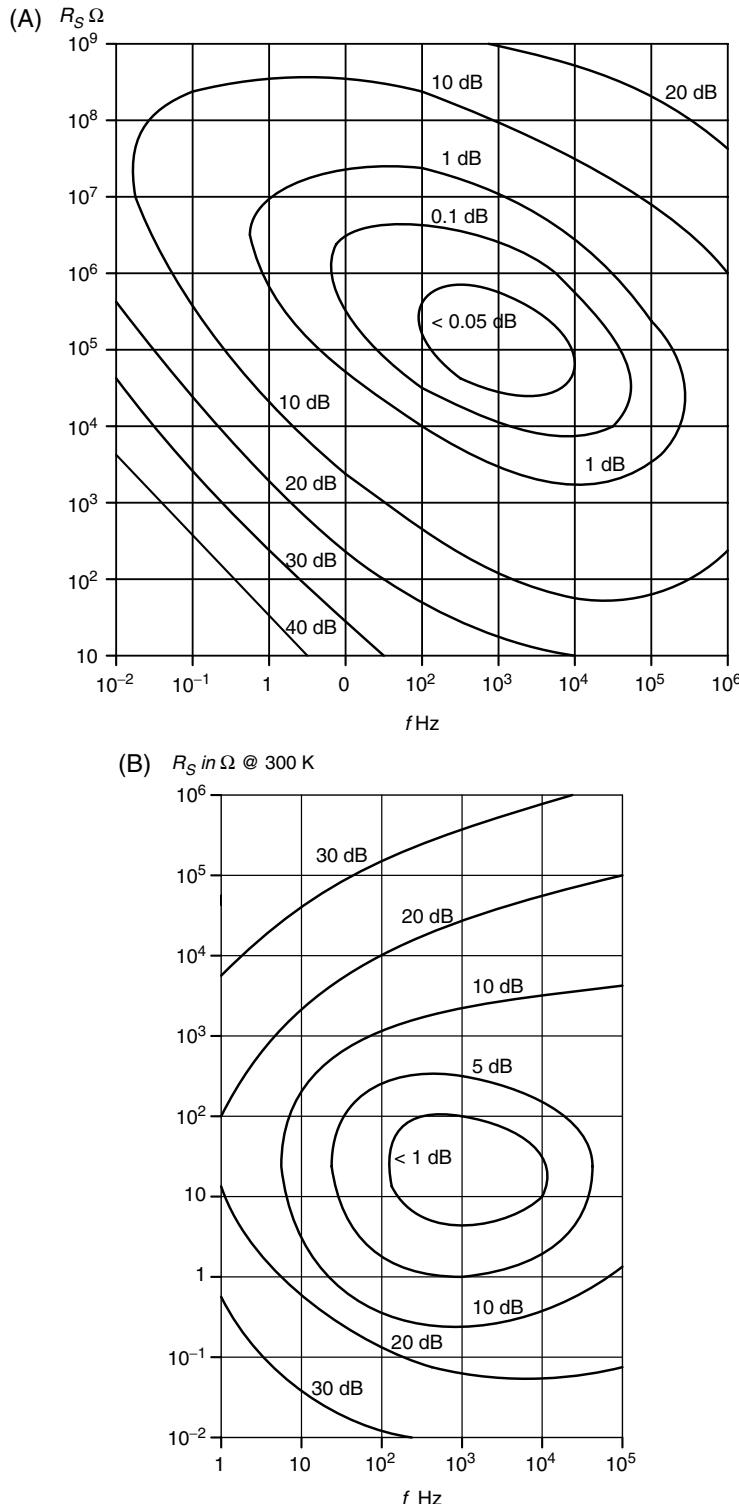
Hence,

$$[e_{na}^2(f_c) + i_{na}^2(f_c)R_s^2] = [e_N^2 - 4kTR_s] \quad (3.50)$$

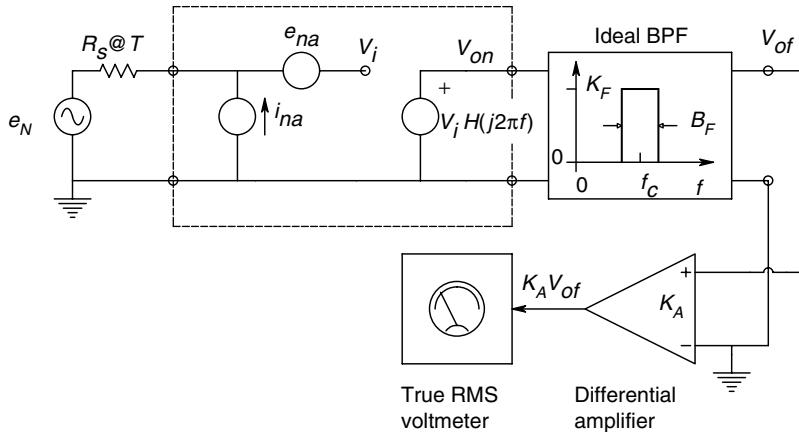
If the left hand side of equation 3.50 is substituted into equation 3.45 for the Noise Factor, F , we find:

$$F_{spot} = e_N^2 / (4kTR_s) \quad (3.51)$$

Note that this expression for the Spot Noise Factor does not contain specific terms for the measurement of center frequency, f_c , or the bandpass filter's Hz noise bandwidth, B_F . However, these parameters should be specified when giving F_{spot} for an amplifier. The F_{spot} is actually calculated by setting f_c and R_s , then determining the e_N^2 value that doubles

**FIGURE 3.13**

A. Spot noise figure contours for a typical amplifier with a direct input having a Thevenin source resistance R_S at temperature T K. B. Noise figure contours for the same amplifier using a $1:100$ turns ratio input transformer. Note the shift of the minimum contour to a lower range of R_S .

**FIGURE 3.14**

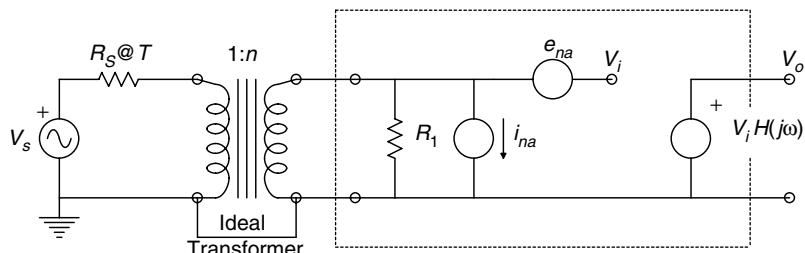
System for measuring the spot noise figure of an amplifier.

the MS output noise. This value of e_N^2 is then divided by the calculated white noise spectrum from the resistor, R_s , to give a value for F_{spot} at a given frequency and R_s .

It is also possible to measure F_{spot} using a sinusoidal source of frequency, f_c , instead of the calibrated white noise source, e_N . Refer to the problems at the end of this chapter for a detailed treatment of this method.

3.6 Transformer Optimization of Amplifier F_{spot} and Output SNR

Figure 3.13A illustrates the fact that for a given set of internal biasing conditions, the amplifier will have an optimum operating region where NF_{spot} is a minimum in R_s, f_c space. In some instances, the input sensor to which the amplifier is connected to has an R_s that is far smaller than the R_s that will give the lowest NF_{spot} on the amplifier's spot NF contours. Consequently, the system (i.e. sensor and amplifier) are not operating to give either the lowest NF or the highest output SNR. One way of improving the output SNR is to couple the input sensor to the amplifier through a low noise, low loss transformer, as shown in Figure 3.15. Such coupling, of course, presumes that the signal coming from the transducer is ac and not dc, for obvious reasons. A practical transformer is a bandpass device, but it loses efficiency at low and high frequencies, limiting the range of frequencies over which output SNR can be maximized. The output MS SNR can

**FIGURE 3.15**

Use of an ideal input transformer to improve the output SNR of an amplifier when R_s is small but finite.

be calculated for the circuit of Figure 3.15 as follows. The MS input signal is simply $\overline{v_s^2}$. In the case of a sinusoidal input, it is well known that $\overline{v_s^2} = v_s^2/2$ MSV. The MS signal at the output is:

$$S_o = \overline{v_s^2} n^2 K_V^2 \quad (3.52)$$

where n is the transformer's *secondary-to-primary turns ratio* and K_V is the amplifier mid-band gain. The transformer is assumed to be ideal (and noiseless). In practice, transformer windings have finite resistance, hence they make Johnson noise and their magnetic cores contribute *Barkhausen noise* to their noisiness. The ideal transformer, besides having infinite frequency response, is lossless. From this latter assumption, it is easy to show that the amplifier 'sees' a transformed Thevenin equivalent circuit of the input transducer having an open circuit voltage of $nv_s(t)$, and a Thevenin resistance of $n^2 R_s$. Thus the MS output noise can be written as:

$$N_o = [n^2 4kTR_s + e_{na}^2 + i_{na}^2(n^2 R_s)^2] K_V^2 B \text{ MSV} \quad (3.53)$$

And the output SNR is:

$$\text{SNR}_o = \frac{\overline{v_s^2}/B}{4kTR_s + e_{na}^2/n^2 + i_{na}^2 n^2 R_s^2} \quad (3.54)$$

The SNR_o clearly has a maximum with respect to n . If the denominator of equation 3.54 is differentiated with respect to n^2 and set equal to zero, then an optimum turns ratio, n_o , exists and is given by:

$$n_o = \sqrt{e_{na}/(i_{na} R_s)} \quad (3.55)$$

If the noiseless (ideal) transformer is given the turns ratio n_o , then it is easy to show that the maximum output SNR is given by:

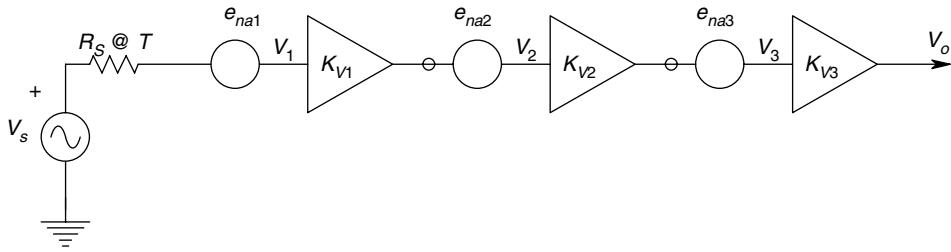
$$\text{SNR}_{o\text{MAX}} = \frac{\overline{v_s^2}/B}{4kTR_s + 2e_{na}i_{na}R_s} \quad (3.56)$$

The general effect of transformer SNR maximization on the system's noise figure contours is to shift the locus of minimum NF_{spot} to a lower range of R_s , although there is no obvious shift along the f_c axis. Also, the minimum NF_{spot} is higher with the transformer. This is because a practical transformer is noisy, as discussed above (Figure 3.13B).

As a rule of thumb, use of a transformer to improve output SNR and reduce NF_{spot} is justified if $(e_{na}^2 + i_{na}^2 R_s^2) > 20$ and e_{na} , i_{na} , R_s are in the range of frequencies of interest.

3.7 Cascaded Noisy Amplifiers

In this section, we examine a rule for the design of low noise signal conditioning systems. Figure 3.16 illustrates three cascaded noisy amplifiers. The i_{na} terms are assumed to be negligible because $i_{na}^2 R_s^2 \ll e_{na}^2$ in each amplifier. We may write the MS

**FIGURE 3.16**

Three cascaded stages of noisy voltage amplifiers.

output signal as:

$$S_o = v_s^2 K_{V(1)}^2 K_{V(2)}^2 K_{V(3)}^2 \text{ MS V} \quad (3.57)$$

The MS output noise can be written as:

$$N_o = (4kTR_s + e_{na1}^2)K_{V(1)}^2 K_{V(2)}^2 K_{V(3)}^2 B + e_{na2}^2 K_{V(2)}^2 K_{V(3)}^2 B + e_{na3}^2 K_{V(3)}^2 B \quad (3.58)$$

The three-amplifier system's MS output SNR is thus:

$$\text{SNR}_o = \frac{v_s^2 B}{4kTR_s + e_{na1}^2 + e_{na2}^2 K_{V(1)}^2 + e_{na3}^2 K_{V(1)}^2 K_{V(2)}^2} \quad (3.59)$$

Note that as long as $K_{V(1)} > 5$, the output SNR is set essentially by the noise characteristics of the input amplifier (headstage, i.e. the two right hand terms in the denominator of the SNR_o expression are not numerically significant).

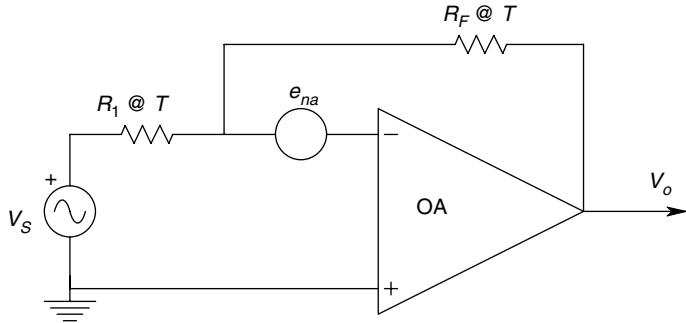
The point to remember from this section is that when it is necessary to design a low noise, signal conditioning system, the lowest noise amplifier in the headstage position should be used, after giving it a gain of at least five.

3.8 Calculation of the Noise Limited Resolution of Certain Signal Conditioning Systems

In designing instrumentation systems, it is often necessary to be able to predict the noise limited threshold resolvable measurand, or alternately, what input measurand level will produce a given system output SNR. In this section as well as in the problems at the end of this chapter, we present some examples of situations in which a noise limited, minimum input signal level is found.

3.8.1 Calculation of the Minimum Resolvable ac Input Voltage to a Noisy, Inverting Op-amp Amplifier

Figure 3.17 illustrates a simple inverting op-amp circuit. The noisy op-amp is modelled as a frequency compensated op-amp with a white, equivalent, short circuit input noise, e_{na} . The i_{na} is not included in the model because $i_{na}R_F \ll e_{na}$. We assume the input signal

**FIGURE 3.17**

A noisy inverting op-amp amplifier.

to be a sinusoidal voltage with peak value V_s and frequency f_s . The MS output signal voltage is:

$$S_o = (V_s^2/2)(R_F/R_1)^2 \text{ MS V} \quad (3.60)$$

The total mean squared output noise is:

$$N_o = 4kTR_1(R_F/R_1)^2B + 4kTR_F B + e_{na}^2(1 + R_F/R_1)^2B \text{ MS V} \quad (3.61)$$

The equivalent Hz noise bandwidth, B , may be calculated as follows. A frequency compensated op-amp operated as a simple inverting amplifier has been shown [Northrop, 1990] to have a frequency response given by:

$$\frac{V_o}{V_s} = A_V(j\omega) = \frac{-A_{vo}R_F/[R_F + R_1(1 + A_{vo})]}{j\omega[\tau_a(R_F + R_1)]/[R_F + R_1(1 + A_{vo})] + 1} \quad (3.62)$$

In this expression, A_{vo} is the open-loop op-amp's dc gain and τ_a is its open-loop time constant. τ_a may be expressed in terms of the op-amp's small signal, unity gain frequency, f_T :

$$\tau_a = A_{vo}/(2\pi f_T) \text{ s} \quad (3.63)$$

Hence the closed-loop, inverting amplifier has a time constant which can be approximately expressed by:

$$\tau_{CL} = \frac{(1 + R_F/R_1)}{2\pi f_T} \text{ s} \quad (3.64)$$

Thus the equivalent Hz noise bandwidth for the amplifier of Figure 3.17 may be found using Table 3.1, which gives:

$$B = 1/(4\tau_{CL}) = \frac{\pi f_T}{2(1 + R_F/R_1)} \text{ Hz} \quad (3.65)$$

Thus the output MS SNR can be written as:

$$\text{SNR}_o = \frac{(V_s^2/2)(R_F/R_1)^2[2(1 + R_F/R_1)]}{[4kTR_1(R_F/R_1)^2 + 4kTR_F + e_{na}^2(1 + R_F/R_1)^2]\pi f_T} \quad (3.66)$$

In this SNR calculation, we assume that the frequency of the input signal, f_s , is much less than the break frequency of the closed-loop amplifier, $1/(2\pi\tau_{CL})$ Hz, so that there is no attenuation of the output signal due to the low-pass filter characteristic of the amplifier.

Using equation 3.66 above, we now calculate the peak value of V_s that will yield a MS SNR of unity, given the following typical system parameters— $R_F = 10^5$, $R_1 = 10^3$, $f_T = 12$ MHz, $4kT = 1.66 \times 10^{-20}$ and $e_{na} = 10$ nV RMS/ $\sqrt{\text{Hz}}$. Calculations yield $V_s = 6.65$ μV , peak. This input value is relatively high because of the broadband nature of the amplifier's output noise ($B = 5.94 \times 10^4$ Hz).

3.8.2 Calculation of the Minimum Resolvable dc Current in White and 1/f Noise

In this example, a very small DC current is to be measured using an electrometer op-amp transresistor circuit, shown in Figure 3.18. The dc current source is represented by a Norton equivalent circuit, having current source, I_s , and conductance, G_s . The thermal noise from G_s , R_F and i_{na} are assumed to have white spectrums. The e_{na} , however, is assumed to have a 1/f component:

$$e_{na}^2(f) = e_{na0}^2 + (b/f) \text{ MS V/Hz} \quad (3.67)$$

The PDS of the amplifier's output noise can be written as:

$$S_{no}(f) = (e_{na0}^2 + b/f)(1 + R_F G_s)^2 + [1kT(G_F + G_s) + i_{na}^2] R_F^2 \quad (3.68)$$

This noise is, of course, broadband and is conditioned by a noiseless, unity gain, quadratic low-pass filter with transfer function:

$$F(s) = \frac{1}{[s^2/\omega_n^2] + [s(2\xi)/\omega_n] + 1} \quad (3.69)$$

Reference to item 3 in Table 3.1 shows that the quadratic low-pass filter has an equivalent noise bandwidth of $B = \omega_n/(8\xi)$ Hz. The total MS output noise voltage can thus be found by integrating the white noise terms of $S_{no}(f)$ from 0 to $\omega_n/(8\xi)$. Integration of the b/f term between the same limits leads to the perplexing result of infinity. To avoid this problem, we must impose a practical low frequency limit for the b/f integral. One arbitrary but effective solution to this problem is to take the low frequency limit to be the reciprocal of the time it takes to confidently read the output voltage from the low-pass

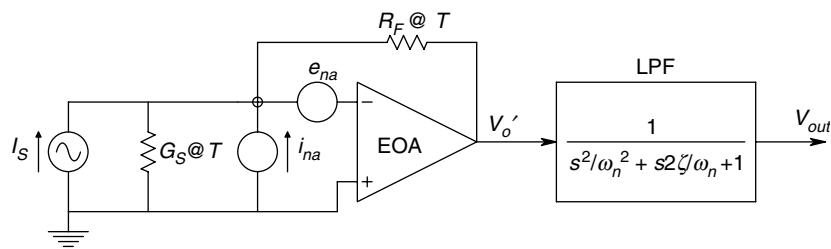


FIGURE 3.18

A noisy operational transresistor circuit followed by a noiseless LPF used to condition small currents.

filter with an analog or digital voltmeter. Let us assume that this takes 4 s, so the low frequency limit is 0.25 Hz. Therefore:

$$N_o = \{(e_{na0}^2)(1 + R_F G_s)^2 + [4kT(G_F + G_s) + i_{na}^2]R_F^2\}[\omega_n/(8\xi)] + b\ln\left(\frac{4\omega_n}{8\xi}\right)(1 + R_F G_s)^2 \text{ MS V} \quad (3.70)$$

The MS dc output signal is $S_o = I_s^2 R_F^2 \text{ MS V}$.

Now let us find the I_s in order to give an output MS SNR of 4. We assume the following parameters for an AD549 electrometer op-amp— $4kT = 1.66 \times 10^{-20}$ at 300 K, $\omega_n/(8\xi) = 2 \text{ Hz}$, $R_F = 10^{10} \Omega$, $G_s = 10^{-7} \text{ S}$, $e_{na0} = 35 \text{ nV RMS}/\sqrt{\text{Hz}}$, $i_{na} = 0.2 \text{ fA RMS}/\sqrt{\text{Hz}}$, $b = 1.36 \times 10^{-17} \text{ MS V}$. Evaluating the noise terms in equation 3.70, we find that the dominant term comes from the Johnson noise in R_F and R_s , which is $3.3233 \times 10^{-7} \text{ MS V}$. The MS output noise from the b/f term in e_{na}^2 is $2.8336 \times 10^{-11} \text{ MS V}$, and the total MS output noise from e_{na0}^2 and i_{na}^2 is $8.2455 \times 10^{-12} \text{ MS V}$. When we set the MS output SNR equal to four, we find that the minimum $I_s = 0.1153 \mu\text{A}$ dc and the dc voltage at the system output is 1.153 mV.

3.8.3 Calculation of the Minimum Resolvable ac Input Signal to Obtain a Specified Output SNR in a Transformer Coupled, Tuned Amplifier

Figure 3.19 illustrates a signal conditioning system used to amplify the 10 kHz, sinusoidal voltage output of a low impedance sensor. A transformer (assumed here to be ideal and noiseless) is used to maximize the output SNR. A noiseless, quadratic bandpass filter, tuned to 10 kHz, follows the low noise op-amp input stage. The peak gain of the tuned filter is unity. Looking toward the source through the ideal transformer, the op-amp summing junction sees a Thevenin equivalent circuit, consisting of an open circuit voltage, $V_{oc}(t) = nV_s(t)$ and an equivalent resistance of $R_{TH} = n^2 R_s$. The MS output

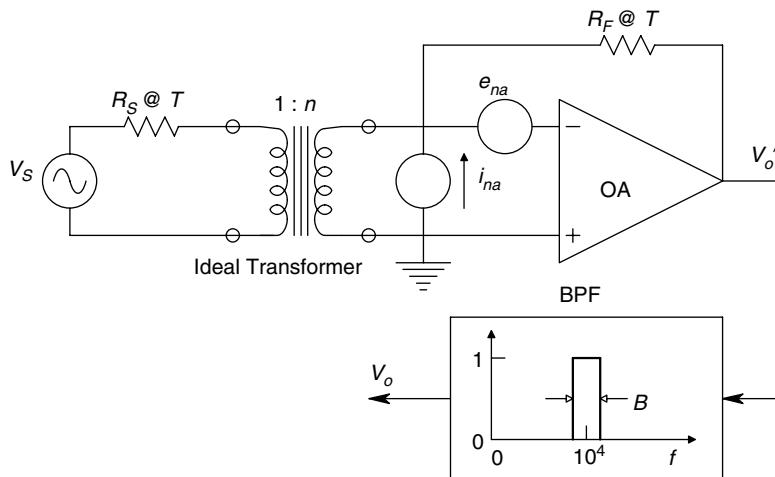


FIGURE 3.19

Use of a transformer to improve the output SNR when conditioning a low level sinusoidal voltage from a low resistance source.

voltage is thus:

$$S_o = (V_s^2/2)[R_F^2/(nR_s)^2] \text{ MS V} \quad (3.71)$$

All sources of noise in the circuit are assumed to be white. The noise bandwidth is determined by the BPF and is found from Table 3.1 to be $B = \omega_n\zeta = \omega_n/2Q$ Hz. The total MS output noise can thus be written as:

$$N_o = [4kTR_s R_F^2/(nR_s)^2 + e_{na}^2(1 + R_F/(n^2 R_s))^2 + i_{na}^2 R_F^2 + 4kTR_F](\omega_n/2Q) \text{ MS V} \quad (3.72)$$

Therefore, the MS output SNR is:

$$\text{SNR}_o = \frac{(V_s^2/2)/B}{4kTR_s + e_{na}^2((nR_s/R_F) + 1/n)^2 + i_{na}^2 R_s^2 n^2 + 4kTG_F R_s^2 n^2} \quad (3.73)$$

The denominator of equation 3.73 has a minimum for some non-negative n_o , hence the SNR has a maximum for that n_o . To find n_o , we differentiate the denominator with respect to n^2 , set the derivative equal to zero and then solve for n_o . This yields:

$$n_o = \frac{\sqrt{e_{na}}}{\sqrt{R_s}(e_{na}^2/R_F^2 + 4kTG_F + i_{na}^2)^{1/4}} \quad (3.74)$$

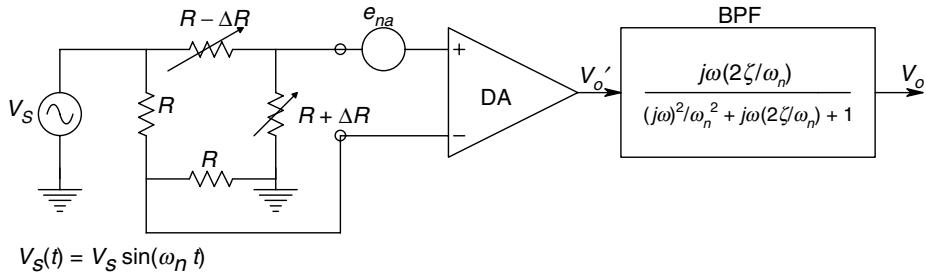
We now examine a numerical solution for n_o and V_s to give a SNR_o equal to unity. The following parameters are used— $e_{na} = 3 \text{ nV RMS}/\sqrt{\text{Hz}}$, $i_{na} = 0.4 \text{ pA RMS}/\sqrt{\text{Hz}}$, $R_F = 10^4 \Omega$, $R_s = 10 \Omega$, the BPF $Q = 100$ and $4kT = 1.66 \times 10^{-20}$. The n_o is found to be 14.73. When n_o is substituted for n in equation 3.73 and a numerical solution obtained, the V_s needed to get unity SNR is 13 peak nV. If the transformer is not used ($n = 1$), then the V_s needed to give $\text{SNR} = 1$ is 76.0 peak nV.

3.8.4 Calculation of the Smallest $\Delta R/R$ in a Wheatstone Bridge to Give a Specified SNR_o

In this example, we calculate the smallest ΔR required to produce MS output SNR of 10 in the ac-powered, Wheatstone bridge circuit, shown in Figure 3.20. Two arms of the bridge are variable, one having increasing resistance while the other's resistance decreases a corresponding amount. Such a configuration is found in certain unbonded strain gauges used to measure forces. The output of the bridge is amplified by a PMI AMP-01 low noise, differential instrumentation amplifier, which is followed by a noiseless, quadratic bandpass filter. The purpose of the bandpass filter is to pass the amplified ac output of the bridge while restricting the bandwidth of the noise, thus achieving an output SNR which is higher than that at the output of the differential amplifier.

The signal output of the system is a sine wave whose amplitude and phase is determined by V_s and $\Delta R/R$, which may be found as follows. Let the bridge excitation be $V_s(t) = V_s \sin(2\pi 400t)$. The voltage at the left hand corner of the bridge, V'_1 , is easily seen to be $V_s(t)/2$. The voltage at the right hand corner of the bridge, V_1 , can be written as:

$$v_1(t) = [v_s(t)/2](1 + \Delta R/R) \quad (3.75)$$

**FIGURE 3.20**

Use of a quadratic BPF to improve the output SNR and resolution of a Wheatstone bridge with ac excitation.

Hence the output of the differential amplifier is:

$$v'_o(t) = K_d(\Delta R/2R)V_s \sin(2\pi 400t) \quad (3.76)$$

The center frequency of the bandpass filter is 400 Hz, so $V_o(t) = V'_o(t)$. Thus the MS signal output is:

$$S_o = (V_s^2/2)(K_d^2/4)(\Delta R/R)^2 \text{ MS V} \quad (3.77)$$

The MS noise at the system output is assumed to be due to the white thermal noise in the bridge resistors and the equivalent input short circuit voltage noise, e_{na} , which is assumed to be white. The total MS output noise is found by adding up the MS noise contributions from all the noise sources. The equivalent Hz noise bandwidth is set by the bandpass filter, which from Table 3.1, is $(2\pi f_n/2Q)$ Hz. Thus the MS output noise voltage can be written as:

$$N_o = (4kTR/2 + 4kTR/2 + e_{na}^2)K_d^2(2\pi f_n/2Q) \text{ MS V} \quad (3.78)$$

The MS SNR at the output of the system is thus:

$$\text{SNR}_o = \frac{(V_s^2/2)(K_d^2/4)(\Delta R/R)^2}{(4kTR + e_{na}^2)K_d^2(2\pi f_n/2Q)} \quad (3.79)$$

We now calculate the $\Delta R/R$ required to give a MS output SNR of 10. Let $f_n = 400$ Hz, $Q = 5$, $R = 1000 \Omega$, $K = 1000$, $4kT = 1.66 \times 10^{-20}$, $e_{na} = 5 \text{ nV RMS}/\sqrt{\text{Hz}}$, $V_s = 4 \text{ V peak}$. Using equation 3.79 above, we find that $\Delta R/R = 7.23 \times 10^{-8}$ or $\Delta R = 7.23 \times 10^{-5} \Omega$. Using equation 3.76, we calculate that this ΔR will produce a 400 Hz sinusoidal output with peak value $V'_o = 1.446 \times 10^{-4} \text{ V}$.

As can be seen, ac operation of a Wheatstone bridge with a tuned output filter can result in considerable sensitivity. Normally, the slow changes in $\Delta R(t)$ would be extracted with a phase sensitive rectifier following the BPF.

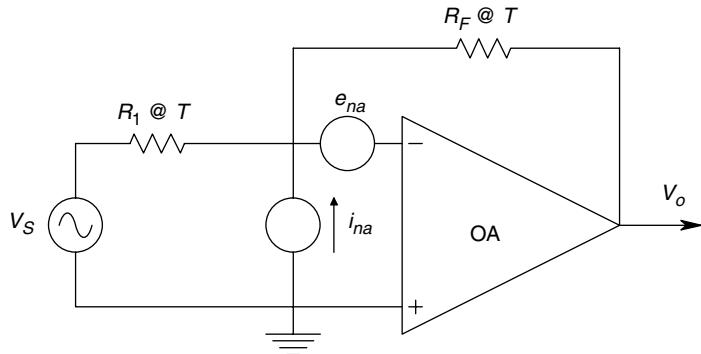


FIGURE 3.21
A simple, noisy, inverting op-amp amplifier.

3.8.5 Determination of the Conditions for Maximum Output SNR Given a Simple Inverting Op-Amp Amplifier with Known e_{na} and i_{na}

Figure 3.21 shows a simple op-amp inverter circuit having a signal source that lies well within the Hz noise bandwidth, B , for the circuit. Both R_1 and R_F are assumed to make white thermal noise. The MS signal at the output can be written as:

$$S_o = \overline{v_s^2} (R_F/R_1)^2 \text{ MSV} \quad (3.80)$$

The total MS noise at the output is:

$$N_o = [4kT R_1 (R_F/R_1)^2 + (i_{na}^2 + 4kT G_F) R_F^2 + e_{na}^2 2(1 + R_F/R_1)^2] B \text{ MSV} \quad (3.81)$$

After some algebra, we can write the MS output SNR as:

$$\text{SNR}_o = \frac{\overline{v_s^2}/B}{4kT R_1 (1 + R_1/R_F) + e_{na}^2 (1 + R_1/R_F)^2 + i_{na}^2 R_1^2} \quad (3.82)$$

Clearly, from equation 3.82, we see that the SNR is larger for high ratios of R_F/R_1 (i.e. for high signal gains). Also, the SNR_o is larger for small R_1 . Small R_1 , of course, means that V_s sees a low effective input resistance (i.e. R_1). A practical limit on the gain is set by the necessary closed-loop bandwidth, which will be approximately $f_T(R_1/R_F)$ Hz. Good low noise design often involves compromise of other design parameters.

3.9 Modern, Low Noise Amplifiers for Use in Instrumentation Signal Conditioning Systems

In the past ten years, as the result of advances which have been made in transistor fabrication technology and in electronic circuit design, there has been a rise in the number of available, low noise, IC amplifiers suitable for use in signal conditioning in measurement systems. Low noise, as used in this section, shall mean amplifiers with e_{nA} s less than $10 \text{ nV RMS}/\sqrt{\text{Hz}}$. It is always risky to attempt to list and categorize low

noise amplifiers in a text for the following reasons:

1. The danger of omitting one or more of the more obscure models
2. By the time the list is read, new, better low noise amplifiers will have been put on the market.

Despite facing these risks, we list below certain low noise operational amplifiers and instrumentation amplifiers which may have application in the design of low noise signal conditioning systems.

In specifying the noise performance of IC amplifiers, manufacturers typically give the following information:

1. e_{na} measured in the *flat* portion of its spectrum
2. The peak-to-peak or RMS output noise measured under short circuited input conditions over a specified, low frequency portion ($1/f$ region) of the $e_{na}^2(f)$ spectrum
3. The value of i_{na} measured in its flat region. For some amplifiers, the manufacturers also give a plot of the spot noise figure contours *vs* R_s and f

TABLE 3.2

Partial listing of commercially available, low noise op-amps and instrumentation amplifiers having low noise characteristics

Amp. Model	Type	e_{na}	i_{na}	Low Freq V_n	R_{in}	f_T	A_{V0}
PMI OP-27	OP/BJT	3	0.4 pA	0.08 μ Vppk	6 M	5	10^6
PMI OP-61A	OP/BJT	3.4	0.8 pA	—	—	200	4×10^5
HA 5147A	OP/BJT	3	0.4 pA	0.08 μ Vppk	6 M	140	10^6
BB OPA111BM	OP/FET	6	0.4 fA	1.6 μ Vppk	10^{13} (DM) 10^{14} (CM)	2	125 dB
LT1028	OP/BJT	0.85	1 pA	35 nVppk	20 M (DM) 300 M (CM)	75	7×10^6
AM-427	OP/BJT	3	0.4 pA	0.18 μ Vppk	1.5 M (DM)	5	120 dB
MAX410/12/14	OP/BJT	2.4	1.2 pA	—	20 k (DM), 40 k (CM)	28	122 dB
LT1115	OP/BJT	0.9	1.2 pA	—	5 k (DM), 250 M (CM)	70	15×10^6
AD8022	OP/BJT	2.5	1.2 pA	—	20 k (DM)	130	72 dB
MAX4106	OP/BJT	0.75	2.5 pA	—	1 M (CM)	350	100 dB
PMI AMP-011	IA/BJT	5	0.15 pA	0.12 μ Vppk	10 G (DM), 20 G (CM)	26	$1-10^3$
PMI AMP-021	IA/FET	9	0.4 fA	0.4 μ Vppk	10 G (DM), 16.5 G (CM)	5	$1-10^3$
AD6251	IA/BJ	4	0.3 pA	0.2 μ V	1 G (DM&CM)	25	$1-10^3$
BB INA1102	IA/FET	10	1.8 fA	1 μ Vppk	2×10^{12} (DM) 5×10^{12} (CM)	12	1-500
PMI AMP-052	IA/FET	16	10 fA	4 μ Vpp	10^{12} (DM&CM)	3	$1-10^3$
ZN459CP	IA/BJT	0.8	1 pA	—	7 k (Single-ended)	15(-3)	60 dB
ZN424	IA/BJT	5.5	0.3 pA	—	200 k	4	2×10^4

TYPE refers to whether the amplifier is an op-amp (OP) or an instrumentation amplifier (IA), and whether it has a BJT or FET headstage. e_{na} is given in nV RMS/ $\sqrt{\text{Hz}}$ measured at 1 kHz, i_{na} is given either in pA or fA RMS/ $\sqrt{\text{Hz}}$ measured at 1 kHz. Low Freq V_n is the equivalent, peak-to-peak, short circuit input noise measured over a standard 0.1 Hz–10 Hz bandwidth. R_{in} is the input resistance, f_T is the unity gain bandwidth in MHz, unless followed by (-3), in which case, it is the high frequency -3 dB frequency. A_{V0} for op-amps is their open-loop, dc gain; the useful gain range is given for IAs.

The noise performance of an IC amplifier is largely set by the designer's choice of headstage transistors. Consequently, low noise amplifiers may be broadly characterized by whether they use low noise BJTs or FETs in their input circuits. Those using FETs generally have high input resistances (over $10^{12} \Omega$) and low i_{nA} ($< 1 \text{ fA RMS}/\sqrt{\text{Hz}}$). Amplifiers with BJT headstages typically have lower input resistances and higher i_{nA} , but in some cases, exhibit e_{nA} an order of magnitude lower than the e_{nA} s from JFET amplifiers. Table 3.2 is a partial listing of the more widely used, low noise, IC amplifiers. It is subdivided into instrumentation amps and op-amps.

3.10 Coherent Interference and its Minimization

In this section, we consider another common form of unwanted input which degrades the resolution and accuracy of measurement systems is considered (i.e. periodic or coherent interference). Coherent interference, unlike random noise, has narrow band power density spectrums, often with harmonic peaks at integral multiples of the fundamental frequency. A serious problem is created when the coherent interference spectrum has power in the signal frequency band. In the following sections, we will discuss the common sources of coherent interference and various means of minimizing its input to a measurement system. The techniques of minimizing coherent interference often appear arcane to the inexperienced, but ultimately, all such techniques follow the laws of physics and circuit theory.

3.10.1 Sources of Coherent Interference

Broadly speaking, there are four major sources of coherent interference:

1. Power line frequency interference, often coupled with the measurement system by an electric field and/or a magnetic field. Obvious sources of power line interference are unshielded power transformers, poorly filtered ac-to-dc power supplies and electromagnetic field pick-up from unshielded, high voltage 60 Hz power lines.
2. Radio frequency electromagnetic interference (RFI) coupled with the measurement system from local radio transmitters, poorly shielded computers and other digital instruments having MHz clock frequencies. RFI may be picked up through the ac power line, by magnetic induction in internal wiring loops, or by magnetic induction in the cables connecting sensors to signal conditioning systems.
3. Of a more transient nature, the periodic noise may be coupled with a measurement system from sparking motor brushes, gasoline engine ignition systems, or high frequency power line transients caused by SCR and triac switching in motor speed controls and ovens. Although hardly periodic, the broadband electromagnetic pulses produced by lightning in thunderstorms are often picked up by EMI-sensitive electronic signal conditioning systems.
4. In a complex electronic measurement system, coherent interference often may arise from within the system. Such interference may be the result of poor grounding and/or shielding practices, allowing the digital portion of the system to 'talk' to the low level analog front end of the system. Internal coherent interference

may also arise through poor decoupling of a common power source which supplies a low level analog front end and a high frequency digital circuit.

3.10.1.1 Direct Electrostatic Coupling of Coherent Interference

To illustrate this phenomenon, we assume that two conductors lie in close proximity. One conductor is at a sinusoidal potential, $v_1(t) = V_1 \sin(\omega t)$. There is a capacitance C_{12} between the two conductors and conductor 2, which picks up the interference from conductor 1, has a resistance R_2 in parallel with a capacitance C_2 to ground, as shown in Figure 3.22A. The discrete component circuit describing this simple capacitive coupling situation can be redrawn as a simple voltage divider, shown in Figure 3.22B. We may write a transfer function for the coherent interference induced on conductor 2:

$$\frac{V_2(j\omega)}{V_1} = \frac{j\omega C_{12} R_2}{j\omega(C_{12} + C_2)R_2 + 1} \quad (3.83)$$

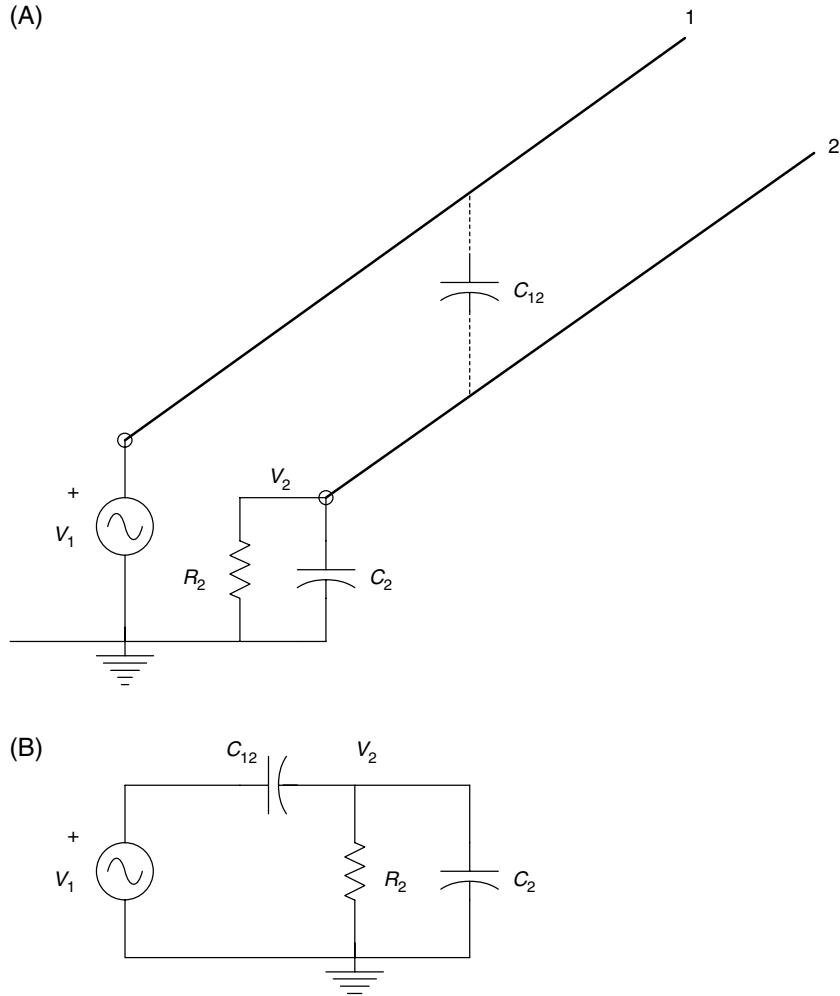


FIGURE 3.22

A. A circuit describing the electrostatic coupling between two adjacent conductors. A sinusoidal voltage source is assumed. B. Lumped parameter equivalent circuit of Figure 3.22A.

By inspection, we see that this is a simple, high-pass transfer function. C_{12} is typically of the order of single pF, C_2 may be in the tens of pF and R_2 may vary widely, depending on the application. For a specific example, let $C_{12}=2.5\text{ pF}$, $C_2=27.5\text{ pF}$ and $R_2=50\Omega$. The break frequency is found to be $6.67 \times 10^8\text{ r/s}$. Hence at frequencies below the break frequency,

$$V_2 = j\omega C_{12} R_2 V_1 \quad (3.84)$$

From the equations above, we gather the lesson that in order to minimize V_2 , the V_1 , C_{12} and R_2 should be made small. C_{12} is minimized by arranging the wiring geometry so that high level or output signal wires, or printed circuit lines, are situated at sufficient distance from sensitive, high impedance, low level circuit wiring. Insulated wire crossings, if necessary, should be at right angles. Also, sensitive wires should be shielded, if possible. Resistance R_2 to ground should be as low as possible. Often the sensor output resistance is several kΩ, so the best protection for the headstage is to shield it as well as the wire connecting it to the sensor. As can be seen below, what would appear to be a simple solution to the capacitive pick-up of coherent interference by coaxial shielding is not that simple—details of shield grounding determine the effectiveness of the shield.

3.10.1.2 Direct Magnetic Induction of Coherent Interference

Figure 3.23A illustrates the basic mechanism whereby a time-varying magnetic field, caused by current in conductor 1, induces an EMF in a second conductor. To model this phenomenon with a lumped parameter circuit, we assume that a mutual inductance, M , exists between the two conductors and that the conductors can be modelled by a transformer, as shown in Figure 3.23B. The EMF induced in wire 2 is:

$$V_2 = j\omega M I_1 \quad (3.85)$$

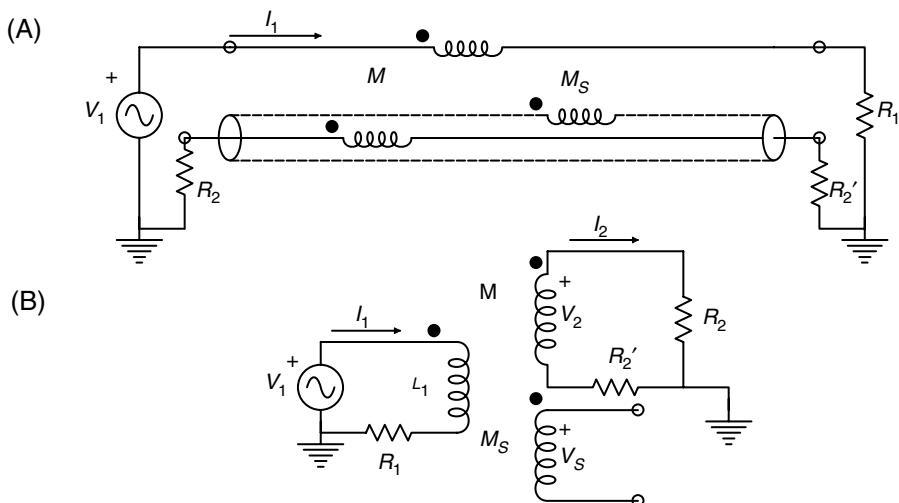
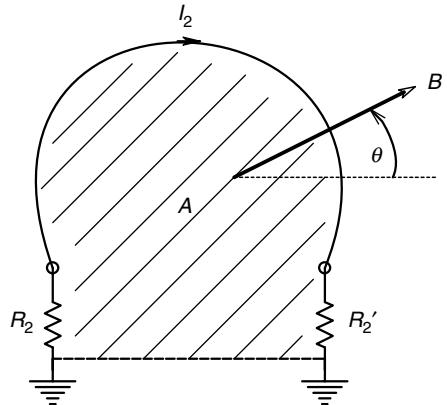


FIGURE 3.23

A. A circuit describing magnetic induction of coherent interference. The shield of the coaxial cable is not grounded. B. Lumped parameter equivalent circuit of Figure 3.23A.

**FIGURE 3.24**

Circuit illustrating the induction of a coherent interference current, I_2 , in a conductive loop enclosing a current carrying conductor.

Thus, the current induced in loop 2 is given by Ohm's law to be:

$$\frac{I_2}{I_1}(j\omega) = \frac{j\omega M/(R_2 + R'_2)}{[1 + j\omega L_2/(R_2 + R'_2)]} \quad (3.86)$$

and the voltage observed across R_2 is $I_2 R_2$. L_2 is the very small self-inductance of wire 2. The break frequency of the pole in equation 3.86 is generally quite large, so for most cases,

$$I_2 = I_1[j\omega M/(R_2 + R'_2)] \quad (3.87)$$

In some cases of induced magnetic interference, a time-varying B field exists in space in the vicinity of a loop of area A , as shown in Figure 3.24. The B field can arise from an unshielded power transformer, or at high frequencies, or can be the result of a strong, radio frequency electromagnetic field. By Faraday's Law, the EMF induced in the loop can be written as the integral of the dot product:

$$E = -\frac{d}{dt} \int B ds \quad (3.88)$$

For the case of a B field which varies sinusoidally in time and which intersects the plane of the loop at an angle θ , we may write:

$$E = j\omega B A \cos \theta \quad (3.89)$$

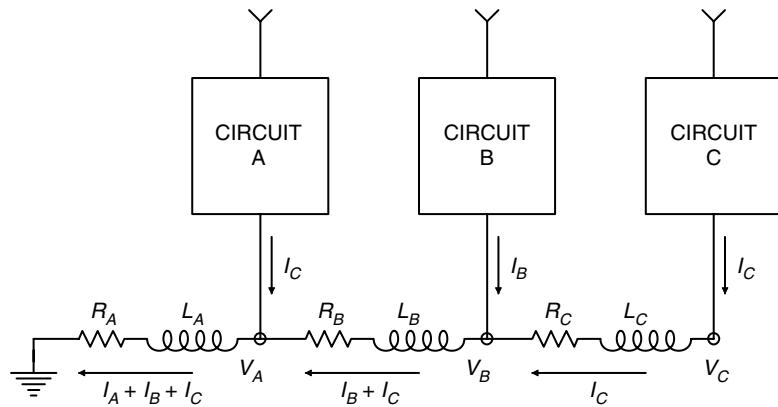
and the interference current induced in the loop is:

$$I_2 = \frac{j\omega B A \cos \theta}{(R_2 + R'_2)} \quad (3.90)$$

Here, we have neglected the inductance in the loop. Clearly, steps taken to reduce B and A will reduce the pick-up of magnetic coherent interference.

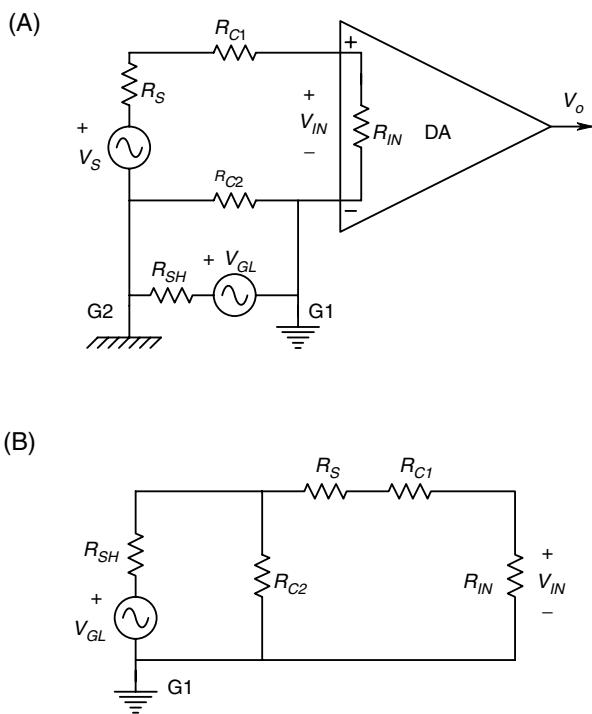
3.10.1.3 Ground Loops

Whenever a ground conduction path is forced to carry a power supply return current which has a significant high frequency component added to the dc, and this ground path

**FIGURE 3.25**

Circuit illustrating how ground current of one circuit can affect the ground potential of other circuits sharing the ground line. This condition is called a ground loop.

is also used in common for a low level signal ground, it is possible for the ac power component of ground current to develop a small voltage across the small but finite impedance of the ground path. This situation is illustrated in Figure 3.25. There are, of course, causes other than power supply currents which cannot make the ground potentials of two subsystems isopotential. The fact that two ground points are not at exactly zero potential means that an unwanted ground loop current can flow in a coaxial cable shield which is grounded at each end to a different chassis or enclosure. The ground loop current in the shield can induce coherent interference on the center conductor. Figure 3.26 illustrates how a difference in ground potential can add a coherent noise

**FIGURE 3.26**

A. Circuit showing how a ground loop voltage can affect the input to an amplifier. B. Equivalent circuit illustrating the two voltage dividers acting on the ground loop voltage.

voltage directly to the desired signal. Here, it is assumed that $R_{IN} \gg R_S \gg R_{C1}, R_{C2}, R_{SH}$. The coherent interference voltage from $V_{GL(OC)}$ appears at the input of the amplifier as V_N and may be shown to be equal to:

$$V_N = V_{GL(OC)} \frac{R_{C2}}{R_{C2} + R} \frac{R_{IN}}{R_{IN} + R_S} \quad (3.91)$$

The numerical value of the product of the two voltage divider terms depends mostly on the values of R_{C2} and R_{SH} , and is typically about 1/5. Thus, we see that an unwanted voltage, $0.2V_{GL(OC)}$, appears in series with the desired signal, V_S , at the input of the amplifier as a result of a ground loop and grounding the shield of the coaxial cable at both ends.

3.10.2 Cures for Coherent Interference

As described above, there are many sources of coherent interference and the means to reduce a given type of interference depends on the physical mechanism(s) by which it is coupled with the measurement system. Comprehensive treatments of the causes of, and means for reducing, coherent interference may be found in texts by Ott (1976) and Barnes (1987).

3.10.2.1 Power Line Low Pass Filters

In many cases, the cure for coherent interference lies in preventing it from escaping from a coherent noise source, such as a computer or digital system, an electronic system with a switching power supply, or a radio transmitter. Obviously, in the latter case, there must be electromagnetic energy radiated from an antenna. However, great care must be taken in all the systems mentioned above to keep the coherent interference from entering the power mains, where it can affect other line powered instruments. Power line filters are generally multistage, LC low-pass filters which attenuate frequencies in the range of about 10 kHz–40 MHz. Maximum attenuation at about 30 MHz ranges from –40 to –65 dB, depending on the number of LC stages in the filter. A schematic of a typical line filter is shown in Figure 3.27. Power line filters are effective in keeping high frequency coherent noise from entering a system on the power line, as well as keeping the internally generated noise from escaping from the system.

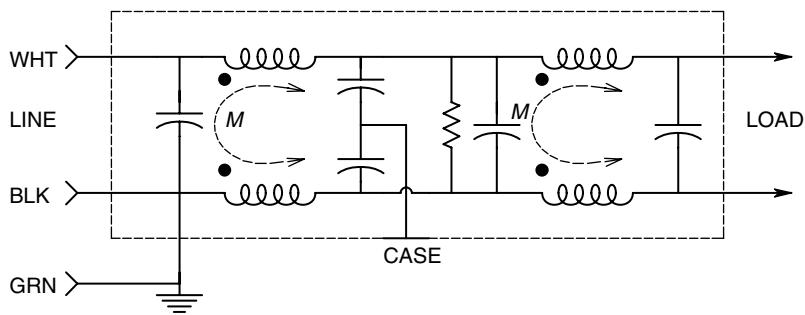


FIGURE 3.27

Schematic of a typical power line low-pass filter, designed to keep high frequency coherent interference on the power line out of a system and to prevent high frequency interference arising within the system from escaping on to the power line.

High frequency interference on power lines can be described in two forms—common-mode (CM) and difference-mode (DM) interference. In CM interference, the unwanted voltage has the same value on both the black and white wires with respect to ground (green wire).

The CM interference voltage can be written as:

$$V_{i(CM)} = (V_{iB} + V_{iW})/2 \quad (3.92)$$

The DM interference voltage is given by:

$$V_{i(DM)} = (V_{iB} - V_{iW})/2 \quad (3.93)$$

High frequency power line filters are designed to attenuate both $V_{i(CM)}$ and $V_{i(DM)}$. Power line filters must be used in conjunction with a robust, grounded metal instrument case—the metal case of the power line filter must have a solid, multipoint electrical connection with the grounded instrument case for best results.

One design trade-off in the application of power line low-pass filters is that the capacitors used to attenuate CM noise inject a current into the green wire power line ground. Generally, this current must be kept $<3.5\text{ mA}$ in non-medical applications. Hence, the size of the line-to-ground capacitors and thus, the amount of CM interference attenuation, is limited by this ground current limit.

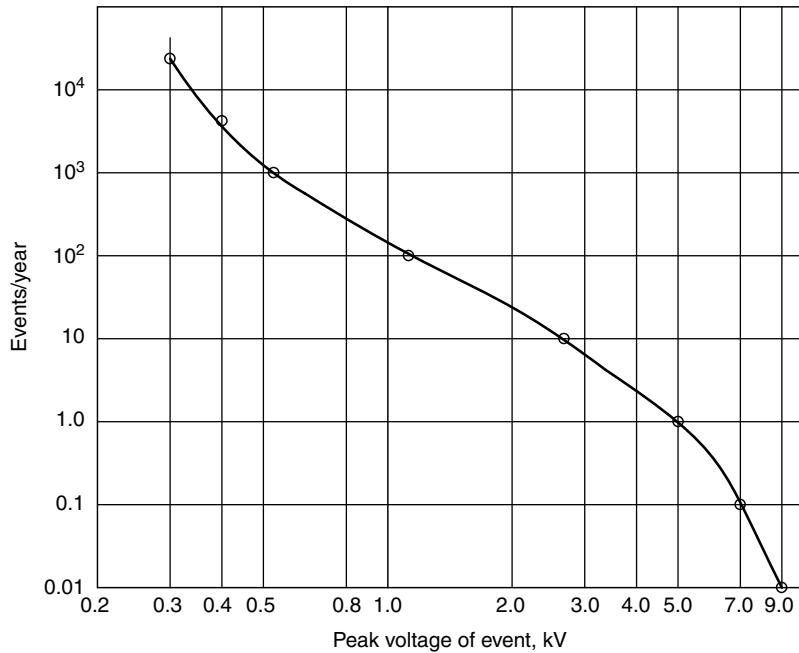
3.10.2.2 Transient Voltage Suppressors

Transient voltage suppressors are used to prevent the high voltage, spike-like transients occurring on the power mains input to an instrument system from causing physical damage to system power supply components and system circuits, or causing damage and anomalous results in associated computer equipment. There are many sources of power line spikes. Some spikes are periodic, occurring once or more every power frequency cycle. These spikes are generally of a biphasic nature and their voltages may be as much as 30–50V above the instantaneous line voltage. They are typically several μs in duration and arise from SCR or triac switching of inductive loads attached to the power line, such as in motor speed controls. Such ‘SCR spikes’ normally do not pose safety problems, but do constitute a major source of coherent interference, if allowed to get into the signal ground path or the power supply outputs. SCR spikes are of course, attenuated by line filters, but in some cases, may be large enough after filtering to pose a problem. Other power line spikes occur randomly as the result of lightning strikes on power lines, or from infrequent switching of inductive loads attached to the power line. Figure 3.28 illustrates the statistical occurrence of singular spikes on 120V, 60Hz power line in the US [Hey and Kram, 1978]. The large, singular, transient spikes are best reduced by the use of nonlinear circuit elements such as varistors or zener diodes. Varistors have steady state current-voltage curves, given by the relations:

$$I = G_{OFF}V \quad 0 < V < V_i \quad (3.94a)$$

$$I = KV^\alpha \quad V > V_i \quad (3.94b)$$

$$I = G_{ON}V \quad I > I_m \quad (3.94c)$$

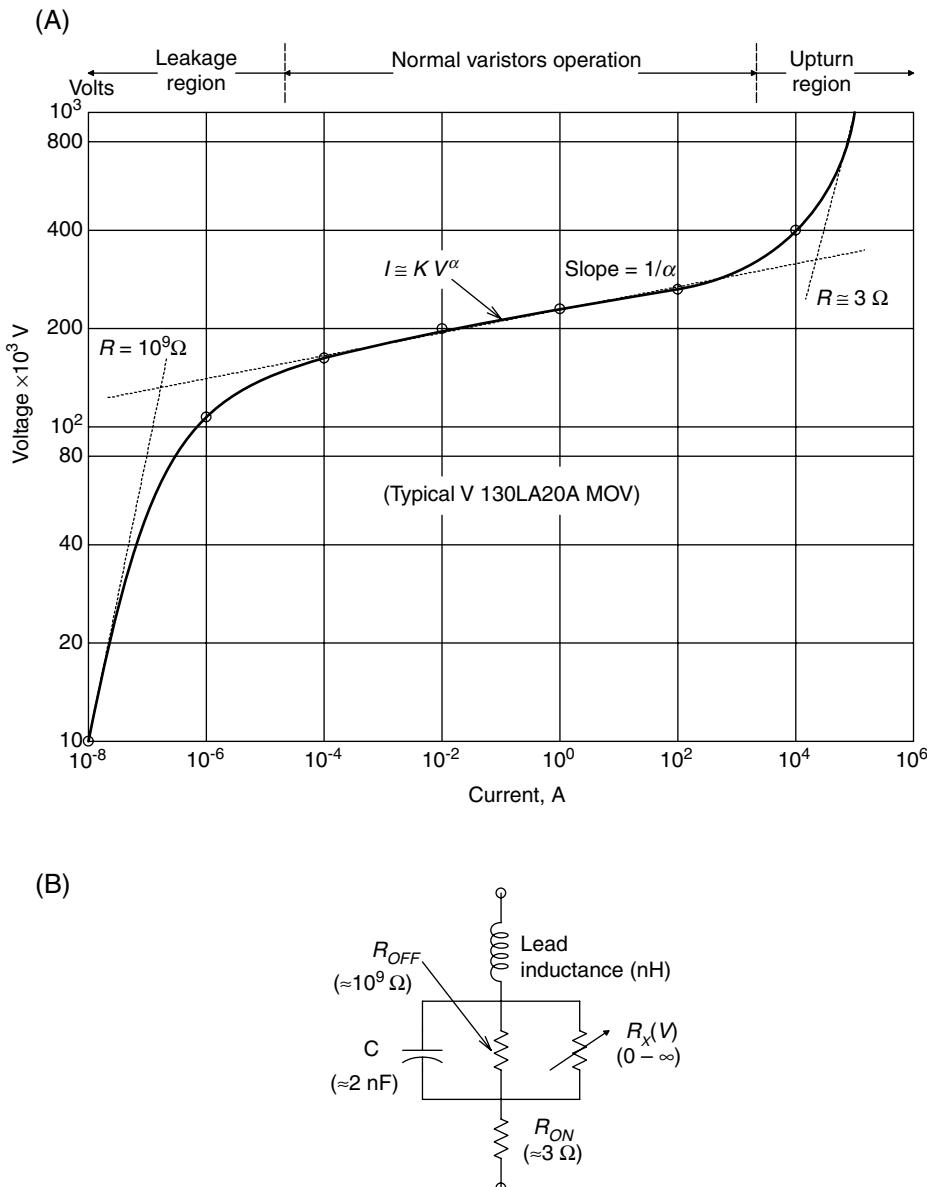
**FIGURE 3.28**

Statistical frequency of occurrence of spike transients on the 120 V, 60 Hz mains *vs* the spike peak voltage.
(Source: Adapted from Figure 1.7, *GE Transient Voltage Suppressor Manual*, 2nd ed., 1978.)

A typical varistor volt-ampere curve is shown in Figure 3.29A. An equivalent, lumped parameter circuit for a GE MOV is shown in Figure 3.29B. Note that in a linear (Ohm's Law) resistor, $\alpha = 1$. Depending on the application and construction of the varistors, α can range 2–30. For efficient transient suppression, preferably, α should be as large as possible. Metal oxide varistors were originally developed by Matsushita Electric Co., and are sintered, polycrystalline semiconductor devices. GE manufactures a family of zinc oxide, metal oxide varistors (MOVTM) under license from Matsushita. The α of GE MOVs are in the range 15–30, measured over several decades of surge current. GE MOVs are designed to operate at ac voltages of 10–1000 V, and can handle current surges of up to 25 kA, with energy capability of over 600 J in the larger units. Of primary consideration in the design of MOV transient suppression systems, is the ability of the MOV device to conduct the transient current that the source can supply under conditions of overvoltage. MOVs are best used in applications where the current rise times (from the power line) are longer than 0.5 μ s.

General Semiconductor Industries, Inc. manufactures the TransZorbTM transient voltage suppressor. This silicon, *pn* junction device acts similar to the MOV with the exception that its α is higher (35) and its response time is faster than MOV devices—the manufacturer claims 60 ps, *vs* one to 50 ns for varistors. The TransZorb has a relatively flat volt-ampere curve as compared to varistors (Figure 3.30). TransZorb AC operates in a voltage range of 5–340 peak volts; 5 volt devices clamp at 9.6 V, 340 volt devices clamp at 548 V.

Figure 3.31 shows a schematic for an extremely simple 6-outlet, power bar or line monitor power conditioner, typically used with personal computers and related equipment to protect them from line spike transients. One MOV-type varistor is used to protect each pair of line outlets. The power bar also has a switch, a neon pilot light and an

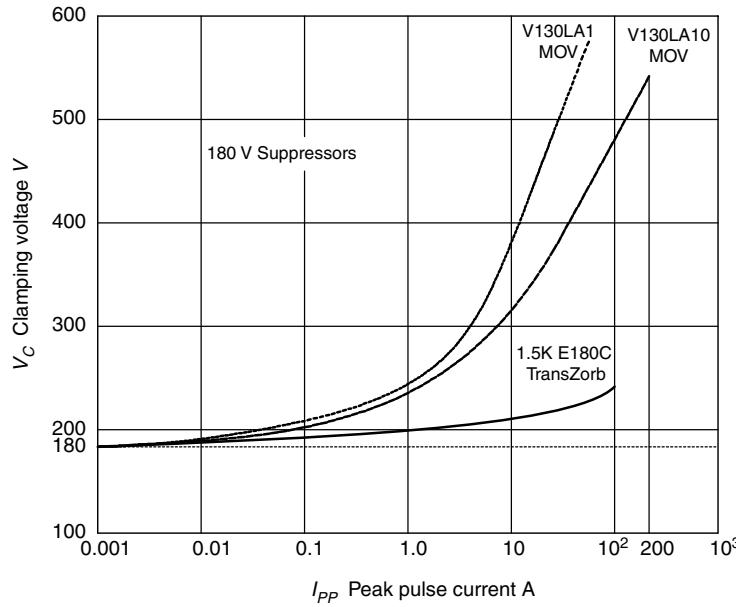
**FIGURE 3.29**

A. Volt-ampere curve for a GE V130LA20A metal oxide varistor (MOV) spike transient clipper. Device is symmetrical; only first quadrant behaviour is shown here. B. Equivalent lumped parameter circuit for the GE MOV device of Figure 3.27A.

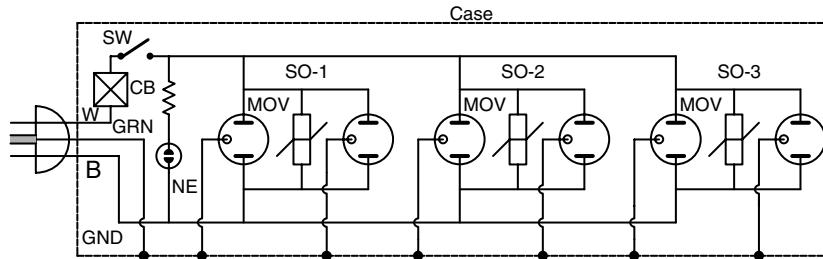
inexpensive circuit breaker for 15 A. MOVs or TransZorb type devices are also used inside sensitive equipment when the presence of a protective power bar cannot be guaranteed.

3.10.2.3 Coherent Interference Induced in Coaxial Cables by Magnetic Coupling

Shielded coaxial cables and twin axial cables are widely used in instrumentation systems to couple sensors to pre-amplifiers and signal conditioning sub-systems together.

**FIGURE 3.30**

Comparison of volt-ampere curves of two MOVs and a TransZorb + *pn* junction transient suppressor made by General Semiconductor Industries, Inc. (Source: Figure used with permission of GSI, Inc.)

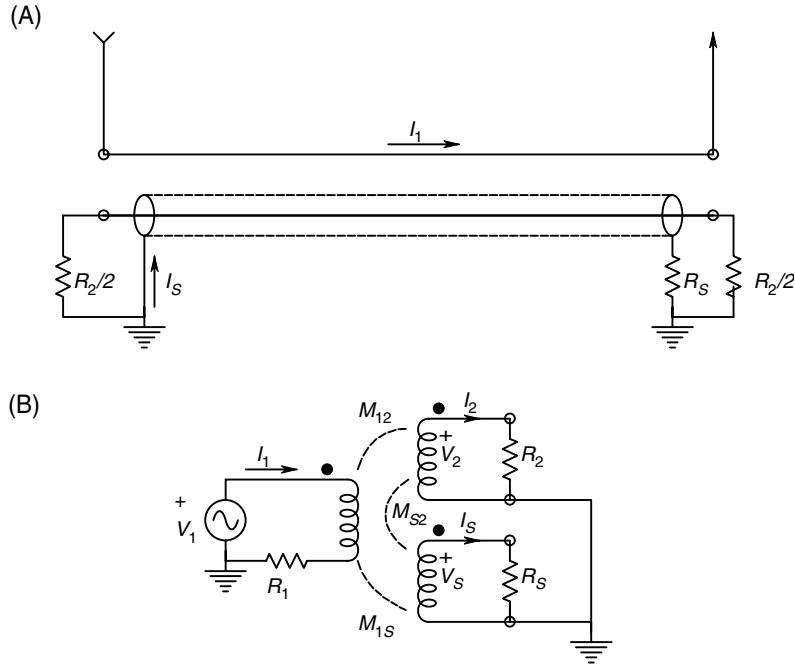
**FIGURE 3.31**

Schematic of a transient-protected, 6-outlet, 'power bar', such as those used with personal computer systems.

Unfortunately, coaxial cables can act as magnetic loop antennas under certain grounding conditions. Referring to Figure 3.32A, we see that an ac current I_1 in an adjacent wire creates flux lines which link the coaxial cable and its shield, creating a mutual inductance between conductor 1 and the coaxial cable. Both ends of the shield are assumed to be grounded in this case. This system can be represented by the lumped parameter circuit in Figure 3.32B. The open circuit coherent interference voltage induced in the center conductor is simply written as:

$$V_{2(OC)} = j\omega M_{12}I_1 - j\omega M_{S2}I_S \quad (3.95)$$

Due to the symmetrical coaxial geometry, it may be argued [Ott, 1976] that the mutual inductance between the center conductor and the shield, M_{S2} , is simply the

**FIGURE 3.32**

A. A current carrying conductor in proximity to a coaxial cable that has both ends of its shield grounded. This architecture is shown to minimize induced interference in the center conductor and to prevent magnetic interference from the cable. B. Lumped parameter equivalent circuit of the coaxial cable and wire shown in Figure 3.32A.

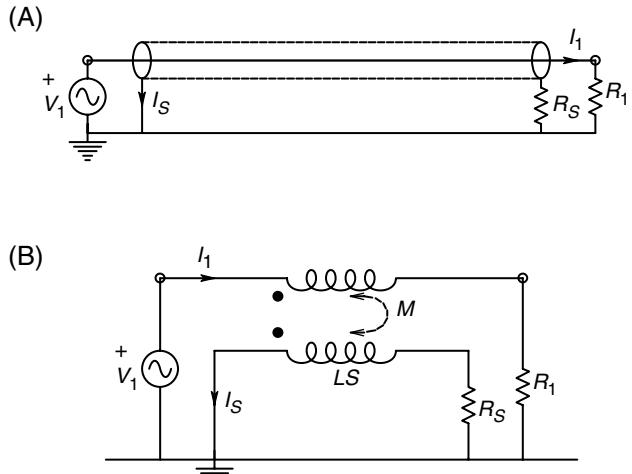
self-inductance of the shield, L_S . Also, we assume that $M_{S1} = M_{12}$. Now the current in the shield loop is given by:

$$I_S = \frac{j\omega M_{12} I_1}{j\omega L_S + R_S} = \frac{(j\omega M_{12}/R_S) I_1}{j\omega L_S/R_S + 1} \quad (3.96)$$

The shield cut-off frequency, $R_S/2\pi L_S$, is generally of the order of 2 kHz for common RG-58C coaxial cable and lower for other types of coaxial cables. If we assume that the radian frequency of the interference lies above the break frequency of the shield, we find that $I_S \approx I_1$. Thus from equation 3.95, we see that $V_{2(OC)} \approx 0$. The lesson here is that, in order to prevent magnetically induced coherent interference on a coaxially shielded line, both ends of the shield must be securely grounded. As seen below, grounding both ends of the coax's shield can, in some cases, cause interference from a ground loop.

Just as we seek to minimize pick-up of magnetically induced coherent interference, it is also important to prevent radiation of magnetic interference from a current carrying coaxial conductor. This situation is illustrated in Figure 3.33A and B. At frequencies above the shield cut-off frequency, we have:

$$I_S = \frac{j\omega L_S I_1}{j\omega L_S + R_S} = \frac{(j\omega L_S/R_S) I_1}{j\omega L_S/R_S + 1} \approx I_1 \quad (3.97)$$

**FIGURE 3.33**

A. A coaxial cable carrying current I_1 on its center conductor. B. Lumped parameter equivalent circuit of the coaxial cable showing the current induced in the shield loop, I_S . At distances greater than several cable diameters from the cable, the magnetic fields from I_1 and I_S effectively sum to zero.

Thus, the induced shield current is equal to the center conductor current and flows in the opposite direction. Hence, outside the coax, the B fields from I_S and I_1 cancel each other out vectorially and magnetic interference is not generated.

3.10.2.4 Single Grounding of Coax Shields to Prevent Ground Loop Interference

In Figure 3.34A, a circuit describing ground loop interference, it can be easily shown using Thevenin's theorem, that the ground loop EMF, V_{GL} , produces a substantial interference voltage at the input of the amplifier, given by:

$$V_{CI} = V_{GL}[R_{C2}/(R_G + R_{C2})] \quad (3.98)$$

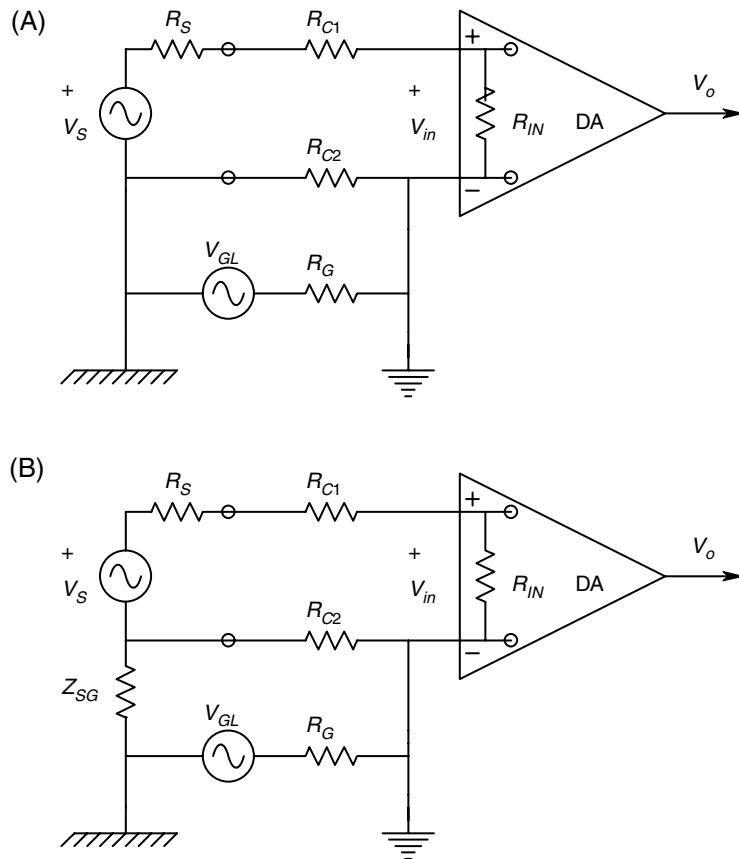
We can easily examine the case where the input source is floating with respect to ground. This case is illustrated in Figure 3.34B. Since there is a very low conductance pathway between $V_{GL(OC)}$ and V_S , there is very little interference from the ground loop EMF. In this case, the coherent interference voltage at the amplifier input is given by:

$$V'_{CI} \approx V_{GL}[R_{C2}/Z_{SG}] \quad (3.99)$$

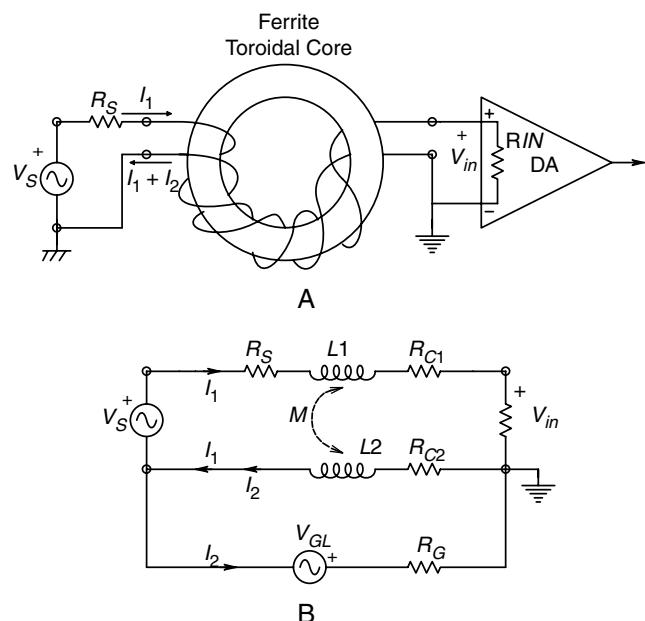
Here, it is obvious that $V_{CI} \gg V'_{CI}$ and it has been assumed that $R_{IN} \gg R_S \gg (R_{C1}, R_{C2}$ and R_G). In this case, a differential or instrumentation amplifier should be used to condition V_S and a shielded, twisted pair ('twinax') cable is used to minimize electrostatic and magnetically induced interference.

3.10.2.5 Use of a Longitudinal Choke or Neutralizing Transformer to Attenuate Common-Mode Coherent Interference

Figure 3.35A illustrates the use of a toroidal ferrite magnetic core to form a *longitudinal choke*. A longitudinal choke is used to reduce coherent interference caused by high frequency, CM interference picked up by wires connecting sensors to signal conditioning modules, or connecting sub-systems in an instrumentation system. The longitudinal choke may be used with pairs of current carrying wires, twisted pairs of wires or coaxial cables. More than one pair of wires or coaxial cable may be wound on the same

**FIGURE 3.34**

A. A differential amplifier is used to condition the voltage of a grounded source. B. In this case, the DA conditions the voltage of a floating source. The effect of V_{GL} is minimal.

**FIGURE 3.35**

A. A toroidal, longitudinal choke. B. Equivalent circuit of the longitudinal choke.

core without the circuits cross-talking. In the simplest form, the wires are passed through the center of a cylindrical ferrite core to form a one-turn longitudinal choke. This geometry is often seen on accessory cables used with PCs. To examine the operation of the longitudinal choke, refer to Figure 3.35B. First, superposition is used to find the signal voltage across R_{IN} of the signal conditioning amplifier. Setting $V_{GL} = 0$, we write two loop equations, where we assume that $L_1 = L_2 = M_{12} = M_{21} = L$.

$$V_S = I_1[R_S + R_{IN} + R_{C1} + R_{C2} + j\omega 2L] + I_2[R_{C2} + j\omega L] - j\omega 2MI_1 - j\omega MI_2 \quad (3.100a)$$

$$0 = I_1[R_{C2} + j\omega L] + I_2[R_G + R_{C2} + j\omega L_2] - j\omega MI_1 \quad (3.100b)$$

Cancelling terms and using Cramer's rule to solve the loop equations, we find:

$$\Delta = (R_S + R_{IN})(R_G + R_{C2} + j\omega L_2) \quad (3.101)$$

and V_{IN} due to V_S is given by:

$$\frac{V_{IN}}{V_S} = \frac{I_1 R_{IN}}{V_S} = \frac{R_{IN}(R_G + R_{C2} + j\omega L_2)}{(R_S + R_{IN})(R_G + R_{C2} + j\omega L_2)} \approx 1 \quad (3.102)$$

Here, we have assumed as before that $R_{IN} \gg R_S \gg R_{C1}, R_{C2}, R_G$. Of great significance, is that the signal component at the amplifier input undergoes no high frequency attenuation due to the longitudinal choke.

Next, we set $V_S = 0$ and consider the amplifier input caused by the ground loop voltage, V_{GL} . And again writing loop equations:

$$0 = I_1[R_{IN} + R_S + R_{C1} + R_{C2} + j\omega 2L] + I_2[R_{C2} + j\omega L_2] - j\omega 2MI_1 - j\omega MI_2 \quad (3.103a)$$

$$V_{GL} = I_1[R_{C2} + j\omega L_2] + I_2[R_G + R_{C2} + j\omega L_2] - j\omega MI_1 \quad (3.103b)$$

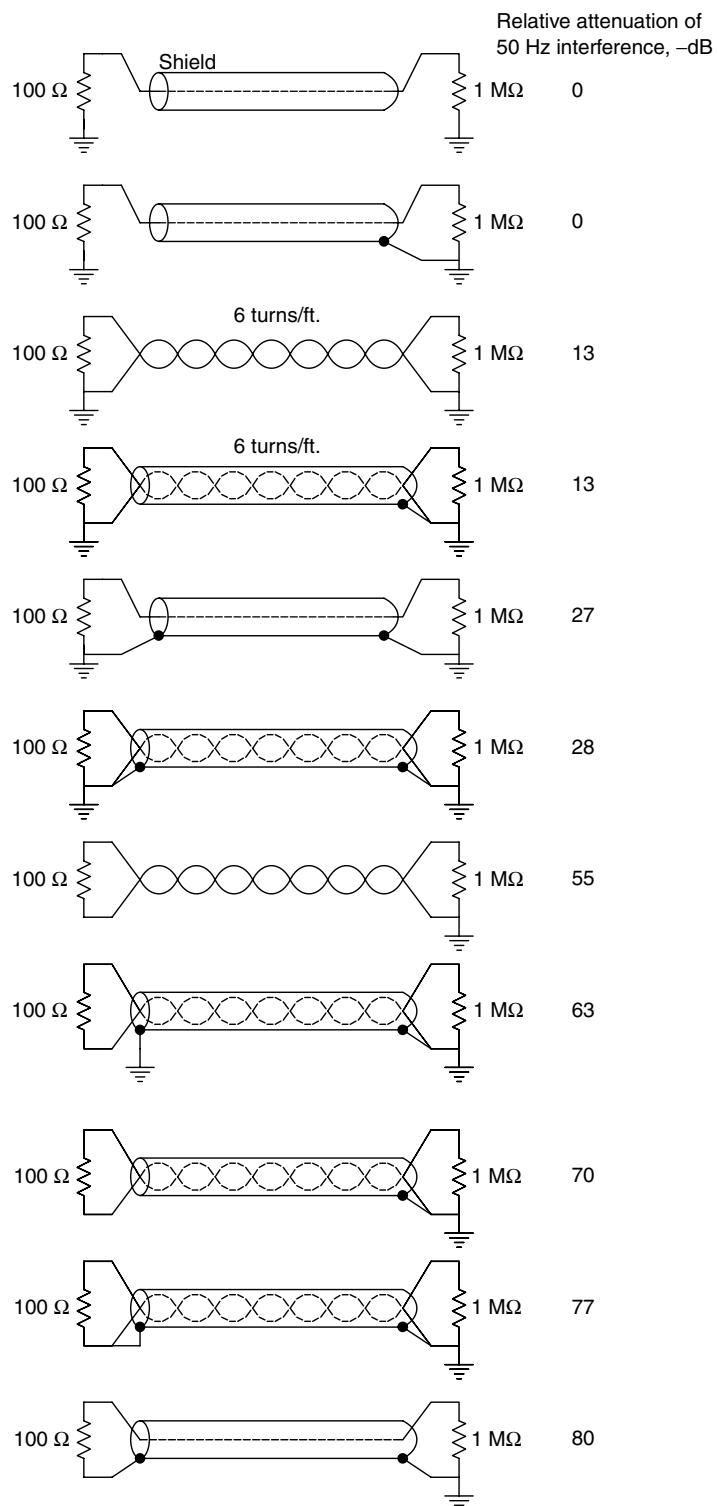
Cancelling terms, we again solve for $V_{IN} = I_1 R_{IN}$ using Cramer's rule:

$$\frac{V_{IN}}{V_{GL}}(j\omega) = \frac{R_{C2}}{(R_{C2} + R_G)[1 + j\omega L_2/(R_G + R_{C2})]} \quad (3.104)$$

From equation 3.104, we see that the interference due to V_{GL} is attenuated and low-pass filtered by the action of the longitudinal choke.

3.10.2.6 Experimental Verification of Cabling and Grounding Schemes to Achieve Minimum Noise Pickup

Ott (1976) described the effectiveness of 12 different cabling schemes in minimizing the pick-up of a 50 kHz magnetic field (Figure 3.36). The unwanted coherent interference was measured across a 1 MΩ resistor. The shielded cables were wound helically (3 turns, 17.8 cm diameter) inside the helical field, generating coil (10 turns of 20 ga. wire, 23 cm diameter, carrying 0.6 A at 50 kHz). Ten volts peak-to-peak was impressed across the field generating coil. From his tests, Ott reported that 0 dB was 0.8 V. We have arranged the cables in Figure 3.36 in the order of increasing magnetic (and electric) shielding effectiveness. Note that the simple coaxial cable with ungrounded shield performed the poorest. The better performing cables had circuits in which the 100 Ω Thevenin source resistance was not tied to ground. Although Ott's evaluation was

**FIGURE 3.36**

Relative effectiveness of 12 different cabling circuits subject to a 50 kHz magnetic field determined experimentally by Ott (1976).

focused on evaluating the shielding from a magnetic field, his system also produced an electric field. Caution should be used in over-generalizing his results. Many subtle factors contribute to the pick-up of coherent interference, including circuit geometry, the frequency and the spatial distributions of E and B fields.

Ott's results show that the use of shielded twinax cables with single grounds on the shield gave the best results (lowest pick-up of coherent interference from a predominantly magnetic source). He attributed the -80 dB attenuation shown by the simple coaxial cable to be anomalous, due to the fact that this circuit had a smaller net loop area, A , than the twinax cables. This anomalous result underscores what often appears to be the arcane nature of coherent noise reduction techniques, which can appear 'arty' to the inexperienced. However, the laws of physics apply to coherent noise reduction practices, as well as other branches of electrical engineering and instrumentation.

3.10.2.7 Circuit Grounding

Probably the single most important principle of reducing the pick-up of coherent noise from internal sources in an electronic system is good grounding practice. It must be appreciated that typical ground currents contain a large dc component on which rides a coherent (ac) component, which is the result of conditioning the desired signal, or is from the high frequency switching of digital (logic) circuits. At high frequencies, ground wires appear as series $R-L$ circuits, whose impedance increases with frequency. If a circuit card contains both digital circuitry and low level analog amplifiers, and has

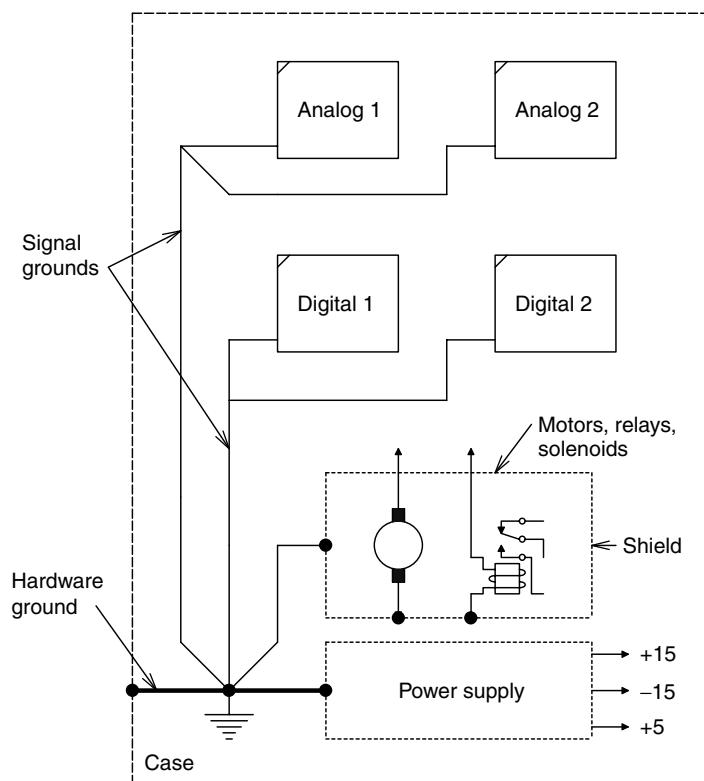


FIGURE 3.37

Good grounding and shielding practice in an electronic system.

a single ground connection, it is almost certain that the analog system will pick up the coherent interference from the ground current flowing through the common ground impedance of the digital circuits. Obviously, separate ground paths will help this situation. Figure 3.37 illustrates a good grounding and electrostatic shielding architecture. Note that there is a common tie point for all grounds at the power supply terminal. This terminal may or may not be tied to the metal instrument case, which is always tied to the green wire (power line earth ground). It is important to note that separate grounds are used for low level analog, high level analog, digital and inductive (arcing) circuits such as motors, solenoids and relays. The ground wire itself should be of heavy gauge; if the + and -15 V dc supplies to a circuit card use #18 wire, then the ground should be #14 stranded wire.

3.10.2.8 Ferrite Beads and Feed through Capacitors

An effective way of attenuating unwanted coherent interference on signal and power lines at frequencies above 1 MHz is through the use of one or more (seldom more than three) ferromagnetic beads strung on the wire carrying the offending interference. Ferrite beads act as small chokes, making the wire upon which they are strung appear to be a series $R-L$ circuit. Due to their internal losses, some ferrite beads cause their wire to have a nearly constant (resistive) impedance of about 30 to 75 Ω and <1–100 MHz. Other beads are made to appear more inductive and their Z increases linearly with frequency, going from 20 Ω at 5 MHz to 110 Ω at 100 MHz (Ott, 1976). The 'type 64' material bead made by Amidon Associates, N. Hollywood, CA, is designed to operate above 200 MHz. Ferrite beads make effective, well-damped, second order low-pass filters when combined with appropriate bypass capacitors to ground. One widely used bypass capacitor is the feed through capacitor, used to decouple RF noisy sources from the power supplies. Typical feed through capacitors have cylindrical, threaded metal bodies which are mounted in the metal case or shield wall and the terminals at both ends permit the dc current to pass through. Capacitance from the center through the conductor to the case can range 0.5–5 nF, depending on the application.

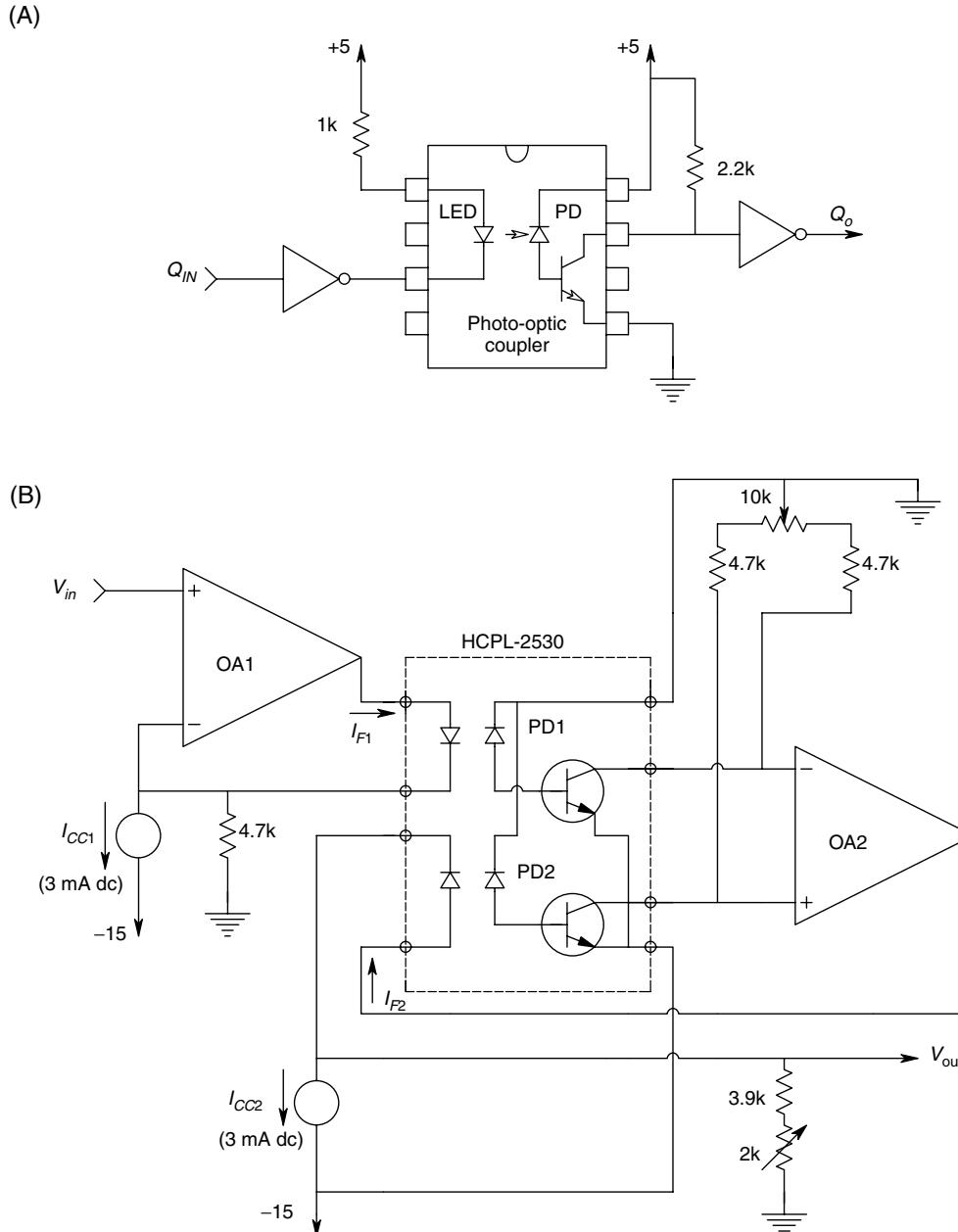
In some cases, the use of ferrite beads will produce more harm than good; misuse of beads can cause unwanted, high frequency resonances and oscillations. Properly applied, ferrite beads and feed through capacitors can provide a simple, low cost means of reducing coherent interference and parasitic oscillations.

3.10.2.9 Interruption of Ground Loops by the Use of Isolation Transformers and Photo-Optic Couplers

Isolation transformers are simple add-ons which can often improve measurement of the system SNR when used between analog signal conditioning subsystems. A typical isolation transformer used with 50 Ω coaxial cables has terminating and source impedances of 50 Ω and a frequency response of 25 Hz–20 MHz. There can be several thousand megohms impedance between the primary and secondary coax shields and hence, the CM EMF has little effect (equation 3.99) on the signal.

3.10.2.10 Photo-Optic Couplers

These are nonlinear devices, primarily used in digital interfaces where it is desired to have complete ground and signal isolation between digital circuits. A typical optocoupler consists of an LED on a chip in optical, but not electrically conductive, intimacy with

**FIGURE 3.38**

A. A photo-optic coupler is used to isolate a TTL digital signal line. B. Use of a dual optocoupler in a feedback circuit in order to realize isolated, linear, analog signal transmission.

an output circuit which consists of a phototransistor or photodiode, which conducts when illuminated by the input LED. The primary use of optocouplers is isolated, high speed, serial digital data transmission. A simple TTL logic circuit coupled with a photo-optic coupler is shown in Figure 3.38A.

Photo-optic couplers can be modified to give optically isolated, analog signal transmission by dc biasing the photodiode and phototransistor to operate in the middle

of their dynamic regions, and using negative feedback to reduce nonlinearity. Such linear photo-optic coupler circuits can have small signal bandwidths from dc to >1 MHz. A feedback analog photo-optic coupler circuit, adapted from a circuit in the Hewlett-Packard Optoelectronics Designer's Catalog, is shown in Figure 3.38B. In this circuit, a HCPL-2530 dual optocoupler is used. The feedback forces the ac signal current through PD2, to track the input signal current in PD1, thus effectively cancelling out much of the nonlinearity that is inherent in the photodiodes and photo-transistors.

3.10.2.11 The Use of Guarding and Shielding to Reduce Common-Mode, Coherent Interference

When extremely low-level signals are being measured, or when very high CM coherent noise is present, as in the case of certain biomedical applications, strain gauge bridge applications, or thermocouple measurements, the use of a guard shield can effectively reduce the input capacitance to ground of the amplifier, hence giving rise to coherent interference. Figure 3.39 illustrates a conventional differential instrumentation amplifier (DA) connected to a Thevenin source (V_S, R_S), by a shielded, twinax cable. The cable shield is tied to the source ground and not to the power supply ground. A large, coherent, CM voltage, V_{GL} , exists between the power ground and the source ground. If no currents flow in the twisted pair wires, then V_{GL} appears as the CM voltage at the amplifier input and is rejected by the amplifier CMRR; there is no DM component of V_{GL} . However, the input capacitances of the DA allow input currents to flow. Note that there is a capacitance C_{1S} between the HI side of the source and the cable shield and conductor 2. Due to asymmetry in the circuit, a DM component of V_{GL} appears across the amplifier input. This DM interference component can be calculated. By inspection of the circuit in Figure 3.39, $V_- = V_{GL}$. V_+ is given by the reactive voltage divider as:

$$V_+ = V_{GL} \frac{1/j\omega C_{1G}}{[1/j\omega C_{1G} + 1/(j\omega C_{1S} + G_S)]} = V_{GL} \frac{j\omega C_{1S} R_S + 1}{j\omega(C_{1S} + C_{1G}) + 1} \quad (3.105)$$

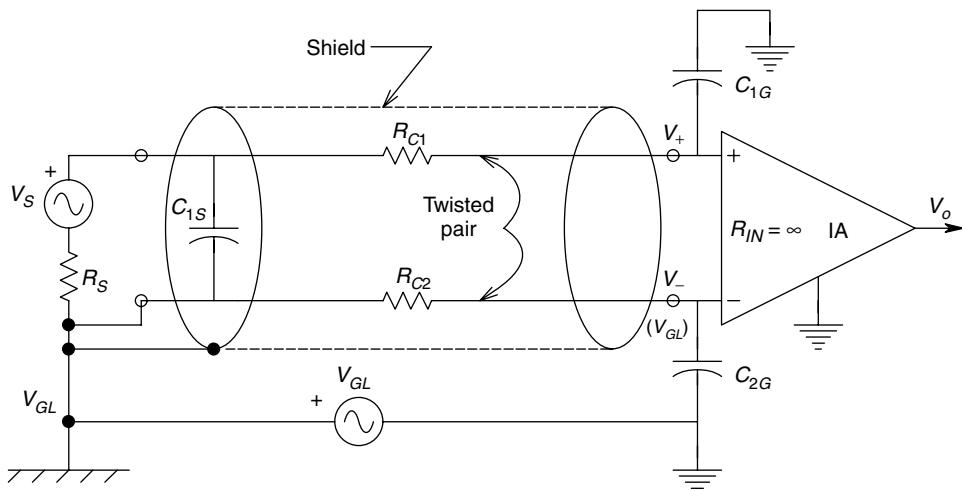


FIGURE 3.39

An instrumentation amplifier (IA) connected to a Thevenin source by twinax shielded cable.

The phasor differential interference voltage at the DA input is:

$$V_d = (V_+ - V_-)/2 \quad (3.106)$$

Substituting from the relations above, we find:

$$V_d = -V_{GL}(j\omega C_{1G} R_S) = 1/[R_S(C_{1G} + C_{1S})] \text{ r/s} \quad \text{for } \omega \ll \omega_b \quad (3.107a)$$

$$= -(V_{GL}/2)\overline{C_{1G} + C_{1S}} \quad \text{for } \omega \gg \omega_b \quad (3.107b)$$

For a numerical example, let $R_S = 1 \text{ k}\Omega$, $C_{1G} = 20 \text{ pF}$, $C_{1S} = 280 \text{ pF}$ and $\omega = 377$ (60 Hz ground loop voltage).

Assume $R_{IN} = \infty$ and $R_{C1} = R_{C2} = R_{GL} = 0$. The f_b is found to be 531 kHz and from equation 3.107a, the magnitude of the DM interference component is:

$$V_d = V_{GL} (3.77 \times 10^{-6}) \quad (3.108)$$

This DM interference can seriously degrade output SNR when V_S is in the single μV range and V_{GL} is in the hundreds of mV. To reduce the interference, V_d , steps are taken to reduce the input capacitances to ground. This can be accomplished by using an isolation amplifier (Section 2.4.2) in conjunction with a *guard shield* and external shield, as shown in Figure 3.40. Here, the guard shield is tied to the amplifier ground and is made to float with respect to 'true ground' as a consequence of the RF power supply of the isolation amplifier. The current in the ground loop and hence, the magnitude of the DM interference voltage, is reduced at least twenty-fold because the capacitance from either amplifier input to true ground (C_{1G} , C_{2G}) is now reduced by the external shield and is of the order of 1 pF. The same voltage divider is used to calculate V_d .

Figure 3.41 illustrates the input circuit of a guarded meter. Note that the following analysis of the guarded meter circuit is also valid for a DA circuit in which the CM

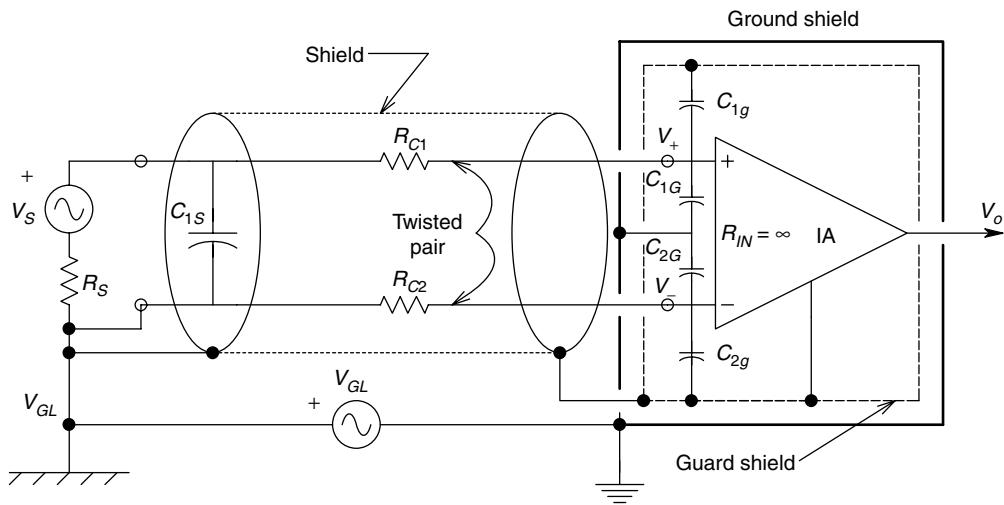
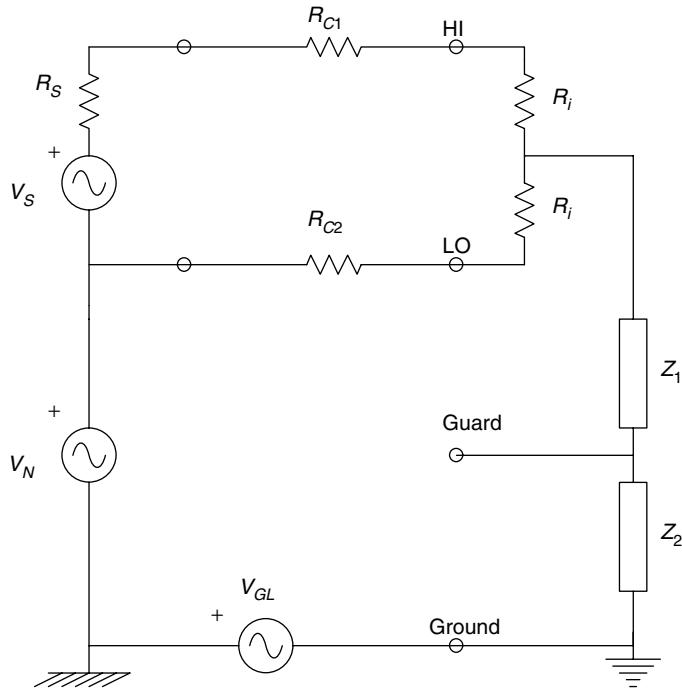


FIGURE 3.40

An isolation amplifier (ISOA) is used with a guard shield and a ground shield to reduce ground loop interference.

**FIGURE 3.41**

Equivalent input circuit of a voltmeter with guard and ground terminals. V_{GL} is the ground-loop voltage and V_N is another common-mode noise source.

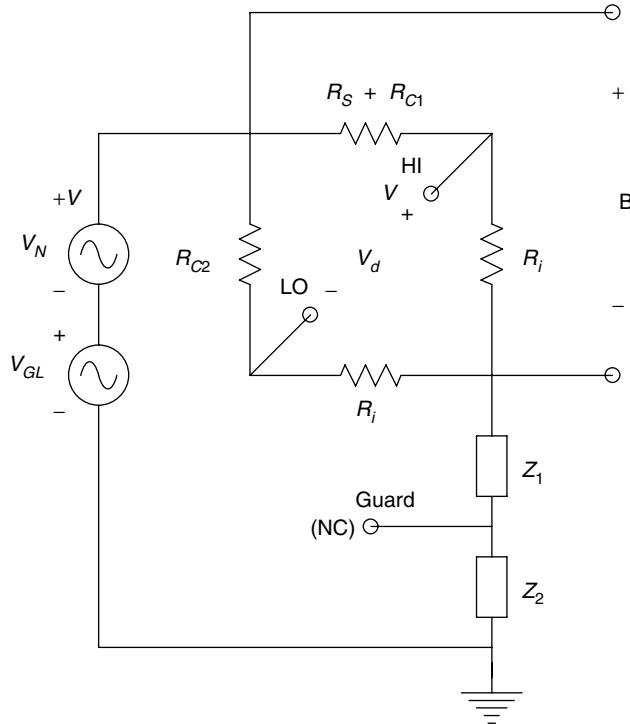
point, CP, is taken to the guard terminal and to the ground through impedances Z_1 and Z_2 , as shown in Figure 3.41. In this circuit, four meter terminals are of interest: HI, LO, GUARD and GROUND. Here, as in the preceding analysis, an unwanted, coherent, CM, ground loop voltage, V_{GL} , is present between true ground and the remote source ground, as well as some other, undesired, CM voltage associated with the source, V_N .

Five ways of connecting the guard of the guarded meter will be considered. The analysis of each case is simplified by calculating the input voltage at the meter, $(V_{HI} - V_{LO})$, as a function of circuit parameters. In the first case, we will leave the guard terminal open-circuited. We assume that the input resistances of the meter or DA, R_i , are very large—at least $10^8 \Omega$. The resistances of the twisted pair wires, $R_{C1} = R_{C2} = R_C$ are about 0.1Ω . The source resistance (R_S) is 1000Ω and Z_1 and Z_2 are typically capacitive reactances of about $10^5 \Omega$. As can be seen from Figure 3.42, the differential input voltage to the meter, V_d , can be considered to be the unbalance voltage of a Wheatstone bridge whose four arms are R_C , $(R_S + R_C)$, R_i and R_i . The voltage across the bridge, V_B , is approximately given by:

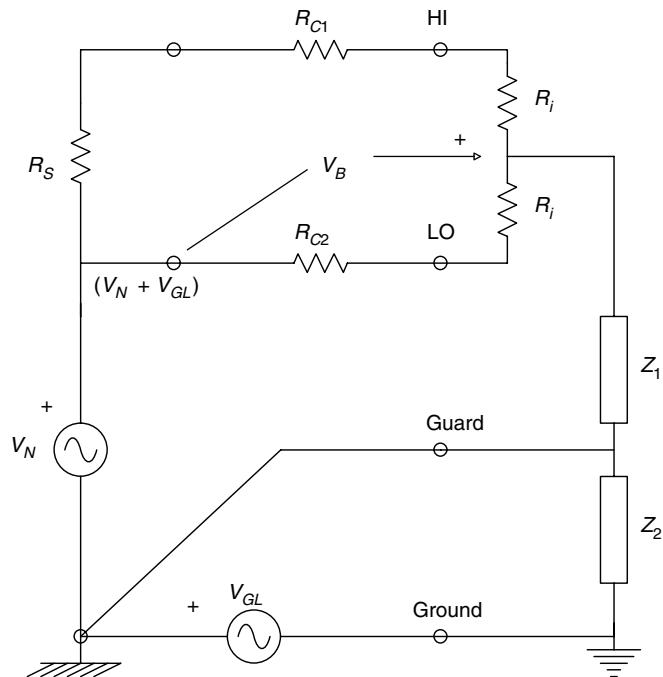
$$V_B = (V_{GL} + V_N) \frac{R_i/2}{R_i/2 + Z_1 + Z_2} \cong (V_{GL} + V_N) \quad (3.109)$$

Again, taking into account the sizes of the resistors in the circuit, it is easy to show that V_d is given by:

$$V_d \cong (V_{GL} + V_N)(-R_S/R_i) = (V_{GL} + V_N)(-10^{-5}) \quad (3.110)$$

**FIGURE 3.42**

Equivalent input circuit of the voltmeter when the guard is not used. $V_S = 0$. Note the bridge architecture.

**FIGURE 3.43**

Equivalent input circuit of the voltmeter when the guard is incorrectly connected to the source ground.

In the second case, we assume that the guard is incorrectly connected to the source ground. This connection effectively shorts out V_{GL} . However, V_N still produces a DM interference voltage at the meter input. The equivalent circuit for this case is shown in Figure 3.43. Again, a bridge circuit is formed with an unbalance voltage, V_d , of:

$$V_d = V_N(-10^{-5}) \quad (3.111)$$

In the third case, shown in Figure 3.44, the guard is incorrectly connected to the meter's LO terminal.

Here, it is easy to see that:

$$V_d = (V_{GL} + V_N)[1 - Z_2/(Z_2 + R_C)] = (V_{GL} + V_N)(R_C/Z_2) = (V_{GL} + V_N)(10^{-6}) \quad (3.112)$$

Figure 3.45 illustrates the fourth case where the guard is incorrectly connected to the meter ground, shorting out Z_2 . The differential interference voltage is found from the bridge circuit to be approximately:

$$V_d = (V_{GL} + V_N)(-R_S/R_i) = (V_{GL} + V_N)(-10^{-5}) \quad (3.113)$$

In the fifth example, the guard terminal is correctly connected to the source LO (i.e. the node between R_{C2} , R_S and V_N). This situation is shown in Figure 3.46. Clearly in this case, $V_B = V_d = 0$, and there is no DM component at the meter input from the coherent noise voltages.

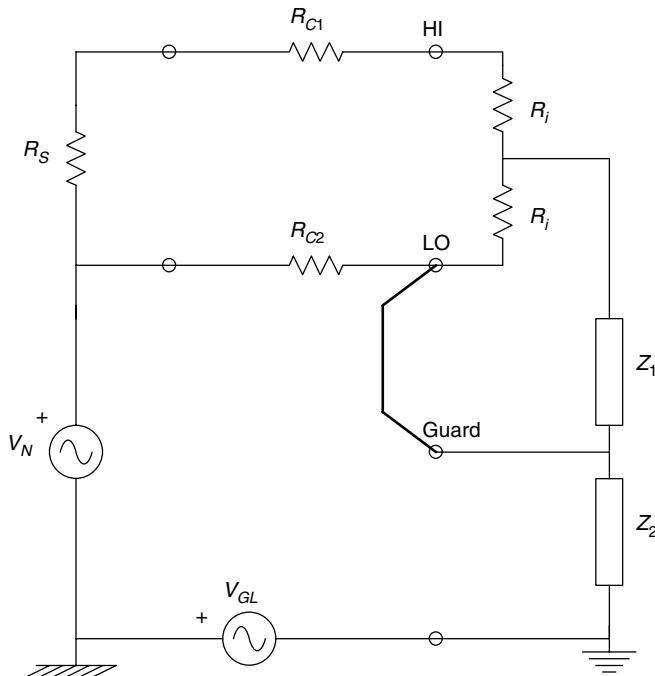
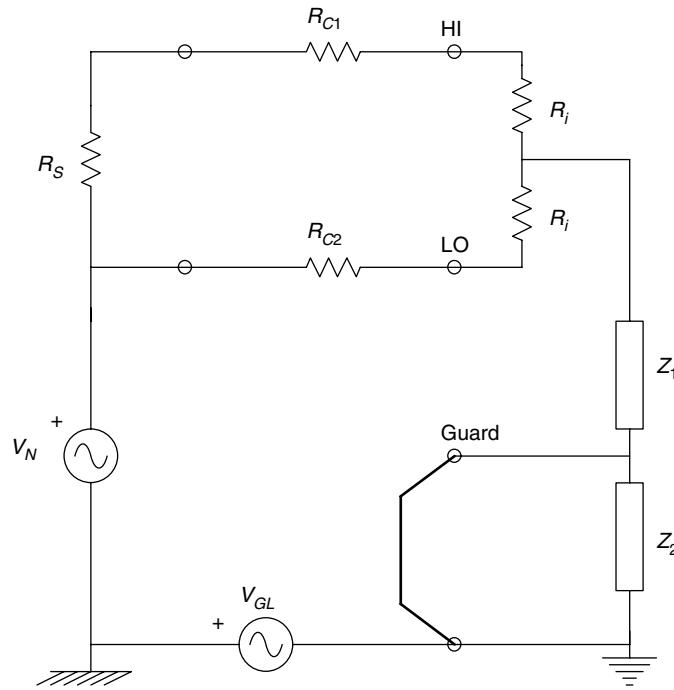
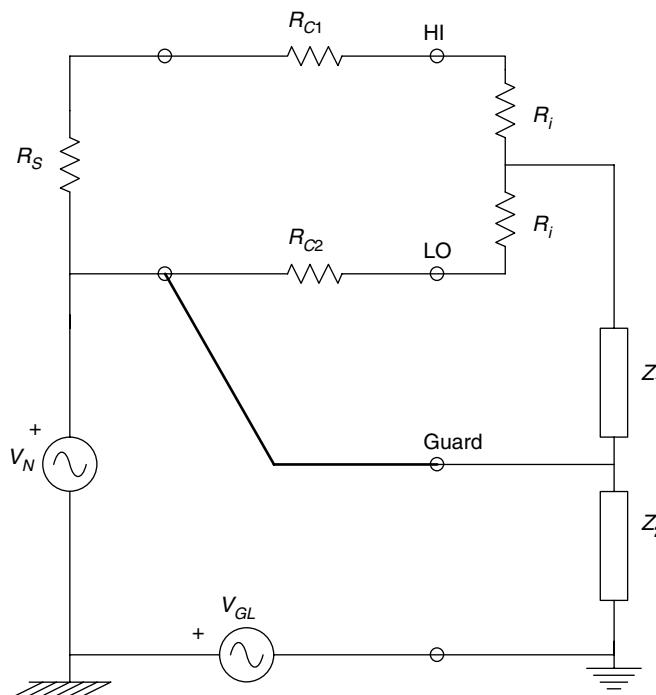


FIGURE 3.44

Equivalent input circuit of the voltmeter when the guard is again incorrectly connected to the LO (-) input terminal.

**FIGURE 3.45**

Equivalent input circuit of the voltmeter when the guard is incorrectly connected to the meter ground.

**FIGURE 3.46**

Equivalent input circuit of the voltmeter when the guard is correctly connected to the source low. There is no difference-mode input component from V_N and V_{GL} .

3.10.3 Summary of Techniques for Coherent Noise Reduction

The practice of the art and science of coherent noise reduction should begin with the elimination of local noise sources. Obviously, these are noise sources which have been identified and which can be controlled. These sources might include digital and computer equipment operating in the same immediate environment, or interference caused by sparking motor commutators, etc. The offending noise sources must be enclosed in appropriate, grounded, electrostatic shields. Their power leads must be low-pass filtered or bypassed as they leave the shield—the use of ferrite beads and feed through capacitors may be indicated. In addition, power leads should be twisted so that minimum magnetic field is created.

Inductive spikes induced on power lines can be minimized by using varistors or diodes across switched relay and solenoid coils. Both ends of coaxial cables carrying RF should be grounded to minimize the radiation of magnetic field interference. As we have seen above, there are many methods for reducing the pick-up of coherent interference in low level analog signal conditioning systems. Lead dress is important—separate low level leads from wires carrying high currents and voltages. Low level leads should be twisted and routed near the grounded case. If twinax (twisted and shielded) cable is used, it should be grounded at one end only, and insulated. Alternate grounds and signal leads on ribbon cables and in connectors must be ensured. High impedance, low level leads should be kept as short as possible—this is especially true in wiring op-amps. The summing junction should never have an exposed lead of more than 1.3 cm connecting components to it.

Often, the pick-up of coherent interference can be traced to the magnetic component of a low frequency electromagnetic field. Such pick-up may occur in spite of rigorous electrostatic shielding practice. In this case, a mu metal or other loss-type, non-saturating magnetic shield may be effective, as will be reducing the area and changing the spatial orientation of the pick-up loop.

We have seen above that ground loops can be a serious source of interference in instrumentation systems. The effect of ground loop EMFs can often be minimized by the use of photo-optical couplers (for digital circuits), isolation transformers (for signals in the 20 Hz–200 kHz range), longitudinal chokes (to attenuate high frequency CM interference), differential isolation amplifiers and guarded inputs. Good grounding practice can significantly reduce ground loops. We have seen that a single ground point is the best practice (except for RF circuits). Individual, insulated grounds from subsystems should be brought separately to the common ground point.

Just as separate grounds are important for interference reduction, so is the use of decoupled, low impedance, dc power supplies with proper power ratings. Dedicated, low noise, analog power supplies for signal conditioning system headstages can help reduce hum and the cross-coupling of interference from other stages.

3.11 Chapter Summary

In this chapter, we have presented a heuristic, yet mathematically sound treatment of random noise in signal conditioning systems based on the one-sided power density spectrum. Sources of random noise in active and passive circuits were considered—thermal (Johnson) noise in resistors, shot noise in bipolar junction transistors and noise in FETs. The dependence of transistor noise performance on dc operating point was

stressed. The standard, two-generator noise source model for amplifiers was introduced, as well as the concepts of noise factor, noise figure and output SNR as figures of merit with respect to noise.

The propagation of Gaussian noise through linear filters was introduced, and the concept of a filter's gain squared bandwidth was applied to finding the MS noise output of signal conditioning systems, given white input noise. We showed how a low noise transformer can be used to increase output SNR when a signal source with a low source resistance drives the input of a noisy amplifier.

Five practical, numerical examples of noise calculations in measurement system design were given. Problems at the end of this chapter also stress calculation of threshold input signal levels to get a specified output SNRs.

We presented, in tabular form, a list of commonly used low noise, IC amplifiers. Here, we noted that the lowest system e_{na} values are generally found in amplifiers having BJT input stages, hence relatively low input resistances, high bias currents and thus, large i_{na} values. FET input amplifiers have slightly higher e_{na} values, but offer much higher input resistances, lower dc bias currents and hence, lower i_{na} values than do BJTs. The wise designer of low noise signal conditioning equipment thus must make compromises between e_{na} , i_{na} , input impedance, bandwidth and other performance factors when choosing an IC amplifier for headstage use.

Section 3.10, we considered coherent interference, its sources and some of the many ways to minimize its effect in instrumentation systems. Coherent interference can appear at the output of a signal conditioning system as the result of electrostatic or magnetic coupling to input leads. We also demonstrated that it can enter an amplifier through the power supply, or through improper grounding. We discussed some of the more common means of mitigating coherent interference and analyzed equivalent circuits relevant to those means.

Problems

- 3.1 (A) Find the numerical value for the RMS noise voltage in a 100 Hz to 20 kHz noise bandwidth made by a $1\text{ M}\Omega$ resistor at 300 K. Use $4kT = 1.656 \times 10^{-20}$ at 300 K.
 (B) The same resistor is shunted by a 23 pF capacitor to ground. Find an expression for $S_{no}(f)$, the one-sided noise power density spectrum across the RC circuit.
 (C) Find the RMS noise voltage across the parallel RC circuit over $0 \leq f \leq \infty$. Note that:

$$\int_0^{\infty} \frac{a \, dx}{a^2 + x^2} = \pi/2$$

- 3.2 A low level sinusoidal signal, $v_s(t) = V_s \sin(2\pi f_o t)$, is conditioned by a noisy amplifier, shown in Figure P3.2, with a gain of $K_V = 10^3$. The amplifier is followed by a noiseless LPF. The amplifier has an equivalent, white SC input voltage noise root power density spectrum of $e_{na} = 14 \text{ nV}/\sqrt{\text{Hz}}$. The $1.5 \text{ k}\Omega$ source resistor also makes thermal white noise @ 300 K $f_o = 100 \text{ Hz}$, $\tau = 0.01 \text{ s}$, $V_s = 10 \mu\text{V pk}$.
- (A) Find an expression for the RMS sinusoidal output signal at the V_o node. Evaluate numerically, considering the LPF gain at f_o .

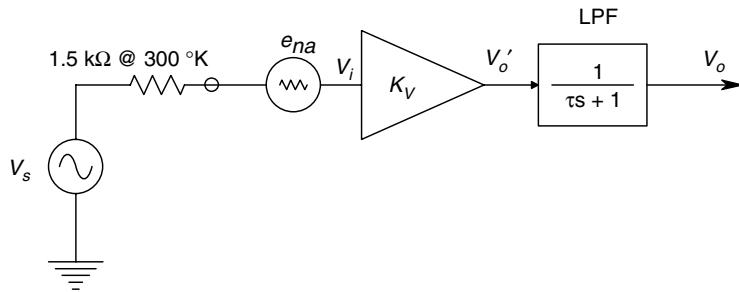


FIGURE P3.2

- (B) Find an expression for the total RMS output noise voltage, V_{oN} , at the V_o node. Use the gain squared BW tables.
- (C) Find the filter time constant, τ_o , that will maximize the output RMS SNR. Calculate this maximum SNR.
- 3.3 A Wheatstone bridge, shown in Figure P3.3, is given sinusoidal excitation: $v_B = 7.071 \sin(2\pi 400t)$. The bridge resistors are $R = 330 \Omega$ at 300 K. An ideal DA with a differential gain $K_D = 10^3$ is used to condition the bridge output. The DA has a white $e_{na} = 15 \text{ nV}/\sqrt{\text{Hz}}$. Following the DA is an ideal, noiseless BPF, with unity gain in the passband centered at 400 Hz and zero gain elsewhere. The passband has width $B = 10 \text{ Hz}$.

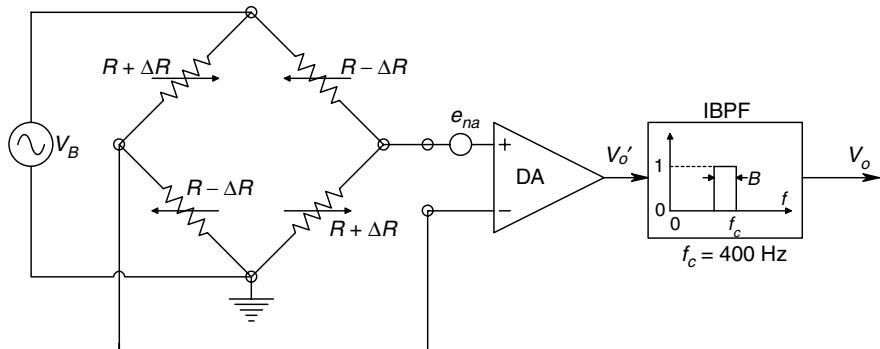


FIGURE P3.3

- (A) Find the ΔR that will give an RMS sinusoidal output voltage at the V_o node equal to the RMS noise at V_o . Neglect ΔR in calculating the output noise.
- 3.4 A noise source has a one-sided power density spectrum (PDS) given by:

$$S_n(f) = \frac{4 \times 10^{-6}}{f^2 + 10^{10}} \text{ MSV/Hz}$$

- (A) Find the value of the equivalent white noise, root PDS that approximates the low frequency portion of $S_n(f)$ above.
- (B) Find the total RMS noise voltage in $S_n(f)$.

3.5 The short circuit input noise PDS of a certain JFET is given by $e_{na}^2(f) = (4kT/g_m)(1 + f_c/f)$ MSV/Hz where $T = 27^\circ\text{C}$, $g_m = 0.003 \text{ S}$, $f_c = 15 \text{ Hz}$. Calculate the RMS noise voltage in $e_{na}^2(f)$ for $0.1 \leq f \leq 100 \text{ Hz}$.

3.6 The resistor in a parallel $R-L-C$ circuit makes thermal white noise.

- (A) Find an expression for the one-sided, voltage noise PDS at the V_o node.
- (B) Find an expression for the total RMS noise voltage at the V_o node over $0 \leq f \leq \infty$. Use the table of gain squared BW integrals.

3.7 The system of Figure P3.7 is used to condition a small 100 Hz sinusoidal voltage accompanied by white noise before it is detected by a PSR/LPF. Assume the transformer is ideal and has a $1:n$ turns ratio. Assume the quadratic BPF is noiseless and is tuned to 100 Hz. A white environmental noise, e_{NS} , is recorded with $v_s(t) = V_S \sin(2\pi 100t)$

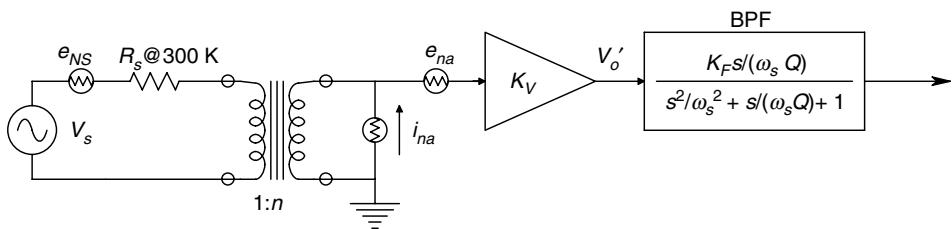


FIGURE P3.7

- (A) Give the Thevenin equivalent circuit the amplifier 'sees' looking toward V_s , R_s .
- (B) Write an expression for the one-sided white noise PDS at the V'_o node, S'_{no} . Note that there are four noise sources in the system. The instrumentation amplifier (IA) has a gain of $K_V = 10^3$.
- (C) Write an expression for the mean squared signal voltage at the V_o node, $\overline{v_{so}^2}$.
- (D) Write an expression for the mean squared noise voltage at the V_o node. Note that the gain squared BW of the quadratic BPF is $K_F^2(\omega_s/2Q)$.
- (E) Write an expression for the MS SNR at the V_o node, SNR_o . Put all n^2 terms in the denominator.
- (F) Derive an expression for the turns ratio, n_o , that will maximize SNR_o . Note that minimizing the denominator of SNR_o maximizes SNR_o .
- (G) Give an expression for the maximum SNR_o .
- (H) Let $e_{na} = 10 \text{ nV RMS}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA RMS}/\sqrt{\text{Hz}}$, $R_s = 52 \Omega$ at 300 K , $K_V = 10^3$, $K_F = 10$, $Q = 10$, $\omega_s = 2\pi 100 \text{ r/s}$, $e_{NS} = 1 \text{ nV RMS}/\sqrt{\text{Hz}}$, $4kT = 1.656 \times 10^{-20}$. All noise sources are white. Calculate n_o to give maximum SNR_o . Also calculate the peak V_s to give maximum $\text{SNR}_o = 1.0$ at the V_o node.

3.8 In order to maximize the SNR_o at V_o , a sinusoidal source with a low R_s is connected to the amplifier through an ideal transformer. The transformer's turns ratio, n , is adjusted to maximize the SNR_o . The op-amp is noisy, but otherwise ideal. Its $e_{na} = 10 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA}/\sqrt{\text{Hz}}$, both white. Assume all resistors are at 300 K and make thermal white noise. The signal is $v_s(t) = V_S \sin(2\pi ft)$.

- (A) Find an expression for the MS output signal voltage, $\overline{v_{os}^2}$.
- (B) Find an expression for the MS output noise voltage, $\overline{v_{os}^2}$. Assume a known noise bandwidth, B Hz, and include the five white noise sources.
- (C) Derive an expression for the transformer turns ratio, n_o , that will maximize the output SNR.
- (D) Find numerical values for n_o and the output $\text{SNR}_{o\text{MAX}}$.
- 3.9 A silicon PIN photodiode is used in the biased mode, as shown in Figure P3.9. An AD549 electrometer op-amp is used. It has $e_{na} = 35 \text{ nV}/\sqrt{\text{Hz}}$ (white), $i_{na} = 0.16 \text{ fA}/\sqrt{\text{Hz}}$ (white), $V_{OS} = 0$, $I_B = 35 \text{ fA}$ dc (inward), $R_{in} \rightarrow \infty$. The photodiode (PD) has $I_{rs} = 0.1 \text{ nA}$, its dark current at 15 V reverse bias is $I_{DK} = I_{rs} + I_{leak} = 4 \text{ nA}$, its photocurrent sensitivity at 850 nm is $[\eta q\lambda/hc] = 0.342 \text{ A/W}$. Assume that the PD makes only white shot noise with power density spectrum: $i_{sh}^2 = 2q(I_{DK} + I_p)$ MS A/Hz. The Hz noise bandwidth is $B = 1 \text{ Hz}$, $4kT = 1.656 \times 10^{-20}$, $q = 1.602 \times 10^{-19} \text{ C}$, $R_F = 10^6 \Omega$ makes white thermal noise and the reverse biased PD presents an infinite dynamic resistance to the summing junction.

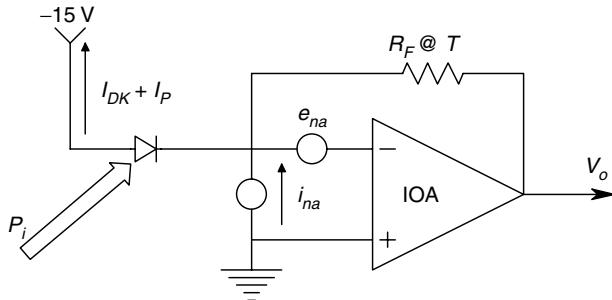


FIGURE P3.9

- (A) Find the dc V_o in the dark, V_{oD} .
- (B) Assume an 850 nm input light power of P_i W. Derive an algebraic expression for the MS SNR_o of the amplifier.
- (C) Calculate the P_i value to produce a dc ΔV_o equal to the RMS noise voltage output in the dark, where $V_o = V_{oD} + V_o$.
- 3.10 Refer to Figure P3.10. When the SNR_o maximization transformer is truly ideal, it has been shown that the optimum turns ratio is given by $n_o = e_{na}/(i_{na}R_s)$. In this problem, consider the finite winding resistances of the transformer, which is otherwise ideal. The primary resistance is $n_w = 10 \Omega$; the secondary resistance is $10 \times n \Omega$. Other parameters are $R_s = 100 \Omega$, $e_{na} = 10 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ pA}/\sqrt{\text{Hz}}$, $K_V = 10^3$ and $B = 10^4 \text{ Hz}$.

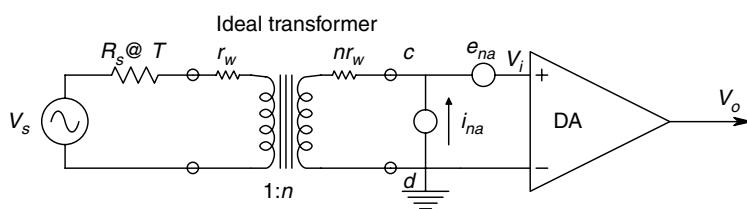


FIGURE P3.10

- (A) Draw the Thevenin equivalent circuit seen by the amplifier looking into nodes C, D.
- (B) Write an algebraic expression for the amplifier's MS SNR_o.
- (C) Put all the n terms in the SNR_o expression's denominator, $D(n)$. Note that $D(n)$ has a minimum at some n_o , giving SNR_o(n_o) a maximum. Set $dD(n)/dn = 0$ and solve numerically for the positive real root, n_o . Note that $dD/dn = 0$ gives a fourth-order algebraic equation in n . Its roots can be found using the Matlab Roots utility.
- 3.11 Gaussian white noise is added to a sine wave of frequency ω_o . The sum, $x(t)$, is passed through a linear filter described by the ODE $\ddot{v}_o = aV_o + Kx(t), x(t) = n(t) + V_S \sin(\omega_o t)$.
- (A) Find an expression for the steady state output MS SNR.
- (B) Find the value of a that will maximize the output MS SNR.
- (C) Give an expression for the maximized output MS SNR.
- 3.12 In an attempt to increase the MS SNR at the output of a signal conditioning system, the circuit architecture of Figure P3.12 is proposed. N low noise pre-amplifiers, each with voltage gain K_V and the same short circuit input voltage noise, e_{na} RMS V/ $\sqrt{\text{Hz}}$, are connected as shown. The ideal op-amp and its gain determining resistors can be considered to be noiseless. R_s at temperature T K makes Johnson noise. Note that $e_{na1}(t) \neq e_{na2}(t)$, etc., and $e_{na1} = e_{na2} = \dots = e_{naN} = e_{na}$ RMS V/ $\sqrt{\text{Hz}}$.

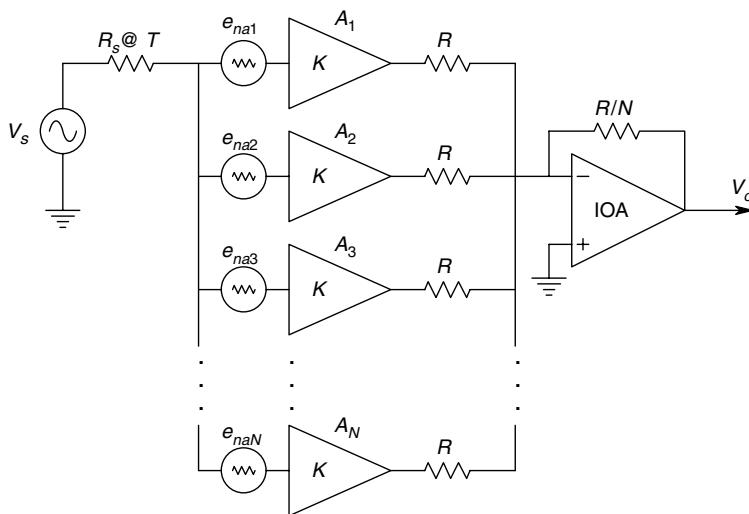


FIGURE P3.12

- (A) Find an expression for the MS output SNR.
- (B) Now assume the resistors R also make Johnson noise. Repeat (A).
- 3.13 The current output of a PIN photodiode, shown in Figure P3.13, operating in the zero voltage mode is conditioned by an OP37 op-amp. The input signal is a

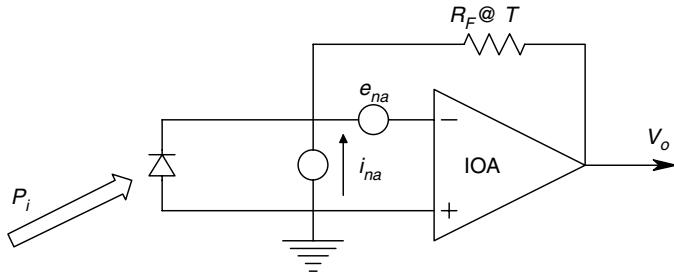


FIGURE P3.13

sinusoidal modulated light power at 640 nm, $P_i(t) = 0.5 P_{pk} [1 + \sin(\omega_o t)]$. Assume that the PD's $I_{rs} = 0.1 \text{ nA}$, $V_T = 0.0259 \text{ V}$, $I_p = [\eta q / (hv)] P_i$, that the PD makes white thermal and shot noise currents of $(4kTg_d + 2qI_{pav}) \text{ MS A/Hz}$, the op-amp has $e_{na} = 3 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 0.4 \text{ pA}/\sqrt{\text{Hz}}$ and R_F makes thermal noise. $4kT = 1.656 \times 10^{-20}$, $R_F = 10^4 \Omega$ and the noise bandwidth, $B = 1 \text{ kHz}$ around $f_o = \omega_o / 2\pi = 10 \text{ kHz}$.

- (A) Write an expression for the MS output SNR.
- (B) Find a numerical value for P_{pk} that will give unity SNR_o .
- (C) What P_{pk} will saturate V_o at 12.5 V?

4

DC Null Measurements

4.1 Introduction

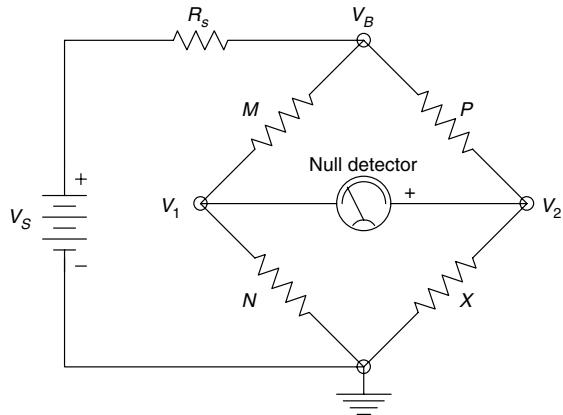
The DC null techniques are used with Wheatstone bridges, Kelvin bridges and potentiometers to obtain increased measurement accuracy based on the fact that the human observer can estimate the occurrence of a dc null on a null meter with greater precision than can be read directly from the meter scale in volts. The accuracy of a dc null measurement system is derived from the known, calibrated accuracy of the resistors making up the bridges, or potentiometer, and has little dependence on the null meter. The human eye is quite sensitive in detecting the microscopic, transient deflections of the pointer of an analog null meter under non-null conditions.

Wheatstone bridges are traditionally used to make precise, accurate resistance measurements (generally in the range of $1\text{--}10^6 \Omega$), or to measure some physical quantity, such as temperature, light intensity or strain, which causes a known change in resistance. Figure 4.1 illustrates the complete circuit of a dc Wheatstone bridge.

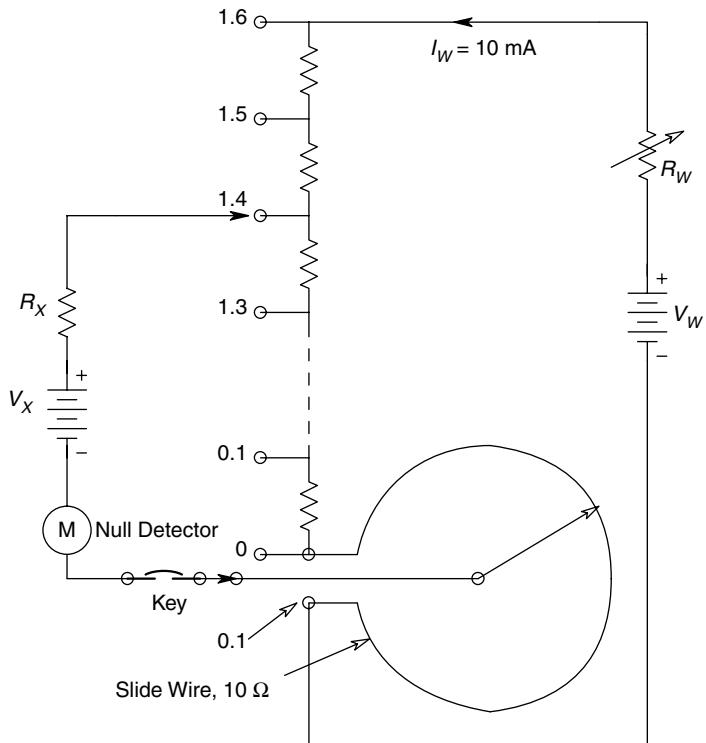
Previously, the bridge null was sensed by a sensitive galvanometer, an electromechanical, current measuring transducer, in which the dc unbalance current of the bridge is passed through a coil which is delicately suspended in a strong permanent magnetic field. The magnetic flux caused by the small coil current interacts with the flux of the permanent magnet and causes a deflection torque which is proportional to the coil current. The magnetic torque acts on the torsion spring of the suspension and produces an angular deflection which is usually sensed optically by reflecting a collimated light beam off a small mirror attached to the galvanometer coil. Needless to say, although sensitive, a mirror galvanometer is not a mechanically robust instrument and consequently, is seldom used in modern dc null measurement systems. Present day dc null sensing devices make use of electronic amplification and display devices. They generally have differential, high impedance ($>10^{12} \Omega$) input stages, so that the unbalance voltage is sensed as an open circuit voltage.

Kelvin bridges are null devices used to measure low values of resistance in the range $10\text{ }\mu\Omega\text{--}10 \Omega$. The null condition for the Kelvin bridge is described in this chapter.

Potentiometers, described in more detail below, are used to make very precise dc voltage measurements in the range of one volt or higher. They are basically accurate voltage dividers that allow differential comparison of their output with an unknown dc EMF. A simplified potentiometer circuit is shown in Figure 4.2. Traditionally, potentiometers were calibrated with the saturated Weston cell, an electrochemical battery of known open circuit voltage or EMF. The same null detection method is used to compare the open circuit voltage output of the calibrated potentiometer with the unknown EMF, similar to Wheatstone and Kelvin bridges.

**FIGURE 4.1**

Basic Wheatstone bridge with dc excitation. The null is detected with a sensitive, center scale, dc voltmeter which draws negligible current.

**FIGURE 4.2**

A basic, dc potentiometer circuit. The null is sensed with an electronic null voltmeter. V_x is the unknown voltage, V_w is the open circuit voltage of the working battery of the potentiometer.

4.2 Wheatstone Bridge Analysis

If we assume that the null detector in Figure 4.1 is a high input impedance voltmeter, then in the null condition is $0 = V_o = V_2 - V_1$. The voltages at the corners of the bridge are given by the voltage divider relations:

$$V_1 = V_B N / (M + N) \quad (4.2)$$

$$V_2 = V_B X / (P + X) \quad (4.3)$$

Hence at null

$$V_o = V_B[X/(P + X) - N/(M + N)] = V_B \frac{XM - NP}{(P + X)(M + N)} \quad (4.4)$$

Since $V_o = 0$ at null, the well-known relation for the unknown resistor, X , is given by:

$$X = NP/M \quad (4.5)$$

The limiting error in determining X at null may be found through application of the limiting error relation from Chapter 1 to equation 4.4, which will give:

$$\Delta X = \left| \frac{\partial X}{\partial M} \Delta M \right| + \left| \frac{\partial X}{\partial P} \Delta P \right| + \left| \frac{\partial X}{\partial N} \Delta N \right| \quad (4.6)$$

Algebraic substitution yields the simple relationship:

$$\frac{\Delta X}{X} = \frac{\Delta M}{M} + \frac{\Delta P}{P} + \frac{\Delta N}{N} \quad (4.7)$$

Thus if, for example, the resistors M , P and N are known to 100 ppm (0.01%), then the LE in X is $\Delta X/X = 0.03\%$. Note that the LE is a worst case error. In practice, the quantities of the form $\Delta Q/Q$ may sum with either sign, reducing the actual $\Delta X/X$.

The null detector resolution can also affect the LE in measuring X . Every null detector has a dead zone of $\pm \Delta V_o$ below which it is impossible to tell whether V_o is non-zero or not. This dead zone is caused by noise and/or the inability of a human operator to tell whether the physical indicator of the null, such as an analog meter pointer, has moved. To see how the dead zone, ΔV_o , affects the LE in X , we examine the Taylor's series for the Wheatstone bridge output relation below:

$$\Delta V_o = (\partial V_o / \partial X) \Delta X_{MIN} \quad (4.8)$$

for the maximum bridge sensitivity case,

$$(\partial V_o / \partial X) = V_B S_V = V_B P / (X + P)^2 \rightarrow V_B / 4X \quad (4.9)$$

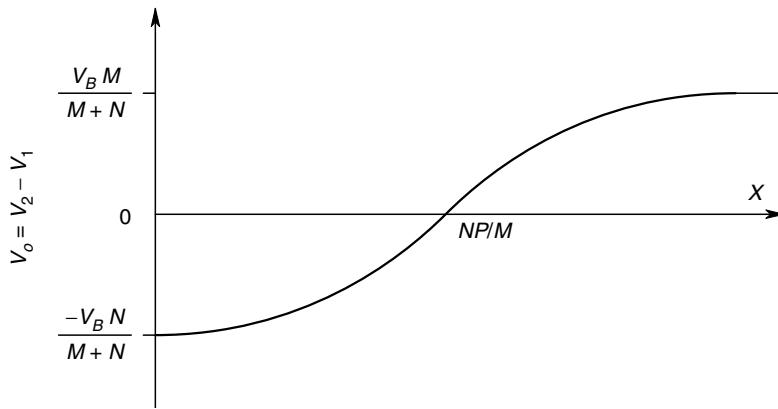
Equation 4.8 for $\partial V_o / \partial X$ is substituted into equation 4.7 and we solve for $\Delta X_{MIN}/X$:

$$\frac{\Delta X_{MIN}}{X} = \frac{\Delta V_o}{V_B} [(P + X)^2 / PX] \rightarrow 4(\Delta V_o / V_B) \quad (4.10)$$

Equation 4.9 for $\Delta X_{MIN}/X$ can be added to equation 4.6 for the LE in X . Generally, this term is numerically much smaller than the $\Delta R/R$ terms in equation 4.6, demonstrating that the null detector is generally the smallest source of error in measuring resistances by a Wheatstone bridge.

Often we are not interested in measuring the value of X *per se*, but rather wish to examine the value of V_o for small changes in X around the value that satisfies equation 4.4. For this case, we plot Eq. equation 4.3 in Figure 4.3. We define a voltage sensitivity for the bridge in equation 4.5:

$$S_V \equiv \frac{\partial(V_o/V_b)}{\partial X} = \frac{P}{(X + P)^2} \quad (4.11)$$

**FIGURE 4.3**

Output voltage of a Wheatstone bridge as a function of the unknown (X) arm.

Clearly, S_V has a maximum when $P=X$. This leads to the conclusion that the maximum sensitivity bridge is one where $P=X=M=N$, which is $S_{VMAX} = 1/4X$ in this case.

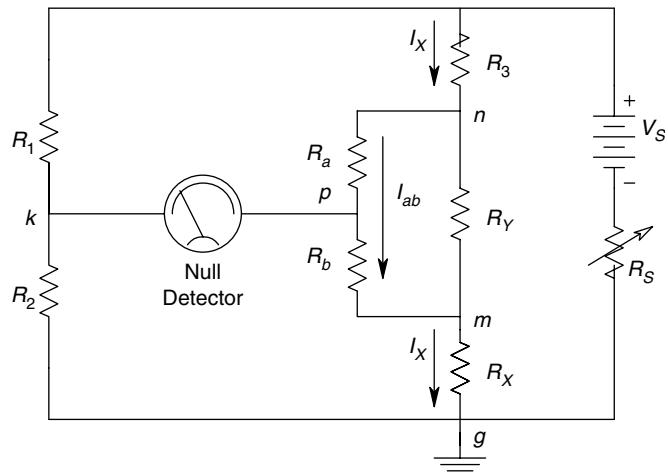
4.3 The Kelvin Bridge

The Kelvin Bridge is a specialized circuit used to measure very low resistances, such as ammeter shunts and motor armatures. The basic circuit for a Kelvin bridge is shown in Figure 4.4. Note that this bridge differs from a Wheatstone bridge by having four resistors in the right half circuit. To derive the conditions for null, it is noted that at null, $V_k = V_p$. V_k can be written by the voltage divider expression as:

$$V_k = V_S R_2 / (R_1 + R_2), \quad (4.12)$$

and

$$V_p = V_{pm} + V_m = I_{ab} R_b + I_x R_X \quad (4.13)$$

**FIGURE 4.4**

Circuit and current paths in a Kelvin bridge used to measure very low resistances, R_X .

where

$$I_{ab} = I_x R_y / (R_a + R_b + R_y) \quad (4.14)$$

and

$$I_x = \frac{V_S}{R_3 + R_x + [R_y(R_a + R_b)/(R_a + R_b + R_y)]} \quad (4.15)$$

Thus at null:

$$V_S R_2 / (R_1 + R_2) = \left[\frac{V_S}{R_3 + R_x + [R_y(R_a + R_b)/(R_a + R_b + R_y)]} \right] \left[R_x + \frac{R_b R_y}{R_a + R_b + R_y} \right] \quad (4.16)$$

Now it turns out that if we let:

$$R_a/R_b = R_1/R_2 = \alpha \quad (4.17)$$

and substitute α in equation 4.15 above, terms cancel and we arrive at the Kelvin bridge null condition:

$$R_x = R_3/\alpha = R_3 R_2/R_1 \quad (4.18)$$

Generally, the resistances in the Kelvin bridge are relatively small, so the battery or dc working voltage source, V_S , must be able to source considerable current. A practical version of a Kelvin bridge is shown in Figure 4.5. Note that the low resistance R_3 is set

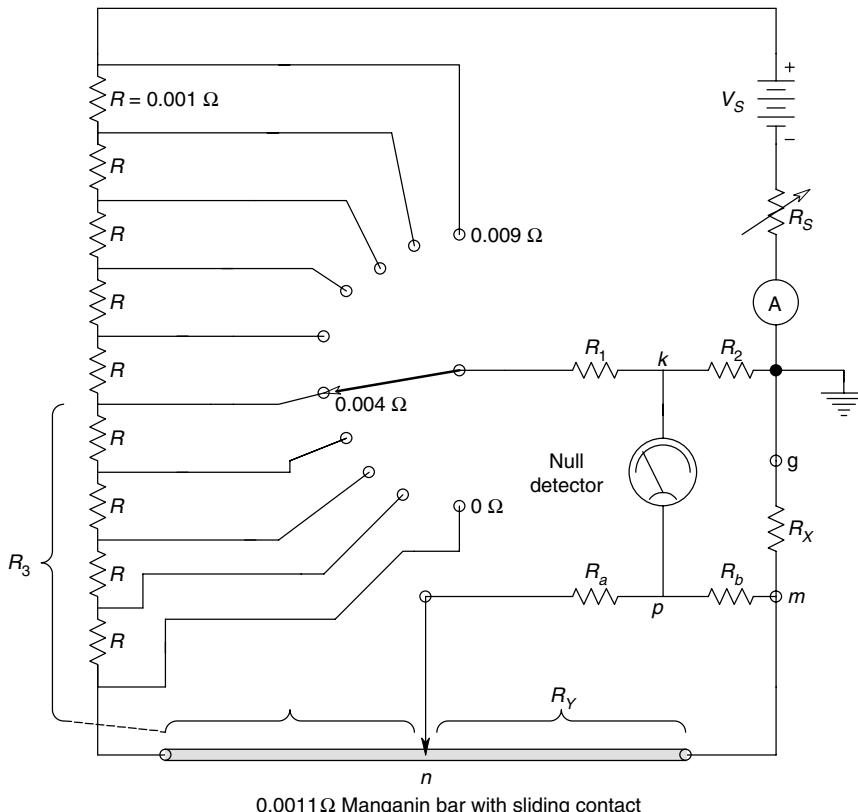


FIGURE 4.5

Practical circuit of a Kelvin bridge.

by a tap switch and 0.001Ω resistors, with interpolation resistance, ΔR_3 , being given by a sliding contact on a manganin bar resistor; the ΔR_3 resistance is proportional to the tap position along the bar.

4.4 The Anderson Constant Current Loop

4.4.1 Introduction

The current loop technology for measuring the ΔR in remote resistive sensors was developed in 1992 at the NASA Dryden Flight Research Center, Edwards, CA, by engineer Karl F. Anderson. It is our purpose in this section to describe how the Anderson Loop works and discuss its benefits and shortcomings, comparing it with circuits such as the traditional, voltage excited Wheatstone bridge. Anderson received a US Patent # 5,731,469, *Constant Current Loop Impedance Measuring System that is Immune to the Effects of Parasitic Impedances* in December 1994.

The advantages of the Anderson circuit architecture are quoted from Anderson (1998):

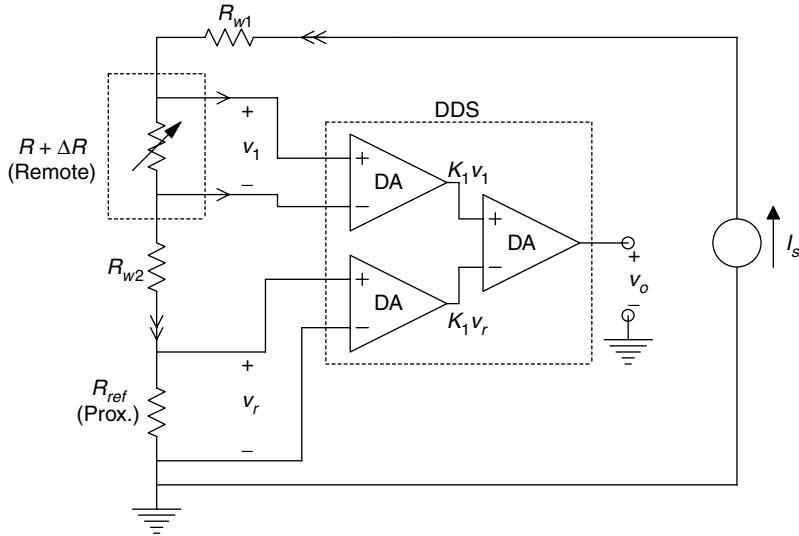
1. Larger and inherently linear outputs that are individually available from each element in a sensor (e.g. a bridge circuit)
2. Power dissipation of the sensor is lowered—beneficial in portable and temperature sensitive applications
3. Even random lead wire and connector (resistance) variations are tolerated without a need for expensive transmitters nearby in hostile environments or the removal of difficult-to-estimate systematic errors from the resulting data
4. Fewer, smaller and less expensive lead wires are required in tight installations
5. Quieter engineering unit readouts are available
6. Smarter sensors can be designed to deliver multi-axis and multi-parameter outputs from simpler sensing structures
7. Stiffer sensors can have twice the frequency response for the same output level
8. Refinements in compensation and calibration can be implemented after installation by adjustments, conveniently away from the sensor, within the signal conditioner

At the heart of the Anderson Loop is a signal conditioning module called the *Dual Differential Subtractor* (DDS). A DDS is shown in Figure 4.6. Note that it uses three high performance, differential instrumentation amplifiers. A DDS is a linear, 3-port voltage amplifier, having two differential inputs and a single ended output. Assuming ideal differential amplifiers (i.e. having ∞ CMRR), the DDS output is:

$$v_o = K_1 K_2 (v_1 - v_2) \quad (4.19)$$

where K_1 is the DM gain of the two input differential amplifiers, K_2 is the DM gain of the output DA, and v_1 and v_2 are DM input signals.

As an illustration of a simple Anderson loop system, consider the circuit of Figure 4.6. A remote sensor transduces the quantity under measurement (QUM) (e.g., temperature) into a change of resistance, ΔR . A current source (here, ideal and referenced to ground) forces a current I_s through a loop that contains the wire resistances (R_{w1} and R_{w2}), the

**FIGURE 4.6**

A simple, single-loop Anderson loop system, showing the dual differential subtractor (DDS) used in all Anderson current loop systems.

sensor ($R + \Delta R$) and a reference resistance, R_{ref} . The DM signal voltage input to the upper DA, according to Ohm's law is $v_1 = I_s(R + \Delta R)$. The DM voltage input to the lower DA is $v_2 = I_s R_{ref}$. Thus the DDS output is:

$$v_o = K_1 K_2 [I_s R(1 + \Delta R/R) - I_s R_{ref}] \quad (4.20)$$

Now if R_{ref} is made equal to R , the output is linearly proportional to ΔR :

$$v_o = K_1 K_2 I_s R(\Delta R/R) \quad (4.21)$$

Note that the dc common-mode voltage for the upper DA is:

$$v_{1CM} = I_s(R_{w2} + R_{ref})$$

Due to its very high CMRR, the upper DA has negligible output from v_{1CM} .

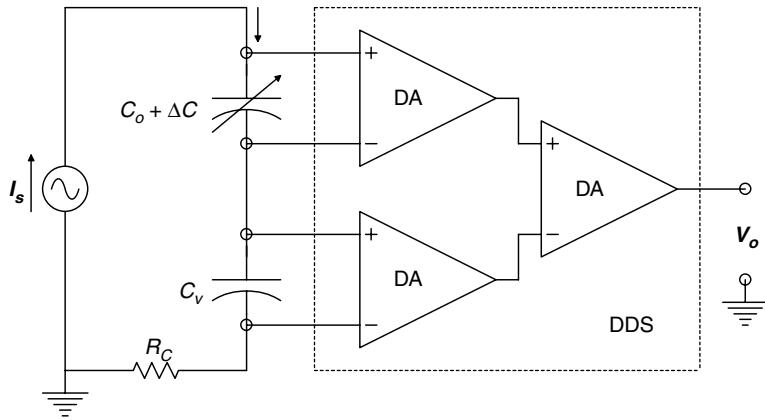
In Figure 4.7 we illustrate an Anderson loop applied to a capacitive sensor in which the QUM causes a change in capacitance. An AC current excitation must be used. By inspection, an expression for the output voltage of the DDS can be written:

$$V_o = I_s \left\{ \frac{1}{j\omega(C_o + \Delta C)} - \frac{1}{j\omega C_v} \right\} \quad (4.22)$$

Now, making $C_v = C_o$ and solving for small $\Delta C/C_o$:

$$V_o = I_s \left\{ \frac{1 - \Delta C/C_o}{j\omega C_o} - \frac{1}{j\omega C_o} \right\} = \frac{I_s}{\omega C_o} \left(\frac{\Delta C}{C_o} \right) \angle -270^\circ \quad (4.23)$$

Note that the magnitude of V_o is proportional to ΔC and calibration depends on knowing I_s , its frequency, ω and C_o .

**FIGURE 4.7**

An Anderson loop applied to a capacitive sensor. See text for analysis.

4.4.2 The Anderson Loop Applied to Groups of Sensors

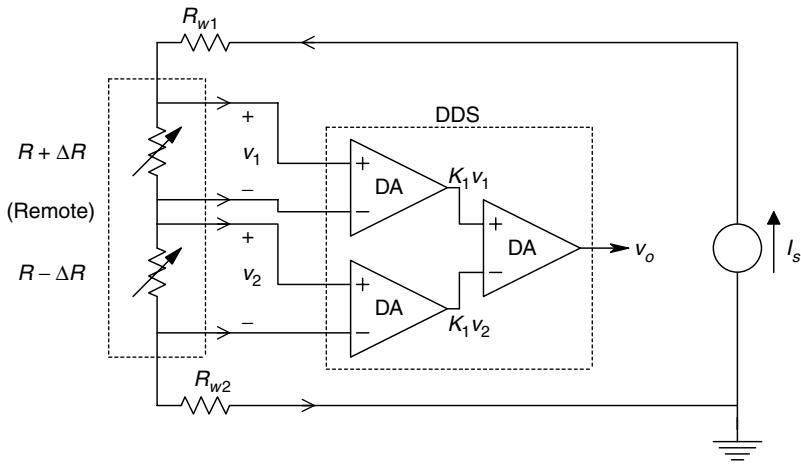
Figure 4.8 illustrates a dual differential, resistive sensor in an Anderson loop. The QUM causes the resistance of the upper sensor to increase the same amount that the resistance of the lower sensor decreases. Again, the ubiquitous DDS is used. The DM input voltages are:

$$v_1 = I_s(R + \Delta R) \quad (4.24a)$$

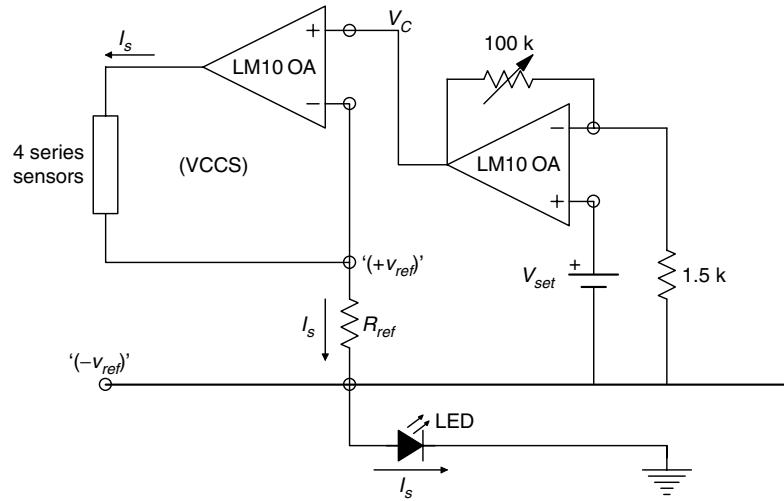
$$v_2 = I_s(R - \Delta R) \quad (4.23b)$$

The DDS output voltage is:

$$v_o = K_2[K_1 v_1 - K_1 v_2] = K_1 K_2 I_s R (2\Delta R/R) \quad (4.25)$$

**FIGURE 4.8**

An Anderson loop system is applied to a remote, differential, resistive sensor. The four wires from the sensor to the DDS carry negligible current.

**FIGURE 4.9**

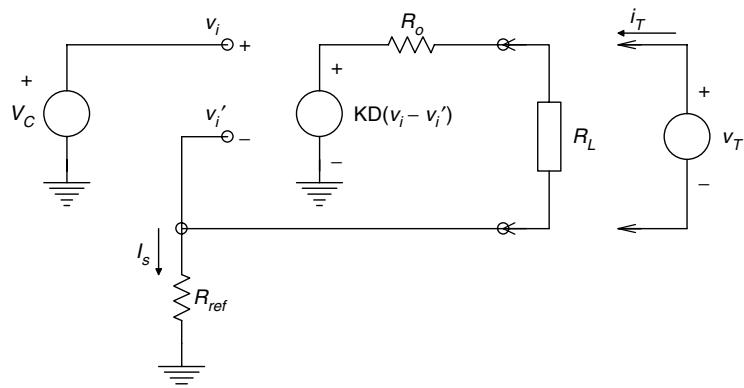
A 'stiff' current source used by Anderson.

The dc CM voltages are:

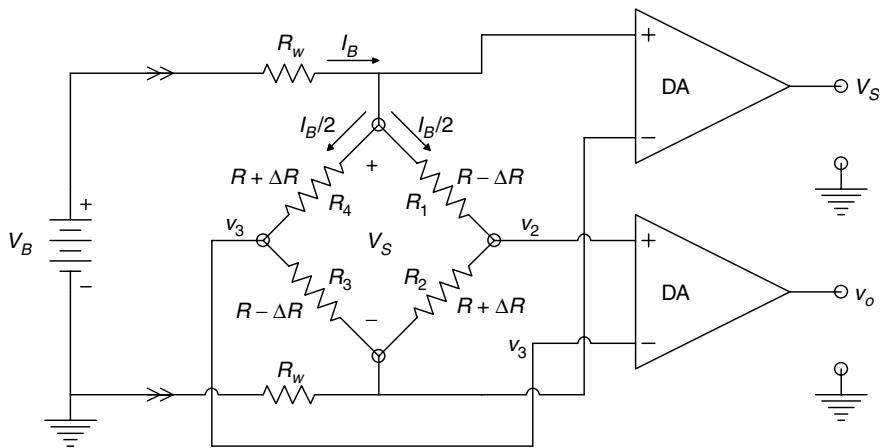
$$v_{1CM} = I_s[(R - \Delta R) + R_{w2}] \quad (4.26a)$$

$$v_{2CM} = I_s R_{w2} \quad (4.25b)$$

Before addressing the problem of adapting a four-sensor Wheatstone bridge to an Anderson current loop, let us examine the problem of designing a stiff current source for the Anderson loop. Figure 4.9 illustrates a simplified schematic of an op-amp VCCS such as the one used by Anderson. Notice that the series sensing elements in the loop are floating. That is, not one is directly conductively coupled to ground. The loop current, I_s , must return to ground through the common (feedback) resistor, R_{ref} , and in this case, also an LED. If it is assumed that the opamp is ideal, then the voltage, $(+v_{ref})$, must equal the control voltage, V_C . Thus by Ohm's law, $I_s = V_C/R_{ref}$. To see how 'stiff' this current source is, a more detailed circuit of the VCCS must be examined using a simple model for a finite gain, real op-amp. Figure 4.10 illustrates such a circuit. Note that the

**FIGURE 4.10**

Model for the current source of Figure 4.9.

**FIGURE 4.11**

Schematic of a remote 4-active-arm Wheatstone bridge. Two DAs are used to condition the bridge output, v_o , and the dc voltage across the bridge, V_S .

output resistance of the op-amp without feedback, R_o , is of the order of 25Ω . Ideally, a VCCS should have an output resistance of $>1M\Omega$. To calculate the output resistance of the VCCS, R_L must be replaced with a small signal test voltage source, v_T . Also, set $V_C=0$. The output resistance of the VCCS is, by Ohm's law, $R_{out}=v_T/i_T$. Simple circuit analysis reveals that $R_{out}=R_o+R_{ref}(1+K_D)\cong R_{ref}K_D$.

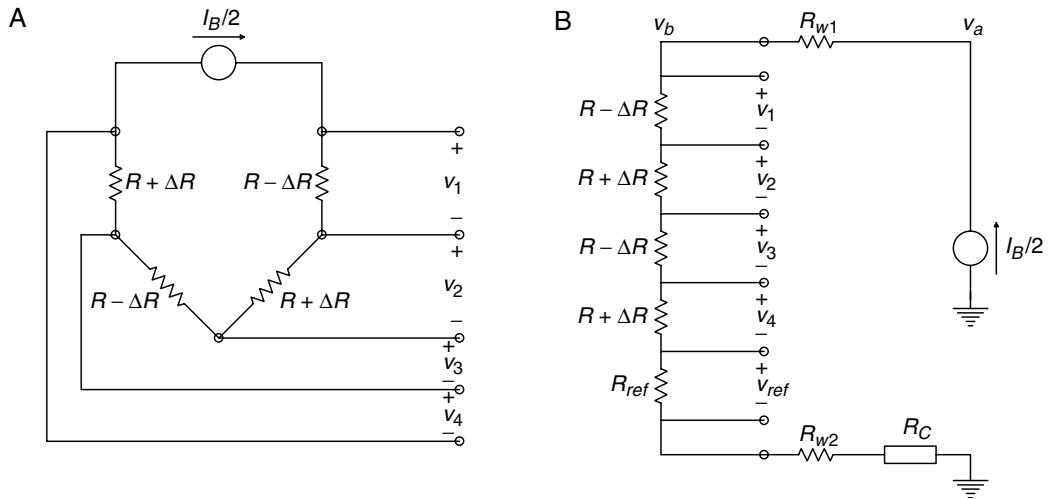
Consider the typical values of $R_{ref}=300$ and $K_D=10^5$, thus $R_{out}=30M\Omega$, which is adequate. It is also easy to show that this VCCS has a transconductance given by $G_M=I_s/V_C\cong 1/R_{ref}S$. A major drawback of the simple Anderson VCCS is that the load is 'floating' (i.e. its elements must be insulated from ground).

The schematic of a standard, 4-active-arm, voltage excited Wheatstone bridge is illustrated in Figure 4.11. This bridge configuration is typical of a force or pressure sensor using four, unbonded strain gauges (Figure 6.4). Assume that the bridge excitation source and the differential amplifier are remote from the bridge. The DA has a very high input impedance (generally over $10^8\Omega$), so it is possible to ignore any small resistance of the wires leading from the bridge to the DA. V_S , the voltage across the bridge is less, by the voltage dropped across the resistances of the wires supplying the bridge, than the excitation voltage, V_B . Elementary circuit theory says that $I_B=V_B/(2R_W+R)$ and thus $V_S=V_BR/(2R_W+R)$. The bridge output can easily be found using voltage divider relations:

$$v_o=(v_2-v_3)=V_S\left[\frac{R+\Delta R}{2R}-\frac{R-\Delta R}{2R}\right]=V_S=\frac{\Delta R}{R} \quad (4.27)$$

The bridge calibration depends on V_S , which in turn, is dependent on R_W which is temperature sensitive. The uncertainty in V_S can be removed if this voltage is measured by another DA whose output is compared to the desired value, the error being used to adjust V_B .

To examine the evolution of the 4-active-arm Anderson circuit derived from a 4-active-arm Wheatstone bridge, examine the circuit of Figure 4.12A. The top of the Wheatstone bridge is opened, and a current source, $I_B/2$, is inserted ($I_B/2$ is used because this is the current in each of the four Wheatstone bridge active elements.) Note that in this circuit, we have lost the ground from the bottom of the bridge. The Anderson circuit is redrawn in Figure 4.12B, illustrating the floating (ungrounded) nature of the Anderson

**FIGURE 4.12**

Steps in 'Andersonizing' a remote 4-active-arm Wheatstone bridge. (A) The bridge is powered by a floating dc current source. (B) The bridge circuit is redrawn.

loop. Note that a reference resistance, R_{ref} , is added, and an additional common-mode resistance, R_C , may also be present.

Figure 4.13 illustrates the final 'Andersonized' 4-active-arm sensor system with the dual differential subtractors (DDSs) added. Note that no less than 15 precision DAs are required with high CMRRs! The CM input voltage for the top DA is:

$$v_{1CM} = (I_B/2)[3R + \Delta R + R_{ref} + R_C]$$

The CM voltage for the DA of the second sensor is:

$$v_{2CM} = (I_B/2)[2R + R_{ref} + R_C]$$

and so on.

The reference voltage is $v_r = (I_B/2)R_{ref}$ and the CM voltage for the four reference input DAs is $v_{rCM} = (I_B/2)R_C$. Assume that the DAs are ideal (i.e. their input resistances, bandwidths and CMRRs are infinite, they have zero output resistance, and their differential gains are finite). So, for the top DDS:

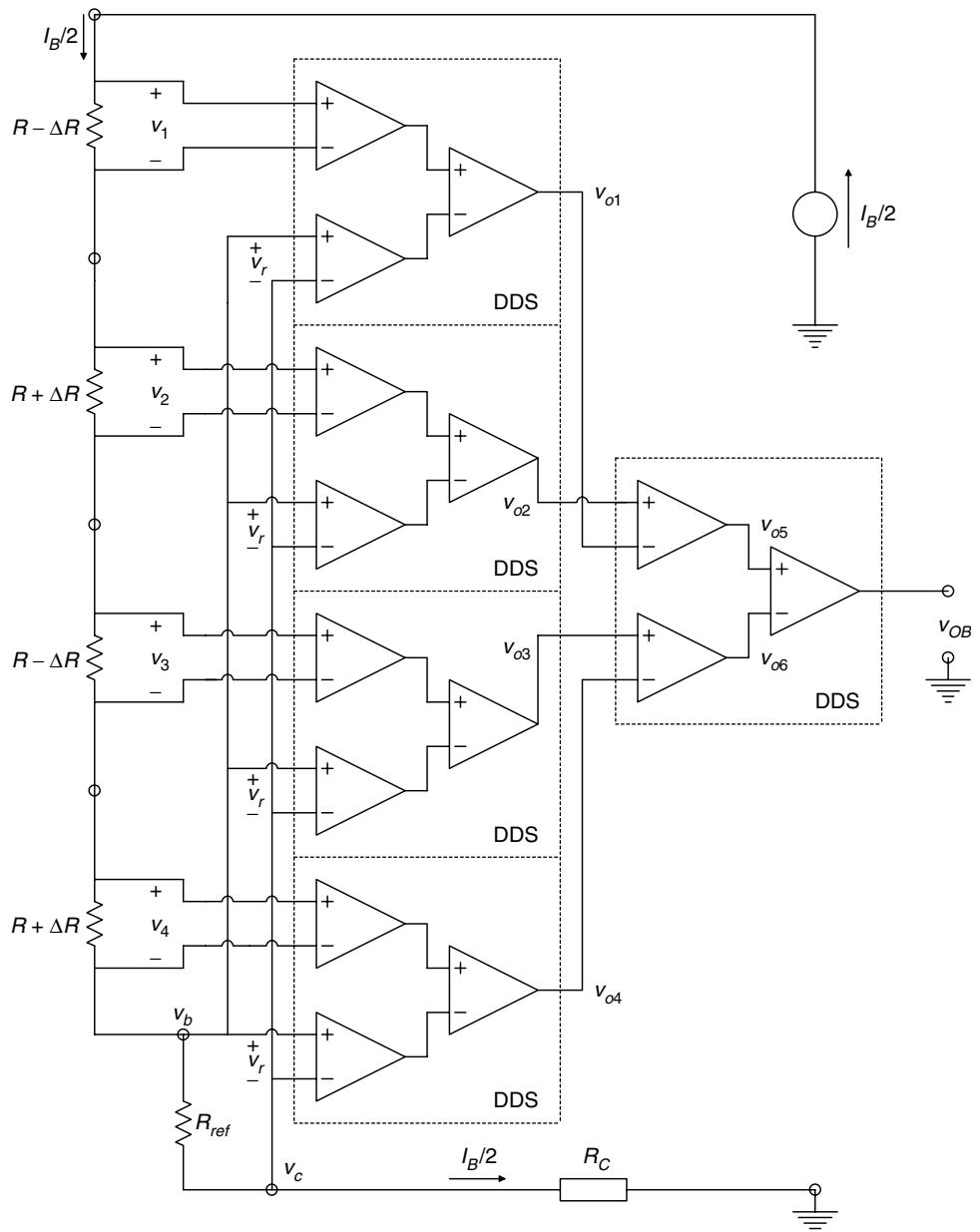
$$v_1 = (I_B/2)(R - \Delta R) \quad (4.28)$$

$$v_r = (I_B/2)R_{ref} \quad (4.29)$$

$$v_{o1} = K_1 K_2 (I_B/2)R[1 - \Delta R/R + R_{ref}/R] \quad (4.30)$$

Similarly, we can show:

$$v_{o2} = K_1 K_2 (I_B/2)R[1 + \Delta R/R + R_{ref}/R] = v_{o4} \quad (4.31)$$

**FIGURE 4.13**

The final ‘Andersonized’, remote, 4-active-arm Wheatstone bridge circuit. Note that a total of five DDS elements are required to obtain an output voltage proportional to ΔR . Compare this circuit to the conventional bridge circuit of Figure 4.11.

and

$$v_{o3} = v_{o1} \quad (4.32)$$

The net output of the ‘Andersonized’ 4-arm bridge is formed by the output DDS:

$$v_{oB} = K_1^2 K_2^2 I_B (2\Delta R) \quad (4.33)$$

Using the DDS modules, there is no need to make $R_{ref}=R$ to obtain equation 4.32, but this condition will allow extraction of $v_{o1}-v_{o4}$ without a dc offset, for example, $v_{o2} = K_1 K_2 (I_B/2) R [\Delta R/R]$ etc.

Note that the complexity of the circuit of Figure 4.13 is not required to measure $\Delta R/R$. Just one sensor works, as shown in Figure 4.6. If R_{ref} is made a sensor without an input and placed adjacent to the active element, then thermal compensation will be inherent in the Anderson circuit. Note that a two-active, or four-active arm Wheatstone bridge is inherently thermally stabilized, as well.

4.4.3 Conclusion

The effectiveness of the Anderson loop architecture is predicated on the use of the DDS module in which all the DAs all have high CMRRs, generally exceeding 120 dB ($>10^6$). However, this requirement is also valid for a conventional Wheatstone bridge supplied by a voltage source. Calibration of the Wheatstone bridge depends on the existence of a known, stable potential difference between the top and bottom of the bridge. Similarly, calibration of an Anderson loop sensor requires the current source (I_s or $I_B/2$) to be stiff, known and stable. Just as in the case of a conventional Wheatstone bridge, an Anderson loop system can be excited sinusoidally, permitting signal amplification out of the $1/f$ range of amplifier noise and allowing synchronous detection.

4.5 Potentiometers

Figure 4.14 illustrates the simplified circuit of a potentiometer. Like Wheatstone and Kelvin Bridges, potentiometers are dc null instruments. Nowadays, they generally find application in standards laboratories. They are used as a means of precision calibrating secondary dc voltage standards. Originally, potentiometers were used with thermocouples to obtain precise temperature measurements and in electrochemical laboratories, to measure cell EMFs. Most of these applications have now been taken over by precision electronic voltmeters.

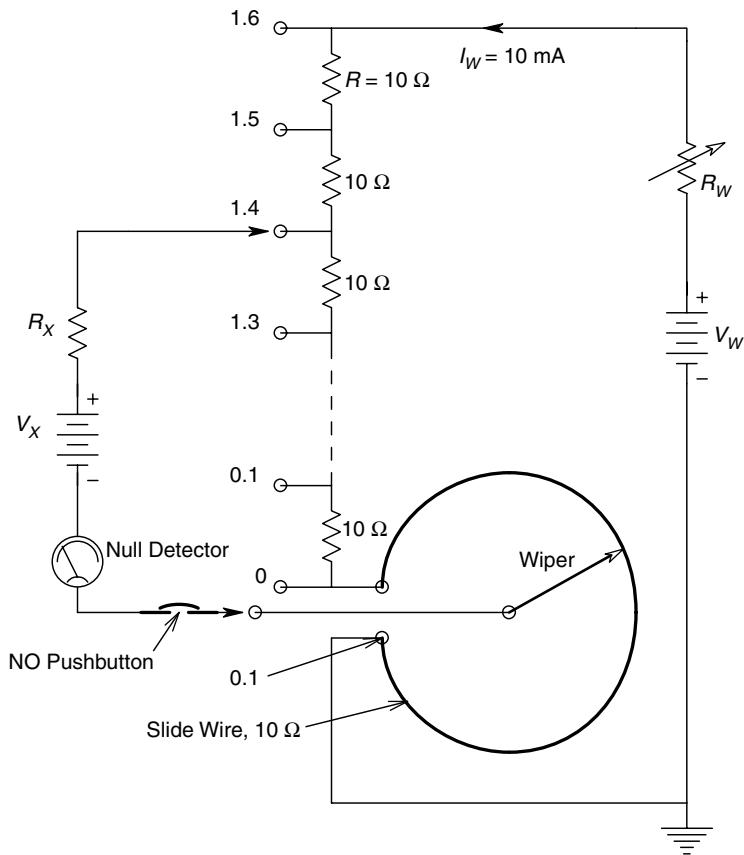
For example, the Leeds & Northrup Model 7556-1 potentiometer has three ranges. The accuracy of each range is given in Table 4.1.

To measure high voltages, a precision voltage divider called a *volt box* is used. For example, the L&N Model 7594-2 volt box was rated at 0–1500 V, with an accuracy of $\pm 0.0025\%$. It had 25 taps and was rated at $667 \Omega/V$ on the 1500 V range, quite a lossy device, compared with modern voltmeter sensitivities.

Potentiometers are often calibrated using Weston Standard Cells, an electrochemical battery which, until fairly recently, was used to define the volt EMF. The Weston cell was developed in 1892 as an EMF standard. Two types of Weston cells exist, saturated and unsaturated. The more widely used saturated (normal) cell is illustrated in Figure 1.6. A Weston cell is generally used in a thermostat controlled oil bath to maintain its internal temperature to within 0.01°C . This is necessary because of the rather high EMF tempco of the normal cell, $-40 \mu\text{V}/^\circ\text{C}$. The EMF of the saturated Weston cell is given by:

$$\text{EMF} = 1.01830 - 4.06 \times 10^{-5}(T - 20) - 9.5 \times 10^{-7}(T - 20)^2 - 10^{-8}(T - 20)^3 \quad (4.34)$$

This EMF is in International Volts; the EMF in absolute volts is 1.01864 at 20°C . Thus the volt used to be defined as 1/1.01830 of the Weston normal cell at 20°C . Weston cells are

**FIGURE 4.14**

Circuit of a simple potentiometer used for precision, comparative measurements of EMFs in the 0–1.6 V range.

TABLE 4.1

Ranges and accuracy of the L&N 7556-1 potentiometer

Range	Accuracy
0–1.61100 V	$\pm(0.0005\% + 1 \mu\text{V})$
0–0.16110 V	$\pm(0.0015\% + 0.1 \mu\text{V})$
0–0.01611 V	$\pm 0.0025\% + 20 \text{ nV}$

very delicate and subject to instabilities if their EMFs, if subjected to large or sudden temperature changes. They also should never have currents in excess of 100 μA drawn from them for any period of time. They also age and should be recalibrated against freshly prepared cells every year, or against Josephson junction primary EMF standards.

Calibration of a potentiometer is done by first adjusting the potentiometer reading to the calculated EMF of the standard cell, then adjusting R_W to obtain a null in the galvanometer or electronic null detector. Next, the unknown EMF is connected to the potentiometer and the null detector, and the calibrated potentiometer output voltage is adjusted to obtain a null.

4.6 Chapter Summary

In this Chapter we have discussed three basic measurement systems that make use of the principle of nulling or zeroing a dc voltage in order to obtain precision measurement of a resistance or an EMF—the Wheatstone bridge, the Kelvin bridge and the potentiometer. The accuracy attainable in a null measurement technique generally depends on the use of accurate resistors, rather than the accurate measurement of a dc voltage or current, *per se*.

The Anderson loop technology was introduced and shown to use electronic constant current sources and a basic three-DA configuration called the dual differential subtractor (DDS), in conjunction with sensors whose resistance changes proportional to the QUM. The Anderson loop allows the voltage output to be linear in terms of the ΔR of the sensor. It is shown to be an alternative to the Wheatstone bridge in certain resistive sensor applications.

Problems

- 4.1 Two ideal op-amps are used to linearize the output of a single active arm Wheatstone bridge, as shown in Figure P4.1. The resistance of the variable resistor, R_V , is given by $R_V = R + \Delta R$, where $\Delta R > -R$.

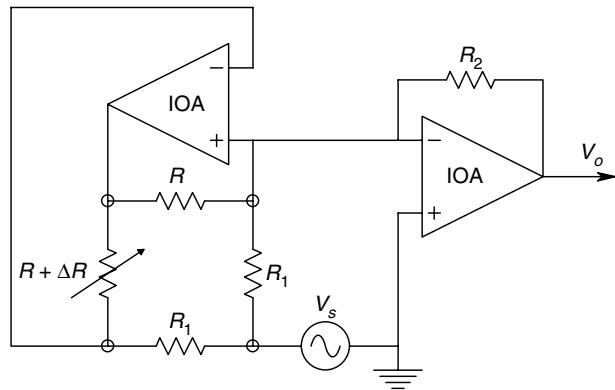


FIGURE P4.1

- (A) Find an algebraic expression for $V_o = f(V_s, R, R_1, R_2, R)$. Hint: Write node equations for V_1 and V_2 ; note that by the ideal op-amp assumption, $V_1 = V_2 = 0$.
- (B) Let $V_s = 5\text{ V}$, $R = R_1 = R_2 = 10^3\text{ }\Omega$. Find the permissible range in R_V so that neither op-amp's output reaches $\pm 12\text{ V}$ (saturation).
- 4.2 A dc Wheatstone bridge contains an RTD whose resistance around 25°C is approximated by:

$$R(T) = R_{25}[1 + \alpha(T - 25)]$$
. The smallest ΔV_o the detector can resolve is 100 nV . $V_s = 3.1\text{ V}$, $R = R_{25} = 300\text{ }\Omega$ and $\alpha = 0.004$. Find the just resolvable ΔT under these conditions (Figure P4.2).

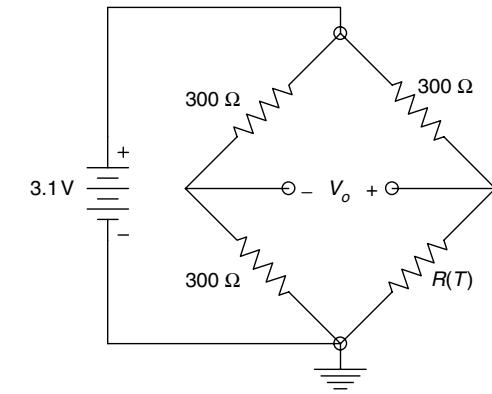


FIGURE P4.2

- 4.3 A double sideband, suppressed carrier modulated (DSBSCM) signal is generated in many sensor systems. A DSBSCM signal is basically the product of two sine waves with different frequencies. In the Wheatstone bridge circuit of Figure P4.3, two strain gauges are subject to a low frequency sinusoidal strain such that: $\delta R(t) = \Delta R \sin(2t)$. $\Delta R = 0.03 \Omega$, $R = 300 \Omega$, $v_b(t) = V_B \sin(1000t)$ and $V_B = 5 \text{ Vpk}$. The DA's gain is 10^3 . Find an expression for $v_o(t)$ and show that it can be resolved into the algebraic sum of sum and difference frequencies.

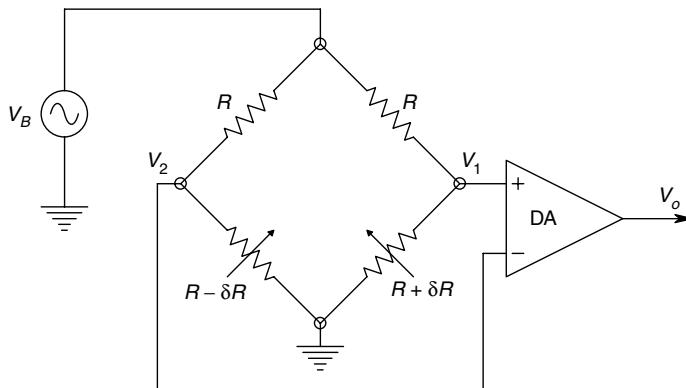


FIGURE P4.3

- 4.4 A Wheatstone bridge (Figure P4.4) is powered with a 1 V RMS, 200Hz sinusoidal source. Fixed resistors A and B are 1000Ω with LEs of 10 ppm. The value of variable resistor, D, is known to 100 ppm. At null, $D = 100 \Omega$.

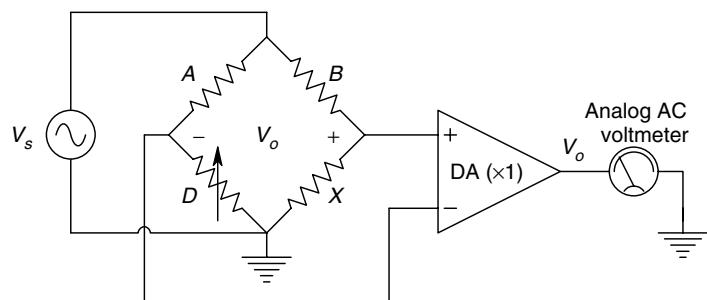


FIGURE P4.4

- (A) Find the nominal value of X.
 (B) The ac null detector can just resolve $\Delta V_o = 1 \mu\text{VRMS}$. The DA is ideal with gain = 1. Calculate the limiting error in X, considering the LEs in A, B, D and V_o .

4.5 Consider Figure P4.5 for the basic Anderson DDS circuit. Assume that the identical input DA's outputs are described by the equations:

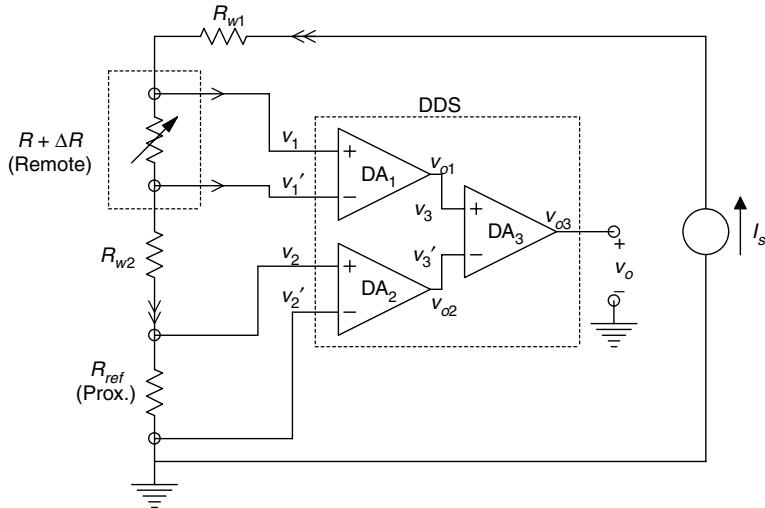


FIGURE P4.5

$$v_{o1} = A_{D1}v_{1d} + A_{C1}v_{1c}, \text{ CMRR}_1 \equiv A_{D1}/A_{C1}$$

$$v_{o2} = A_{D2}v_{2d} + A_{C2}v_{2c}, \text{ CMRR}_2 \equiv A_{D2}/A_{C2}$$

Let $A_{D1} = A_{D2}$, $A_{C1} = A_{C2}$, so that $\text{CMRR}_1 = \text{CMRR}_2$. The output subtractor DA's output is described by:

$$v_o = v_{o3} = A_{D3}v_{3d} + A_{C3}v_{2d}, \text{ CMRR}_3 \equiv A_{D3}/A_{C3}$$

Note that $A_{C1} = A_{D1}/\text{CMRR}_1$, $A_{C1} = A_{D1}/\text{CMRR}_1$, etc. and R_{ref} is made equal to R_o of the sensor.

- (A) Write expressions for v_1 , v'_1 , v_2 and v'_2 in terms of the circuit parameters and ΔR .
 (B) Find expressions for the DA outputs, v_{o1} and v_{o2} .
 (C) Find an expression for $v_o = v_{o3}$. Give $v_o = v'_o$ when all of the CMRRs $\rightarrow \infty$.
 (D) Give an expression for the Anderson DDS circuit's voltage error, $v_e = v_o - v'_o$ for finite CMRR_1 and CMRR_3 .

4.6. Figure P4.6 (after Stout, 1960) illustrates the schematic of a potentiometer. R_W is used to adjust the working current, I_W , in the loop to 50.000 mA when setting the pot up with a standard cell. The standard cell's EMF is defined as 1.01830 V at 20°C. A separate, high resolution, standard cell slide wire is used to initially set I_W . The EMF slide wire has 100 divisions over its range, and can be read to 0.1 division. The maximum range of E_X that can be read is 1.7 V.

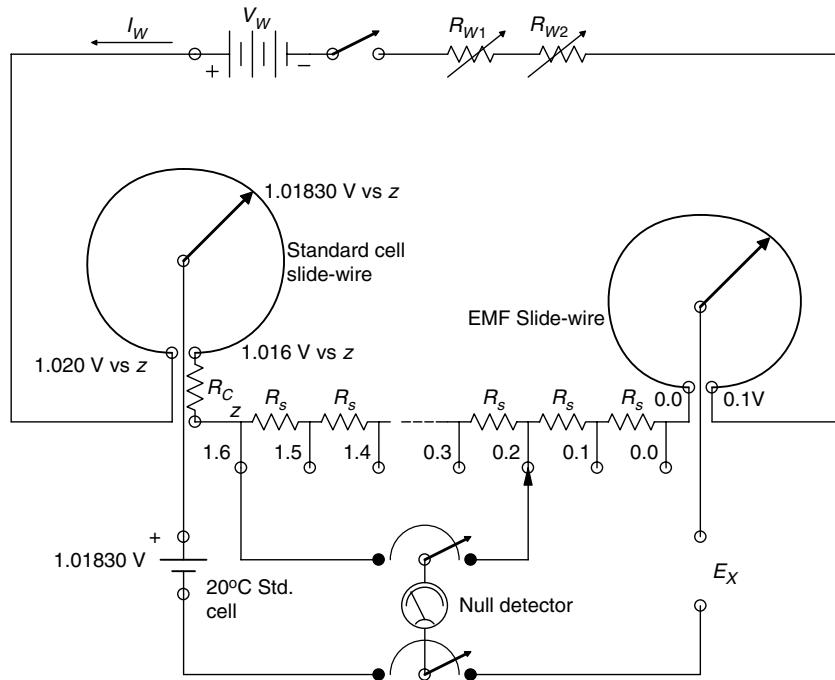


FIGURE P4.6

- Find the smallest ΔE_X that can be read.
- Give ohm values for each of the 16 fixed resistors, R_s , and the EMF slide wire resistance, the standard cell slide wire resistance and R_c .
- Show the polarity of the E_X terminals.

4.7 A Kelvin Bridge schematic is shown in Figure P4.7. At null, $V_k = V_p$.

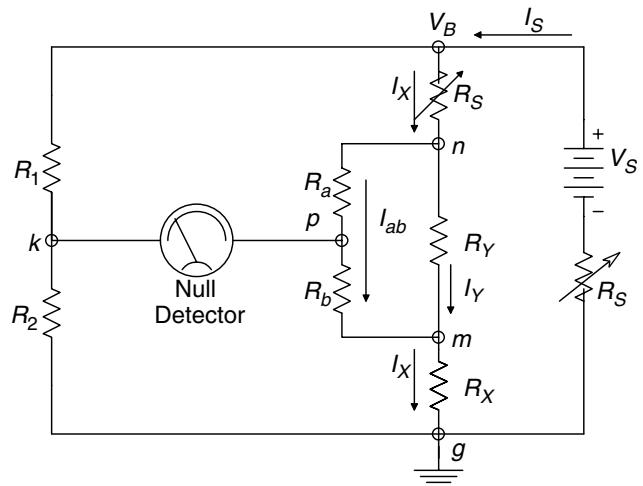


FIGURE P4.7

- Find expressions for V_k , I_X and I_{ab} .
- Derive an expression for R_X at null (note that it has two terms).

- (C) At null: $R_1 = 100.0 \Omega$, $R_2 = 1000 \Omega$, $R_Y = 0.1000 \Omega$, $R_3 = 0.00377 \Omega$, $R_b = 99.2 \Omega$, $R_a = 1000.6 \Omega$. Calculate R_X using the simple first term in the equation of part B). Calculate R_X using the whole expression from B).
- (D) Calculate I_X assuming $V_B = 1.500 \text{ V}$. Find the power dissipated in R_X .
- 4.8 Consider the Anderson DDS used to measure the small ΔC of a capacitive sensor, as shown in text Figure 4.7. Consider the DAs of the DDS to have flat frequency responses (scalar gains) with outputs given by:

$$\begin{aligned} v_{o1} &= A_{D1}v_{1d} + A_{C1}v_{1c}, \text{ CMRR}_1 \equiv A_{D1}/A_{C1} \\ v_{o2} &= A_{D2}v_{2d} + A_{C2}v_{2c}, \text{ CMRR}_2 \equiv A_{D2}/A_{C2} \end{aligned}$$

Let $A_{D1} = A_{D2}$, $A_{C1} = A_{C2}$, so that $\text{CMRR}_1 = \text{CMRR}_2$. The output subtractor DA's output is described by:

$$v_o = v_{o3} = A_{D3}v_{3d} + A_{C3}v_{2d}, \text{ CMRR}_3 \equiv A_{D3}/A_{C3}$$

Note that $A_{C1} = A_{D1}/\text{CMRR}_1$, etc. and C_v is made equal to C_o of the sensor.

- (A) Write phasor expressions for V_1 , V'_1 , V_2 and V'_2 in terms of the circuit parameters and ΔC . Give I_S zero phase.
- (B) Find expressions for the phasor DA outputs, V_{o1} and V_{o2} .
- (C) Find an expression for $V_o = V_{o3}$. Give $V_o = V'_o$ when all of the CMRRs $\rightarrow \infty$.
- (D) Give an expression for the Anderson DDS circuit voltage error, $V_\varepsilon = V_o - V'_o$ for finite CMRR_1 and CMRR_3 .

5

AC Null Measurements

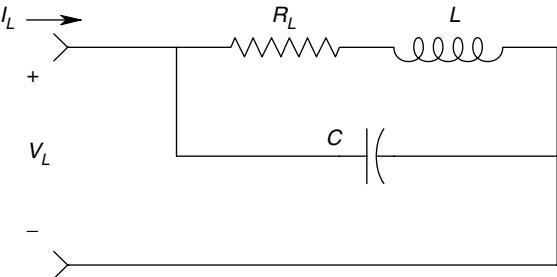
5.1 Introduction

In this chapter, we examine measurement techniques that make use of sensing an ac null. These include, but are not limited to, ac operation of (resistive) Wheatstone bridges, various types of ac bridges used to measure capacitance, *capacitor dissipation factor* (D), inductance, *inductor quality factor* (Q), mutual inductance and the *small signal transconductance* (g_m) of bipolar junction transistors (BJT) and field effect transistors. As in the case of dc null methods, ac methods are used to obtain accurate measurements of component values based on the accuracy of the bridge components.

It should be noted that in the real world of circuit components, there is no such thing as a pure or ideal resistor, capacitor or inductor. All real world devices have parasitic parameters associated with them. In some cases, these parasitic components may be treated as lumped parameter circuits, while in other situations, they are best described as distributed parameter networks, similar to transmission lines. For example, depending on the frequency of the ac voltage across a resistor, the resistor may appear to have a pure inductance in series with a resistor in parallel with a capacitor, and distributed capacitance to ground along its length. The resistance of a resistor has a value which is an increasing function of frequency due to skin effect at VHF – UHF. The presence of parasitic components at high frequencies makes the operation of bridges and null circuits at high frequencies more prone to errors. To minimize the effects of parasitic components, most simple ac bridges operate at 1 kHz.

5.2 Inductor Equivalent Circuits

Most practical inductors are made from one or more turns of a conductor, wound as a solenoid, either on an air core, or a ferromagnetic core, made of ferrite ceramic or laminated iron. If the coil is wound around a ferromagnetic ring or ‘doughnut’, it is called a toroidal inductor. The use of ferromagnetic cores concentrates the magnetic flux and produces a higher inductance than would be attainable with an air core and the same coil geometry. Since the conductor used to wind an inductor has a finite resistance, the simplest, low frequency equivalent circuit of a practical inductor is a resistor in series with a pure inductance. At very high frequencies, the small, stray capacitance between adjacent turns of the inductor coil produces a complex, distributed parameter, RLC circuit. If the coil is wound with several layers of turns, capacitance between the layers, as well as between adjacent turns and inner turns to the core, produces a complex equivalent circuit of the inductor at very high frequencies.

**FIGURE 5.1**

High frequency, lumped parameter model for an inductor.

One compromise to the problem of representing an inductor realistically at high frequencies is shown in Figure 5.1. Here, the distributed capacitance between windings is modelled by a single, equivalent, lumped capacitor in parallel with the series R - L circuit. This model is probably reasonably valid in the low audio frequency range where most ac inductance bridges operate. To characterize and measure inductors at radio frequencies where the lumped model is not valid, we generally use a *Q meter*, which is discussed below.

The quality factor, or Q , of an inductor is defined at a given frequency as the ratio of the inductor's inductive reactance to the real part of its impedance. The higher the Q , the 'purer' the inductor and the lower are its losses. High Q inductors allow us to build more sharply tuned RLC frequency selective circuits as compared to low Q inductors of the same inductance. For a given inductance, high Q inductors are generally more expensive. Note that Q is a dimensionless function of frequency. For a simple series R - L circuit, the Q is given by:

$$Q \equiv \frac{X_L}{R_L} = \frac{\omega_L}{R_L} \quad (5.1)$$

The Q of the RLC inductor model of Figure 5.1 may be found by finding its impedance:

$$\begin{aligned} Z_L(j\omega) &= \frac{(j\omega L + R)}{1 + j\omega RC + (j\omega)^2 LC} \\ &= \frac{R}{(1 - \omega^2 LC) + (\omega RC)^2} + j \frac{\omega L[(1 - \omega^2 LC) - (CR^2/L)]}{(1 - \omega^2 LC) + (\omega RC)^2} \end{aligned} \quad (5.2)$$

From the basic definition of Q given in equation 5.1 and the impedance given by equation 5.2, the Q of the series model inductor can be written as:

$$Q_s(\omega) = \frac{\omega L[(1 - \omega^2 LC) - (CR^2/L)]}{R} \quad (5.3)$$

Interestingly, $Q_s(\omega)$ rises to a maximum at $\omega = \omega(\sqrt{3})/3$ and then drops to zero at ω_s , the resonance frequency of the lumped RLC model of the inductor at the resonance frequency, $Z_L(j\omega) = \text{Real}$. From equation 5.2, this is seen to occur when:

$$\omega_s = \frac{1}{\sqrt{LC}} - \frac{R^2}{L^2} \text{ r/s} \quad (5.4)$$

Although the lumped RLC model for the inductor is crude, its behavior does mimic that observed in real inductors at high frequencies. Their Q does increase to a maximum and

then decreases to zero as the frequency is further increased. Some practical inductors exhibit multiple peaks in their Q curves at high frequencies due to the distributed nature of the stray capacitance. At 1 kHz, the frequency used in most ac bridges used to measure inductors, we are operating well to the left of the peak of the $Q(\omega)$ curve and capacitive effects are generally second order.

5.3 Capacitor Equivalent Circuits

There are many physical types of capacitors having a wide variety of geometries and using many different types of dielectrics, all depending on the selected application. The basic capacitor consists of a pair of parallel metal plates, separated by a dielectric (insulating substance), which can be vacuum, air, sulfur hexafluoride, oil, oil impregnated paper, glass, mica, metal oxides, or various plastics or ceramics.

Each dielectric has its own unique properties of dc leakage, dielectric constant, losses at high frequencies, temperature coefficient, etc. At high frequencies, capacitors can be modelled by a pure capacitance surrounded by parasitic inductors due to leads, and resistances modelling dielectric power loss and ohmic leakage. Figure 5.2 illustrates a general, lumped parameter, high frequency equivalent circuit for a non-polarized (non-electrolytic) capacitor. As in the case of the inductor's equivalent circuit, we may write an expression for the impedance of the model as:

$$Z_C(j\omega) = R_s + j\omega L + \frac{R_p(1/j\omega C)}{R_p + (1/j\omega C)} \quad (5.5)$$

$Z_C(j\omega)$ may be put into the form, $Z_C = Re + jIm$, using complex algebra:

$$Z_C(j\omega) = \left[R_s + \frac{R_p}{1 + (\omega R_p C)^2} \right] + j \frac{(\omega L - \omega R_p^2 C + \omega^3 R_p^2 L C^2)}{1 + (\omega R_p C)^2} \quad (5.6)$$

This relation implies that the complex RLC circuit model for the capacitor at high frequencies can be reduced to a simple, equivalent model with a resistor given by the real term of equation 5.6 being in series, with a reactance given by the second term in equation 5.6. It is possible to find an expression for the equivalent series capacitance, $C_{s(EQ)}$. From the second term in equation 5.6:

$$X_c = \frac{-1}{\omega C_{s(EQ)}} = \frac{\omega(L - R_p^2 C) + \omega^3 R_p^2 C^2 L}{1 + (\omega R_p C)^2} \quad (5.7)$$

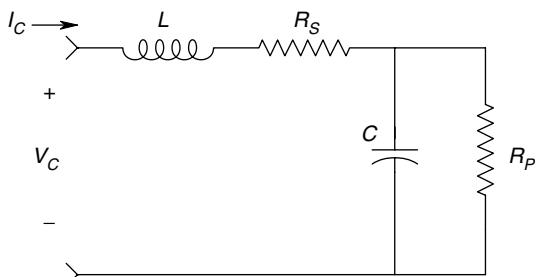


FIGURE 5.2

High frequency, lumped parameter model for a capacitor. Note that a capacitor exhibits inductance at high frequencies and an inductance exhibits capacitance between its windings.

from which it is easy to find:

$$C_{S(EQ)} = \frac{1 + (\omega R_p C)^2}{\omega^2 [(R_p^2 C - L) - \omega^2 R_p^2 C^2 L]} \quad (5.8)$$

At low frequencies, where L is negligible, the capacitive reactance reduces to $-1/\omega C$. Note that the series high frequency capacitor model exhibits resonance at a frequency where the reactance magnitude goes to zero. The resonant frequency of the capacitor is easily found to be

$$\omega_0 = \sqrt{1/LC - 1/(R_p C)^2} \text{ r/s} \quad (5.9)$$

Thus, the reactance of the equivalent circuit appears as $-1/\omega C$ at low frequencies. The reactance magnitude decreases with ω , until it reaches zero at ω_0 and then, at $\omega \gg \omega_0$, appears inductive, as $+\omega L$.

Most laboratory bridges do not measure capacitive reactance, rather they are calibrated in capacitance units and dissipation factor, D . D_s is defined as the ratio of the series equivalent resistance to the capacitive reactance. In algebraic terms,

$$D_s = \frac{R_{S(EQ)}}{X_{S(EQ)}} = \omega C_{S(EQ)} R_{S(EQ)} \quad (5.10)$$

Note that at any fixed frequency, a series R - C circuit, such as we have been discussing above, has an equivalent parallel R - C circuit which has the same impedance. We may also define a *dissipation factor* for the parallel equivalent circuit the ratio of the capacitive reactance to the equivalent parallel resistance. In algebraic terms,

$$D_P \equiv \frac{X_{P(EQ)}}{R_{P(EQ)}} = 1/(\omega C_{P(EQ)} R_{P(EQ)}) \quad (5.11)$$

By equating the impedances or conductances of the series and parallel R - C models, it is possible to derive relations relating one circuit to the other as given below:

$$R_{P(EQ)} = R_{S(EQ)} \left(\frac{D_s^2 + 1}{D_s^2} \right) \quad (5.12)$$

$$C_{P(EQ)} = C_{S(EQ)} / (1 + D_s^2) \quad (5.13)$$

$$R_{S(EQ)} = R_{P(EQ)} \left(\frac{D_P^2}{D_P^2 + 1} \right) \quad (5.14)$$

$$C_{S(EQ)} = C_{P(EQ)} (1 + D_P^2) \quad (5.15)$$

As you will see, certain capacitance bridge configurations make the assumption that the capacitance under measurement is represented by the series equivalent circuit

model, while others use the parallel equivalent circuit. At a given frequency, since the impedances of the two circuits are equal, the reason for using one model or the other lies in the practical derivation of the bridge balance equations in a form that allows two bridge elements to be uniquely calibrated in C_S and D_S , or in C_p and D_p .

5.4 AC Operation of Wheatstone Bridges

In this mode of operation, a low frequency ac excitation signal is used, usually ranging 100–1000 Hz. The bridge arms are resistors and the effects of stray capacitance to ground and inductance are generally negligible. Balance conditions are usually the same as for a dc Wheatstone bridge. However, one advantage of the ac excitation is that an electronic null detector with a lower uncertainty voltage, ΔV_o , can be used. This is probably not important, because bridge accuracy in measuring the unknown resistor, X , is largely due to the accuracy to which the arms M , N and P are known, as we demonstrated in the previous chapter. However, when the bridge is used in the voltage output mode, as with strain gauges, greater threshold sensitivity in measuring small changes in X are realized because the amplification of V_o can be done with a low noise, narrow band, ac amplifier working above the 1/f noise region.

Following amplification, the ac output voltage is generally converted into a dc signal proportional to ΔX through the use of a phase sensitive rectifier (PSR, discussed in detail in Section 2.7).

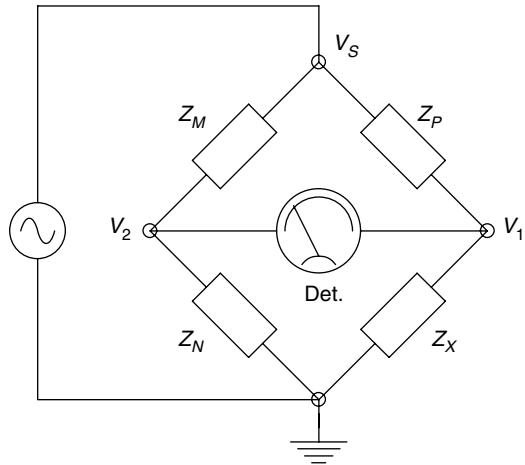
Another application for ac Wheatstone bridge operation is where the passage of dc through X will alter the magnitude of X . An example of this phenomenon is the measurement of the resistance of electrochemical electrodes (such as used in electrocardiography) attached to the body. The passage of direct current causes polarization of the electrodes due to ion migration in the dc electric field. Polarization increases electrode impedance. The use of ac generally avoids this phenomenon.

5.5 AC Bridges

The design, analysis and application of ac bridges is a large field, with rich, old literature. Most of the texts dealing in detail with this topic were written over 50 years ago. Great emphasis was placed on how an ac bridge approached its null. Vector (circle) diagrams were used to illustrate how the complex V_o behaves due to the interaction of the two variable bridge elements when reaching a null. In the following sections, we do not consider the details of how nulls are approached. Rather, we summarize the conditions at null and the applications of each type of bridge. Note that most modern ac bridges are designed to work at 1 kHz, although other frequencies may be used for special applications.

Figure 5.3 illustrates a general ac bridge, in which the arms are impedances, having real and imaginary parts. In general, the ac bridge output voltage can be written as a complex (vector) equation:

$$V_o = V_s \left[\frac{Z_X}{Z_X + Z_P} - \frac{Z_N}{Z_N + Z_M} \right] \quad (5.16)$$

**FIGURE 5.3**

General configuration for an ac bridge. Det. is the null detector; it is a sensitive ac voltmeter assumed to have infinite input resistance.

From equation 5.16, we find that at null, where $V_o = 0$, we can write the vector equation as:

$$Z_X Z_M = Z_N Z_P \quad (5.17)$$

which leads to the vector equation for Z_X in polar form, assuming bridge null:

$$Z_X = |Z_X| \angle \theta_x = \left| \frac{Z_N Z_P}{Z_M} \right| \angle (\theta_N + \theta_P - \theta_M) \quad (5.18)$$

Thus, finding unique expressions for the unknown L and Q , or C and D , requires solving vector equations of the form above. Generally, this is done by independently equating the real and imaginary terms on both sides of equation 5.17.

It can be shown for all conventional, four-arm ac bridges that the same conditions for null exist if the null detector is exchanged for the ac source and equation 5.17 still applies. Presented below is a summary of bridge designs used to measure capacitance, inductance and mutual inductance using the null method.

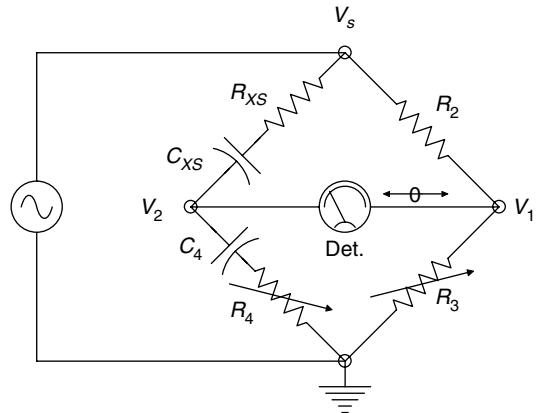
5.5.1 Bridges Used to Measure Capacitance

In this section, we examine the designs of bridges useful for the measurement of capacitance. Capacitance measuring bridges can be subdivided into those designs suitable for low loss (low D) capacitors and those giving best results for lossy (high D) capacitors.

5.5.1.1 The Resistance Ratio Bridge

The resistance ratio bridge is shown in Figure 5.4. It is best used to measure capacitors with low D and it uses the series equivalent capacitor model. It is used in the venerable, General Radio Model 1650A bridge. At null, $V_o = 0$ and we may write:

$$R_3(R_{XS} + 1/j\omega C_{XS}) = R_2(R_4 + 1/j\omega C_4) \quad (5.19)$$

**FIGURE 5.4**

A resistance ratio bridge, used to measure capacitors with low D_S .

Equating the real terms, we find

$$R_{XS} = R_2 R_4 / R_3 \quad (5.20)$$

Equating the imaginary terms, we obtain

$$C_{XS} = R_3 C_4 / R_2 \quad (5.21)$$

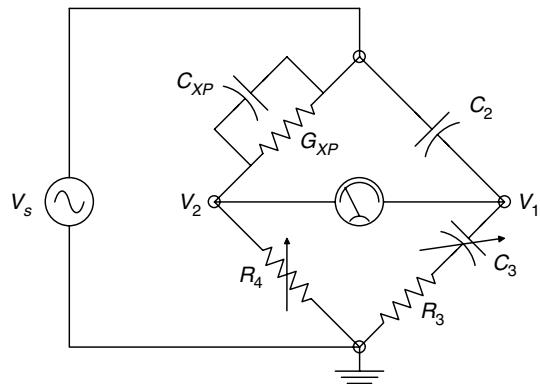
The dissipation factor for the series capacitor model was shown to be $D_S = \omega C_{XS} R_{XS}$. If we substitute the relations for C_{XS} and R_{XS} into D_S , then:

$$D_S = \omega C_4 R_4 \quad (5.22)$$

Thus, we see that R_4 can be uniquely calibrated in low D_S and R_3 should be calibrated in C_{XS} units.

5.5.1.2 The Schering Bridge

The Schering bridge is useful for measuring capacitors with high losses (high D_s). In finding the balance conditions for this bridge, it is expedient to use the parallel R-C equivalent circuit, as shown in Figure 5.5. At null, we can write, as before, a relation

**FIGURE 5.5**

A Schering bridge, used to measure capacitors with high D_p .

similar to equation 5.17:

$$\mathbf{Z}_x \mathbf{Z}_3 = \mathbf{Z}_2 \mathbf{Z}_4 \quad (5.23)$$

or

$$\mathbf{Z}_3 \mathbf{Y}_2 = \mathbf{Y}_x \mathbf{Z}_4 \quad (5.24)$$

Thus

$$(R_3 + 1/j\omega C_3)j\omega C_2 = (G_{XP} + j\omega C_{XP})R_4 \quad (5.25)$$

Now by equating real terms and imaginary terms, the conditions at balance are obtained as:

$$C_{XP} = R_3 C_2 / R_4 \quad (5.26)$$

$$R_{XP} = R_4 C_3 / C_2 \quad (5.27)$$

$$D_P = 1/(\omega C_{XP} R_{XP}) = \omega C_3 R_3 \quad (5.28)$$

Notice that to obtain uniqueness in finding C_{XP} and D_P , C_3 can be calibrated in D_P and R_4 can be calibrated in C_{XP} . The other components are fixed for a given range of C_{XP} .

5.5.1.3 The Parallel C Bridge

This bridge design, shown in Figure 5.6, is also used in the GR 1650A bridge to measure high D (lossy) capacitors ($0.1 < D_P < 50$). At null, we can write $R_2 \mathbf{Z}_4 = R_3 \mathbf{Z}_x$. This leads to:

$$R_2(G_{XP} + j\omega C_{XP}) = R_3(G_4 + j\omega C_4) \quad (5.29)$$

from which the following relation can be easily find:

$$C_{XP} = C_4 R_3 / R_2 \quad (5.30)$$

$$R_{XP} = R_2 R_4 / R_3 \quad (5.31)$$

$$D_P = 1/(\omega C_{XP} R_{XP}) = 1/(\omega C_4 R_4) \quad (5.32)$$

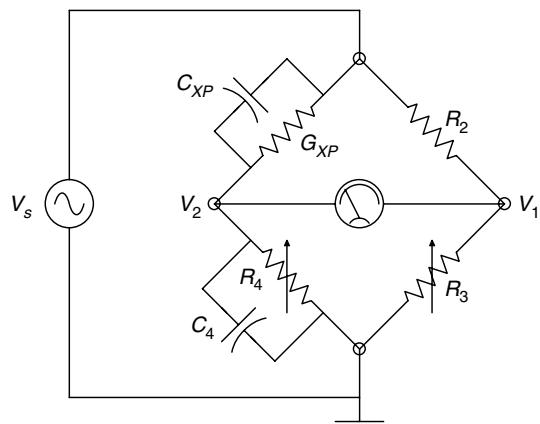


FIGURE 5.6

A parallel C bridge used to measure capacitors with $0.1 < D_P < 50$.

From the above equations, it is easy to see that R_4 should be calibrated in D_P and R_3 or R_2 can give C_{XP} .

5.5.1.4 The De Sauty Bridge

The De Sauty bridge, shown in Figure 5.7, is a deceptively simple bridge which is often used to produce an output voltage that is proportional to a small change, δC , in one of the capacitors, rather than to measure capacitance. The δC may be caused by a variety of physical phenomena, such as the deflection of a diaphragm due to a pressure difference across it, a change in capacitor plate separation due to a change in material thickness or an applied force, and so on.

Since the output voltage is a nonlinear function of large capacitance changes, the De Sauty bridge is usually operated such that $\delta C/C_0 \ll 1$ and the is linearly proportional to the QUM. Also, the resistors R_1 and R_2 are made equal to R and null is achieved by setting $C_4 = C_3 = C$. Now, the output of the bridge can be expressed as a function of R , C , ω and δC as:

$$V_o = V_b = \frac{j\omega RC(\delta C/C)}{(1 + j\omega RC)(1 + j\omega R(C + \delta C))} \quad (5.33)$$

From equation 5.33, we see that three approximate relations for V_o can be written, depending on the operating frequency. First, let $\omega \gg 1/RC$. The transfer function reduces to:

$$V_o/V_b = (\delta C/C)(1/j\omega RC)[1 - (\delta C/C)] \quad (5.34)$$

Generally, the second order term can be neglected. However, a square law nonlinear distortion will occur for $\delta C/C \rightarrow 0.1$.

Second, we let the bridge be excited at a low frequency so that $\omega \ll 1/RC$. Under this condition, we observe that:

$$V_o/V_b = (\delta C/C)(j\omega RC) \quad (5.35)$$

Third, if $\omega = 1/RC$ (tuned bridge condition), the output can be written as:

$$V_o/V_b = \delta C/2C \text{ (Real)} \quad (5.36)$$

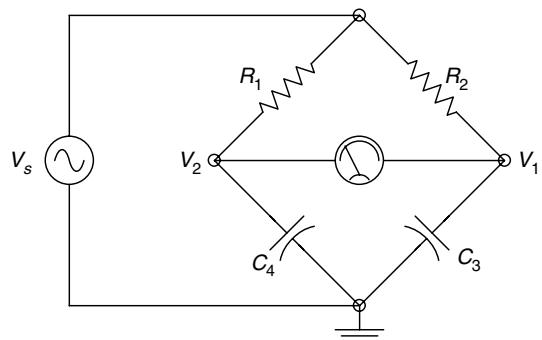


FIGURE 5.7

A De Sauty bridge, used to measure small changes in capacitance. $C_4 = C$, $C_3 = C + \Delta C$, $R_1 = R_2 = R$, and $\omega = 1/RC$.

Maximum sensitivity generally occurs for the third case, when $\omega = 1/RC$. In this case, the δC , which is just detectable, can be estimated by using the series expansion as follows:

$$\Delta V_o = \left(\frac{\partial V_o}{\partial C} \right) \delta C = V_b \delta C / (2C) \quad (5.37)$$

For example, the just detectable δC is found by assuming the bridge detector ac voltmeter resolution to be $\Delta V_o = 0.1 \mu V$ and the bridge excitation, V_b , to be 5 V. If we let $C = 100 \text{ pF}$, then δC_{MIN} is given by:

$$\delta C_{MIN} = (0.1 \times 10^{-6} \times 2 \times 100 \times 10^{-12}) / 5 = 4 \times 10^{-6} \text{ pF} = 4 \text{ aF} \quad (5.38)$$

This is an incredible theoretical sensitivity, which is not reachable in practice because of stray capacitances associated with the bridge arms, detector noise, etc.

5.5.1.5 The Wien Bridge

The Wien bridge, illustrated in Figure 5.8, is generally not used to measure capacitors because of the complexity of its solution at null. The Wien bridge is a frequency dependent null network and as such, finds application in the design of tuned band-pass and band-reject (notch) filters, as also oscillators. Using Figure 5.8, the general form of the bridge balance equation at null is expressed as:

$$Z_1 Z_3 = Z_2 Z_X \quad (5.39)$$

This can be modified as:

$$Z_1 Y_X = Z_2 Y_3 \text{ or } (R_1 / 1/j\omega C_1)(G_X + j\omega C_X) = R_2 G_3 \quad (5.40)$$

To find the null conditions, we must solve the simultaneous equations derived from equation 5.40 must be solved by separately equating the real and imaginary terms:

$$G_X R_1 + C_X / C_1 = R_2 G_3 \quad (5.41)$$

$$-G_X / \omega C_1 + \omega C_X R_1 = 0 \quad (5.42)$$

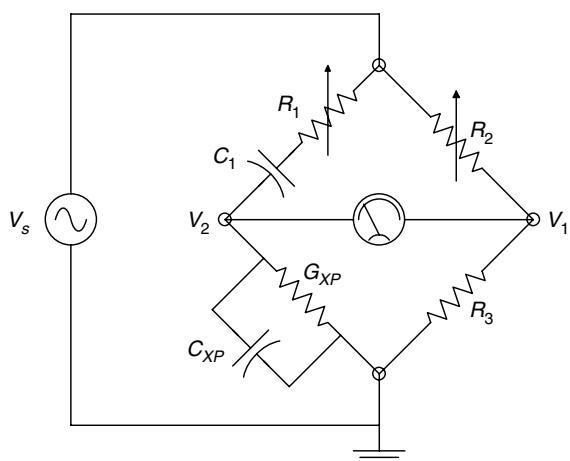


FIGURE 5.8

The Wien bridge is frequency dependent. Here, it is used to measure capacitance with $D_p \ll 0.1$. R_2 can be calibrated in C_X ; R_1 in D_{PX} .

Their solution leads to:

$$C_X = \frac{(R_2/R_3)C_1}{(\omega^2 C_1^2 R_1^2 + 1)} = \frac{(R_2/R_3)C_1}{(D_{1S}^2 + 1)} \rightarrow (R_2/R_3)C_1 \quad (5.43)$$

$$G_X = \frac{(\omega R_1 R_2 / R_3)(\omega C_1^2)}{(\omega^2 C_1^2 R_1^2 + 1)} \quad (5.44)$$

$$D_{XP} = 1/(\omega C_X R_X) = \omega R_1 C_1 = D_{1S} \quad (5.45)$$

Provided $D_{XP} = D_{1S} \ll 0.1$, R_2 can be uniquely calibrated in C_X units and R_1 in D_{XP} units. The use of the Wien bridge as a frequency sensitive null network can be demonstrated by writing its transfer function:

$$\mathbf{V}_o/\mathbf{V}_b = \frac{1/(G_X + j\omega C_X)}{1/(G_X + j\omega C_X) + R_1 + 1/j\omega C_1} - \frac{R_3}{R_3 + R_2} \quad (5.46)$$

which reduces to:

$$\mathbf{V}_o/\mathbf{V}_b = \frac{j\omega C_1 R_X}{1 + j\omega(C_1 R_X + C_1 R_1 + C_X R_X) + (j\omega)^2 C_1 R_1 C_X R_X} - \frac{R_3}{R_3 + R_2} \quad (5.47)$$

Now, if we tune the source frequency to

$$\omega = \omega_o = (1/C_1 R_1 C_X R_X) \text{ r/s} \quad (5.48)$$

then the transfer function becomes:

$$\mathbf{V}_o/\mathbf{V}_b = \frac{C_1 R_X}{C_1 R_X + C_1 R_1 + C_X R_X} - \frac{R_3}{R_3 + R_2} \quad (5.49)$$

This real transfer function goes to zero at $\omega = \omega_o$ if:

$$R_1 = R_X = R \quad (5.50)$$

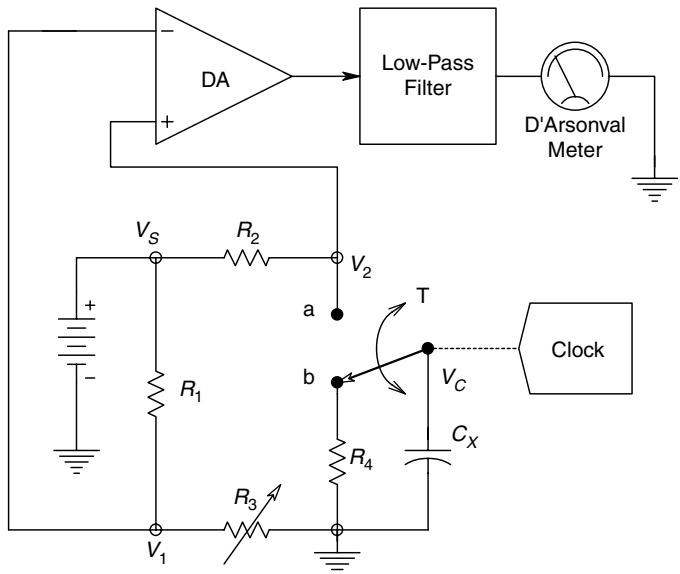
$$C_1 = C_X = C \quad (5.51)$$

$$R_2 = 2R_3 = R \quad (5.52)$$

then we have:

$$\mathbf{V}_o/\mathbf{V}_b = 1/3 - 1/3 = 0 \quad (5.53)$$

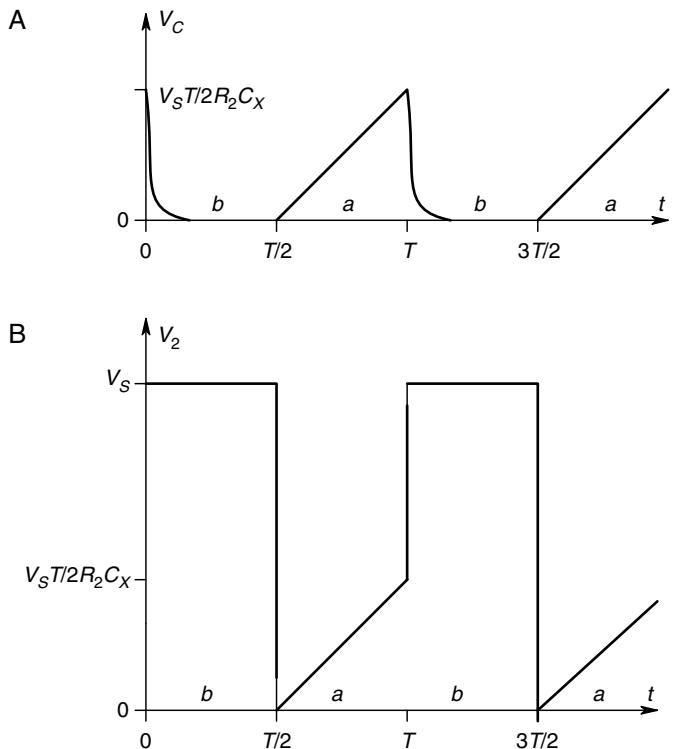
As mentioned above, the frequency dependent null of the Wien bridge can be exploited to create tuned filters and oscillators that are useful in instrumentation systems.

**FIGURE 5.9**

A commutated capacitor bridge. Unlike other ac bridges, the detector is a dc null meter which responds to the average voltage.

5.5.1.6 The Commutated Capacitor Bridge

A commutated capacitor bridge is shown in Figure 5.9. Unlike a conventional ac bridge, this bridge uses a dc V_s . Such bridges do not allow measurement of the capacitor D . However, they are often used for high accuracy, low frequency applications. A MOS switch is driven by a square wave clock with period T . For $T/2$ seconds, the capacitor is connected to node a and charges toward V_s through R_2 . For an alternate $T/2$ seconds,

**FIGURE 5.10**

A. Voltage waveform across C_X , $v_C(t)$. Note $RC_X \ll R_2C_X$. B. Voltage waveform, $v_2(t)$.

C_X is allowed to discharge through R_4 to ground. The waveforms at nodes a and V_c are shown in Figures 5.10A and 5.10B, respectively. We assume that the clock period and resistors R_2 and R_4 are chosen so that $R_4C_X \ll T/2$ and $R_2C_X \gg T/2$. We also assume that the null detector responds to the average of $(V_2 - V_1)$. Hence at null, we find that:

$$V_sR_3/(R_1 + R_3) = (V_s/2)[1 + (T/4R_2C_X)] \quad (5.54)$$

From which we can solve for C_X :

$$C_X = \frac{T(R_3 + R_1)}{4R_2(R_3 - R_1)} \quad (5.55)$$

where obviously, $R_3 > R_1$. R_3 or R_2 can be varied to obtain null and for a given range setting (fixed values of T , R_2 and R_1), can be calibrated according to equation 5.55 above.

5.5.2 Bridges Used to Measure Inductance and Mutual Inductance

As in the case of bridges used for measuring capacitance, the inductance bridges can be subdivided into bridges that are optimal in terms of reaching null, for measuring high Q inductors and those best suited for the measurement of low Q inductors. In addition, we describe below the design of two specialized bridges used to measure the mutual inductance of power and audio frequency transformers.

5.5.2.1 The Maxwell Bridge

As the name implies, this is a well-known bridge which is used to measure low Q inductors having Q s in the range of 0.02–10. The Maxwell bridge is used in the General Radio model 1650A multi-purpose bridge; its circuit is shown in Figure 5.11. At null:

$$Z_X = Z_2 Y_3 Z_4 \quad (5.56)$$

or

$$(R_X + j\omega L_X) = R_2(G_3 + j\omega C_3)R_4 \quad (5.57)$$

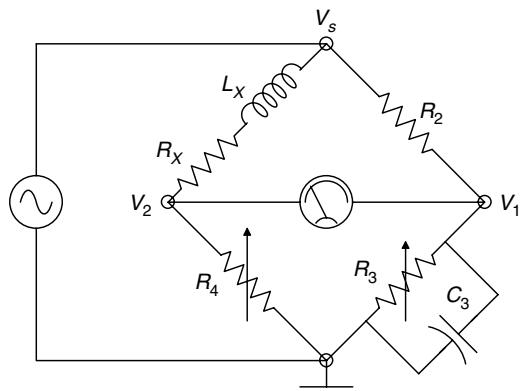


FIGURE 5.11

The Maxwell bridge for low Q inductors ($0.02 < Q < 10$).

From equation 5.54, we easily find:

$$L_X = C_3 R_2 R_4 \quad (5.58)$$

$$R_X = R_2 R_4 / R_3 \quad (5.59)$$

$$Q_X = \omega L_X / R_X = \omega C_3 R_3 \quad (5.60)$$

Thus R_4 is calibrated in inductance units and the R_3 scale reads Q . In the GR 1650A bridge, R_4 can be varied in the range 0–11.7 kΩ and R_3 can be set at 0–16 kΩ.

5.5.2.2 Parallel Inductance Bridge

This bridge, shown in Figure 5.12, is also used in the GR 1650A bridge to measure high Q inductors ($1 < Q_P < \infty$). It is somewhat unusual in that it uses a parallel equivalent circuit for the inductor. At a given frequency, any series $R-L$ circuit can be made to be equal in impedance to a parallel $R-L$ circuit. We describe these equivalences below. At balance, we can write:

$$R_2 R_4 = (R_3 + 1/j\omega C_3) \frac{j\omega L_X R_X}{R_X + j\omega L_X} \quad (5.61)$$

From which we obtain:

$$j\omega L_X R_2 R_4 + R_X R_2 R_4 = j\omega L_X R_X R_3 + L_X R_X / C_3 \quad (5.62)$$

Equating the real and imaginary terms in equation 5.62 independently, we finally obtain the expressions for L_{XP} , R_{XP} and Q_P as follows:

$$L_{XP} = C_3 R_2 R_4 \quad (5.63)$$

$$R_{XP} = R_2 R_4 / R_3 \quad (5.64)$$

$$Q_P = R_X / \omega L_X = 1 / \omega C_3 R_3 \quad (5.65)$$

Thus, it is expedient to make R_4 the L dial and R_3 the high Q dial on this bridge.

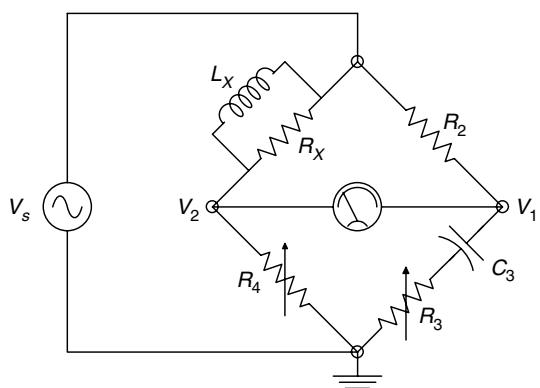
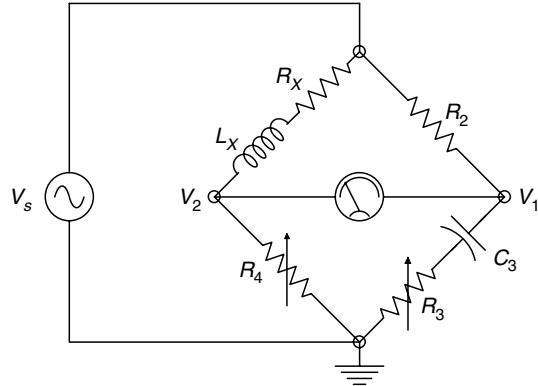


FIGURE 5.12

The General Radio parallel inductance model bridge for high Q inductors ($1 < Q < \infty$). R_3 is calibrated in Q_P , R_4 in L_{XP} .

**FIGURE 5.13**

The Hay bridge is used to measure high Q inductors. R_3 is calibrated in Q_S , and R_4 in inductance units.

5.5.2.3 The Hay Bridge

The Hay bridge, shown in Figure 5.13, uses the series $R-L$ model for an inductor to measure the inductance and Q of high Q coils. Assuming null, we find:

$$(j\omega L_X + R_X)(R_3 + 1/j\omega C_3) = R_2 R_4 \quad (5.66)$$

The balance equation, equation 5.66, can be broken down to two simultaneous equations:

$$L_X/C_3 + R_X R_3 = R_2 R_4 \quad (5.67)$$

$$\omega L_X R_3 - R_X/\omega C_3 = 0 \quad (5.68)$$

Their solution yields:

$$L_X = \frac{R_4 R_2 C_3}{(\omega^2 R_3^2 C_3^2 + 1)} = \frac{R_4 R_2 C_3}{(Q^{-2} + 1)} \quad (5.69)$$

$$R_X = \frac{\omega R_3 R_2 R_4 (\omega C_3^2)}{(\omega^2 R_3^2 C_3^2 + 1)} \quad (5.70)$$

$$Q_S = \omega L_X / R_X = 1 / \omega C_3 R_3 \quad (5.71)$$

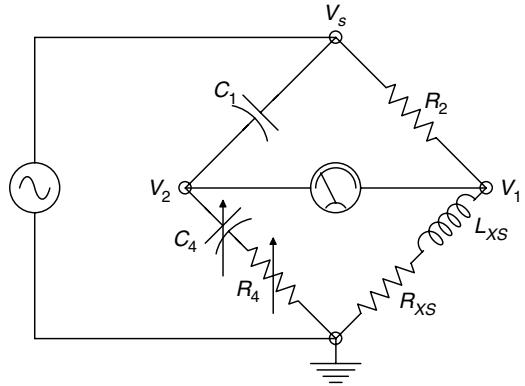
Here, R_3 may be calibrated in Q_S and R_4 in inductance units. Inductance calibration is substantially independent of $Q_S(R_3)$ as long as $Q_S > 10$.

5.5.2.4 The Owen Bridge

The Owen bridge, shown in Figure 5.14, uses the conventional series inductance model and is best used on large, low Q inductors. This bridge is somewhat unique in that the null may be obtained by varying both elements in the 4-arm, R_4 and C_4 . At null we have:

$$(j\omega L_X + R_X)/j\omega C_1 = R_2(R_4 + 1/j\omega C_4) \quad (5.72)$$

By equating the real and imaginary terms in equation 5.72 separately, it is easy to find that

**FIGURE 5.14**

The Owen bridge is best used to measure large, low Q inductors. R_4 can be calibrated in inductance and C_4 in series equivalent resistance, R_{xs} .

$$R_X = R_2 C_1 / C_4 \quad (5.73)$$

$$L_X = R_2 C_1 R_4 \quad (5.74)$$

Hence, R_4 may be uniquely calibrated in inductance and C_4 in resistance (nonlinear scale). The major disadvantages of the Owen bridge are that it requires a precisely calibrated variable capacitor (C_4) and that for high Q , low inductance coils, impractically large C_4 values may be required.

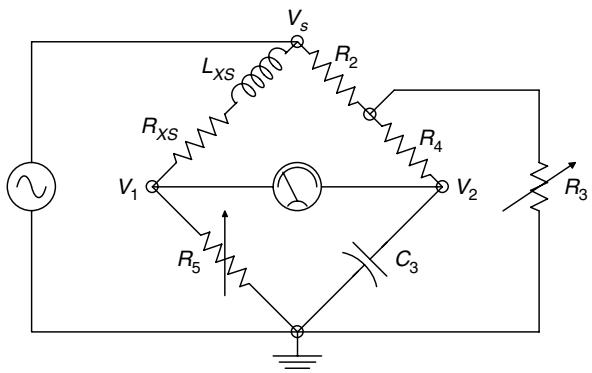
5.5.2.5 The Anderson Bridge

The Anderson bridge is illustrated in Figure 5.15. According to Stout (1950), the Anderson bridge gives the best convergence to nulls for low Q coils. To analyze this bridge, we note that at null, $\mathbf{V}_o = \mathbf{V}_1 - \mathbf{V}_2 = 0$ and the voltage divider relations can be used to find \mathbf{V}_1 and \mathbf{V}_2 . Thus we have:

$$\frac{\mathbf{V}_o}{\mathbf{V}_s} = \frac{R_5}{R_5 + R_X + j\omega L_X} - \frac{R_3}{R_3 + R_2 + j\omega C[R_3 R_4 + R_3 R_2 + R_4 R_2]} \quad (5.75)$$

Setting $\mathbf{V}_o = 0$, we find:

$$R_X = R_2 R_5 / R_3 \quad (5.76)$$

**FIGURE 5.15**

The Anderson bridge works best with low Q coils. It has relatively complex balance equations.

$$L_X = CR_5[R_4 + R_2 + (R_4R_2/R_3)] \quad (5.77)$$

$$Q_X = \omega CR_3[1 + R_4/R_2 + R_4/R_3] \quad (5.78)$$

It is easily seen that the price we pay in using this bridge, which is easily balanced with low Q inductors, is a relatively complex set of balance equations. R_5 can be calibrated for L_X and R_3 used to read Q_X .

5.5.2.6 The Heaviside Mutual Inductance Bridge

Before discussing the circuit and balance conditions for the Heaviside mutual inductance bridge, we should review the meaning of *mutual inductance* of a transformer or a pair of magnetically coupled coils. Figure 5.16 illustrates a simple equivalent circuit for a transformer operating at low frequencies. Under ac steady state conditions, we may write the loop equations:

$$V_1 = I_1(R_1 + j\omega L_1) + I_2(j\omega M) \quad (5.79)$$

$$V_2 = I_1(j\omega M) + I_2(R_2 + j\omega L_2) \quad (5.80)$$

where M is the mutual inductance in H, given a positive sign for the dot convention shown in Figure 5.16.

Figure 5.17 illustrates the circuit of a Heaviside bridge. At null, $V_o = V_1 - V_2 = 0$, hence we can write:

$$I_1R_1 = I_2R_2 \quad (5.81)$$

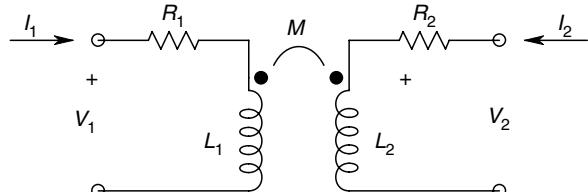


FIGURE 5.16

Circuit model for a transformer with mutual inductance at low frequencies. L_1 and L_2 are the self-inductances of the primary and secondary windings, respectively.

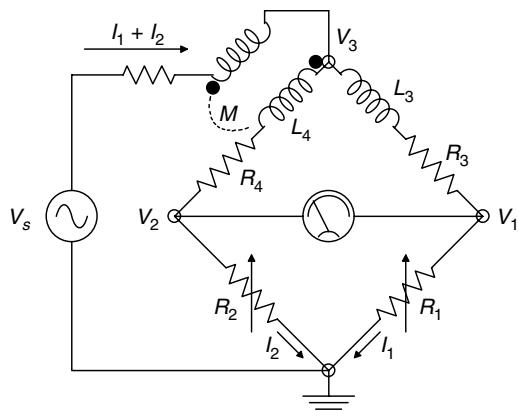


FIGURE 5.17

The Heaviside mutual inductance bridge.

and also,

$$V_3 = I_1(R_1 + R_3 + j\omega L_3) = I_2(R_2 + R_4 + j\omega L_4) + j\omega M(I_1 + I_2) \quad (5.82)$$

If we solve for I_1 in equation 5.81, and substitute it into equation 5.82, and then equate the real and then imaginary terms, we find that at null:

$$R_4 = R_2 R_3 / R_1 \quad (5.83)$$

$$M = \frac{R_2 L_3 - R_1 L_4}{R_1 + R_2} \quad (5.84)$$

Note that M may have either sign, depending on the dots. The transformer's primary self-inductance, L_4 , must be measured with the secondary open circuited. There is no need to know the secondary self-inductance, or the primary resistance when using the Heaviside mutual inductance bridge.

5.5.2.7 The Heydweiller Mutual Inductance Bridge

The Heydweiller means of measuring transformer mutual inductance uses a 'bridge' in which one arm is a short circuit (Figure 5.18). Thus the conventional method of examining $Z_1 Z_3 = Z_2 Z_4$ will not work with this bridge! Instead, it is seen that at null, $V_2 = V_1 \rightarrow 0$, thus by use of KVL:

$$V_2 = 0 = I_1(R_3 + j\omega L_3) - j\omega M(I_1 + I_2) \quad (5.85)$$

and also,

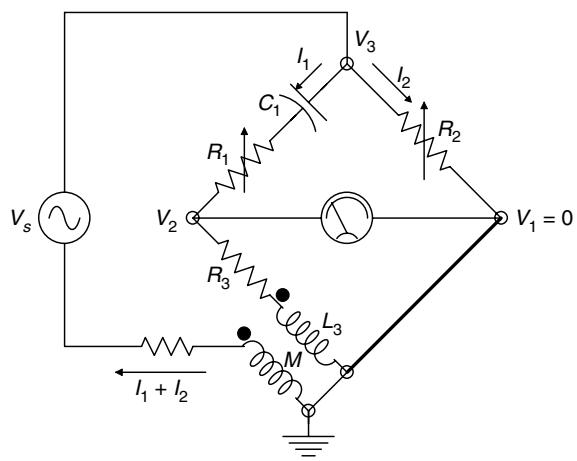
$$I_2 = I_1(R_1 + 1/j\omega C_1)/R_2 \quad (5.86)$$

If equation 5.86 is substituted into equation 5.85 and the sum of the real terms is set equal to zero, we find:

$$M = C_1 R_2 R_3 \quad (5.87)$$

FIGURE 5.18

The Heydweiller mutual inductance null circuit.



Equating the sum of the imaginary terms to zero leads to:

$$L_3 = C_1 R_3 (R_1 + R_2) \quad (5.88)$$

Note that R_3 must be known, as it appears in both the expression for M and the primary self-inductance, L_3 . R_1 and R_2 are manipulated to obtain the null.

5.5.3 Null Method of Measuring Transistor Small Signal Transconductance and Feedback Capacitance

The null circuits discussed below are used to measure BJT or FET small signal transconductance at a given dc quiescent operating point. An audio frequency ac signal, usually 1 kHz, is used. Measurement of small signal transconductance is important to verify if transistors are good and to match pairs for like characteristics when building a discrete differential amplifier.

First, we illustrate the small signal, modified hybrid pi model for a BJT operating in its linear region (i.e., neither cut-off or saturated). This model, shown in Figure 5.19, is generally accepted to be valid for frequencies up to the BJT's $f_T/3$. The small signal transconductance, g_m , is defined as:

$$g_m \equiv \left(\frac{\partial i_c}{\partial v_{b'e}} \right), v_{ce} = 0 \quad (5.89)$$

The BJT's g_m may be approximated by:

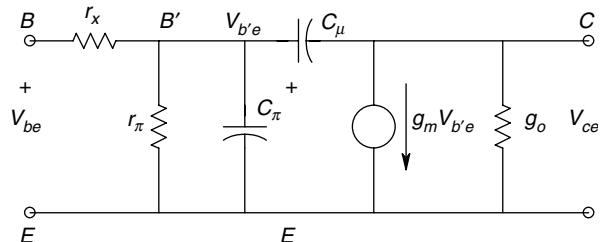
$$g_m \approx h_{fe}/r_\pi = I_{CQ}/V_T \quad (5.90)$$

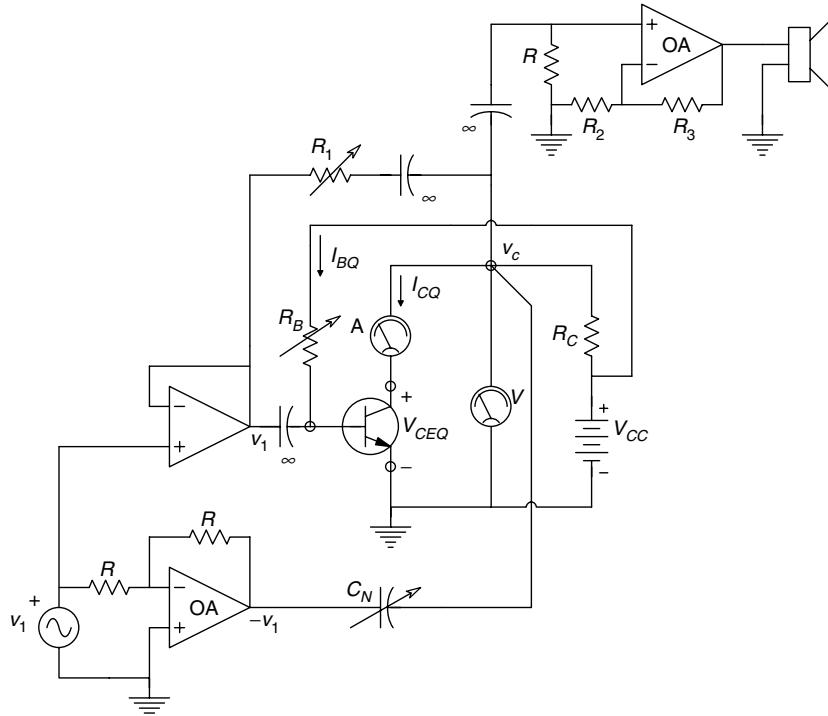
The capacitance C_μ is the dc voltage variable capacitance of the reverse biased, collector base junction evaluated at the dc operating point of the BJT. C_π is the capacitance of the forward biased base emitter junction, measured at the operating point, which is also a function of the dc voltage across the junction. r_π is the small signal base input resistance. It is approximately equal to V_T/I_{BQ} . The base input spreading resistance, r_x , is generally less than 100 Ω and we set it equal to zero to simplify analysis. The output conductance, g_o , is generally very small, but its exact value is not important in determining g_m .

Figure 5.20 illustrates the complete circuit of the transconductance bridge. Resistors R_B and R_C and dc source V_{CC} are used to set up the dc quiescent operating (Q) point of the BJT. A is a milliammeter to read I_{CQ} and V is a dc voltmeter to read V_{CEQ} . A small, audio frequency, sinusoidal signal, v_1 , is applied to the transistor's base through a large dc blocking capacitor, C_1 . Simultaneously, the inverted ac signal, $-v_1$, is applied to the BJT's collector. Capacitor C_N and resistor R_1 are adjusted to get a null or minimum ac

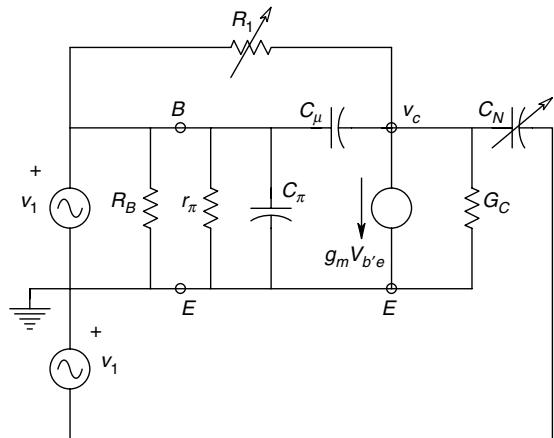
FIGURE 5.19

Simplified, hybrid pi, high frequency, small signal circuit model for BJTs.



**FIGURE 5.20**

Circuit of null system used to measure the small signal transconductance of BJTs. OAs are op-amps. Capacitors marked ∞ are large and have negligible reactance at the operating frequency. A is a dc milliammeter used to measure I_{CQ} . V is a dc voltmeter used to measure V_{CEQ} . The null is sensed acoustically with a loudspeaker or headphones.

**FIGURE 5.21**

Small signal, equivalent circuit of the g_m measuring circuit of Figure 5.20.

signal at the v_c node. A node equation for v_c can be written for the small signal equivalent circuit of Figure 5.21. Note that $v_{be} \cong v_1$. At null:

$$0 = V_c[G_1 + G_c + j\omega(C_\mu + C_N)] - V_1(j\omega C + G_1) - j\omega C_N(-V_1) + g_m V_1 \quad (5.91)$$

for $V_c \rightarrow 0$, it is evident that:

$$g_m = G_1 \quad (5.92)$$

and

$$C_\mu = C_N \quad (5.93)$$

A similar g_m null circuit can be used to measure the small signal transconductance of various types of FETs (see the figure in the problems at the end of the chapter). Note that FET g_m at the dc operating point can be shown to be given by:

$$g_m = \left(\frac{\partial i_d}{\partial v_{gs} v_{ds} \rightarrow 0} \right) = g_{m0} \sqrt{I_{DQ}/I_{DSS}} = g_{m0}(1 - V_{GS}/V_P) \quad (5.94)$$

where g_{m0} is the small signal transconductance evaluated for $V_{GS}=0$, V_P is the pinch-off voltage (for JFETs) and I_{DSS} is the JFET's dc drain current measured for $V_{GS}=0$ and $V_{DS} \gg V_{GS} + V_P$.

5.6 Chapter Summary

In this chapter, we reviewed the more important audio frequency ac bridges and null systems used for making precision measurements of capacitance, capacitor dissipation factor (D), inductance, inductor (Q), mutual inductance and transistor small signal transconductance. General conditions for bridge null were derived, and the conditions at null for each circuit presented. No attempt has been made to discuss the effects of stray capacitances between bridge elements and the capacitance to ground offered by the two detector terminals. These are generally not significant sources of measurement error for most ac bridges operating at 1 kHz. A common means of compensating for detector capacitance to ground is by the Wagner Earth circuit (Stout (1950), Sec. 9-11).

Measurement of circuit parameters at high frequencies (including video and radio frequencies) requires special apparatus such as the Q-Meter, or instruments such as the Hewlett-Packard model 4191A RF Impedance Analyzer.

Problems

- 5.1 The bridged T circuit provides an alternate means to four-arm bridges for the measurement of circuit element parameters. Figure P5.1 illustrates a bridged T circuit used to measure the parallel model for an inductor. A null at the V_o node is obtained by varying the two capacitors together with ω of the sinusoidal source and R . Assume $V_o=0$. Find expressions for ω , L_p , R_p and Q_p in terms of R and C . The expressions for

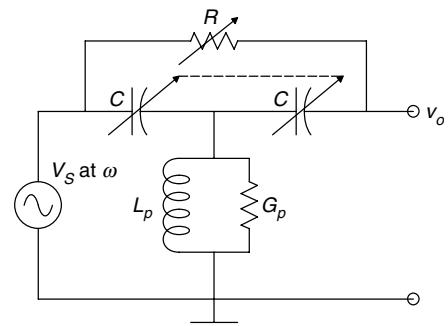


FIGURE P5.1

L_p and R_p should not contain terms in R_p and L_p , respectively. In general, $Q_p = R_p / (\omega L_p)$.

- 5.2 Repeat Problem 5.1 for the bridged T circuit of Figure P5.2, and find expressions for ω , L_s , R_s and Q_s . $Q_s = \omega L_s / R_s$.

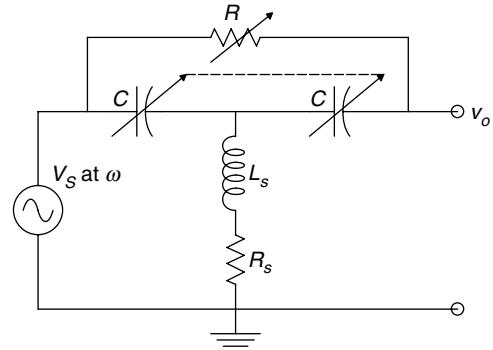


FIGURE P5.2

- 5.3 In the bridged T circuit of Figure P5.3, find ω_o for null. Also find a numerical value for R_p .

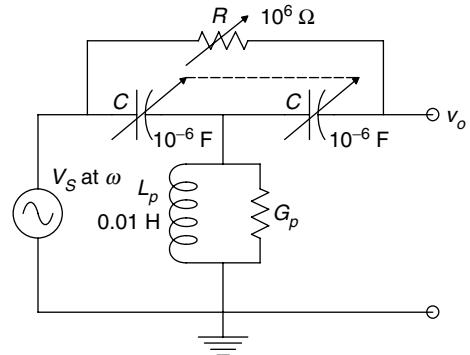


FIGURE P5.3

- 5.4 Figure P5.4 illustrates a parallel T null circuit. Find the frequency ω_o at which the $V_o \rightarrow 0$.

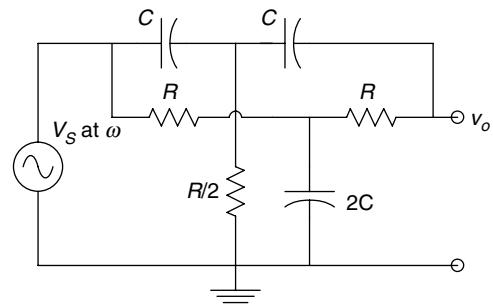


FIGURE P5.4

- 5.5 Prove that the vector conditions for null (bridge balance) are unchanged if you exchange detector and source nodes, as shown in Figure P5.5.

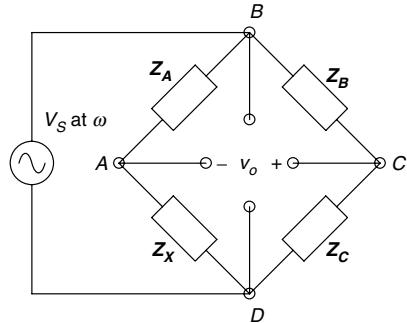


FIGURE P5.5

- 5.6 At a given frequency, ω , a series $R-L$ circuit can be exactly replaced by a parallel $R-L$ circuit. Find expressions for L_p and R_p in terms of L_s , R_s and $Q_s = \omega L_s / R_s$.
- 5.7 Figure P5.7 illustrates a De Sauty bridge in which the capacitors change differentially. The source is $v_s(t) = 5 \sin(\omega t)$. Assume $\omega = 1/RC_o$ r/s. The detector can only resolve $\delta V_o = 5 \mu\text{V}$ pk. Find the smallest resolvable $\Delta C/C_o$.

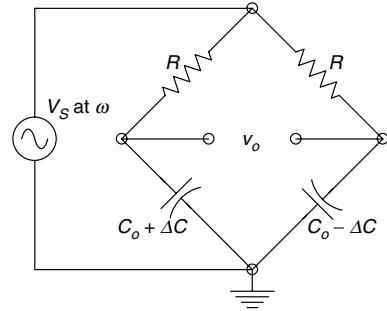


FIGURE P5.7

- 5.8 A four-arm ac bridge, operated at 1 kHz, is at null. $Z_{AB} = 1,500 \Omega$, $Z_{BC} = 1,000 \Omega$, $Z_{CD} = Z_X$, $Z_{DA} = 10 + 1/j\omega 10^{-7} \Omega$. Z_X is a series resistance and R_x is a reactive element. Find the numerical value of R_x and the value and kind of reactance (L or C).
- 5.9 A four-arm ac bridge is operated at $\omega = 2\pi f = 10^4$ r/s. Node D is grounded. Arm AB is a 10^{-10} F capacitor. Arm BC contains the unknown. Arm CD is a pure $10^4 \Omega$ resistor and arm DA is a $12.5 \text{ k}\Omega$ resistor in parallel with a variable capacitor, C_{DA} . Null is obtained with $C_{DA} = 120 \text{ pF}$. Find the two series elements of Z_{BC} .
- 5.10 A Wien bridge is shown in Figure P5.10. Find f_o of the source and the R_x that will give null.

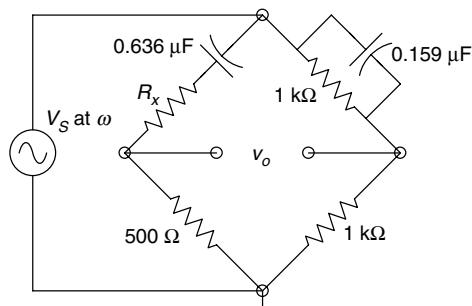


FIGURE P5.10

5.11 A De Sauty bridge is shown in Figure P5.11.

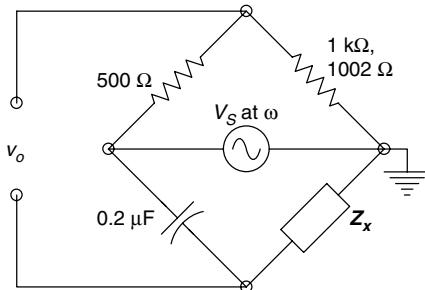


FIGURE P5.11

- (A) Find Z_x .
 - (B) Now Z_{BC} is made $1,002 \Omega$, unbalancing the bridge. Find an expression for $v_o(t)$.
- 5.12 A four-arm ac bridge is shown in Figure P5.12. The oscillator frequency, $\omega = 50,000 \text{ r/s}$. Find Z_x at null. Assume a series, $R_x + jX_x$ model.

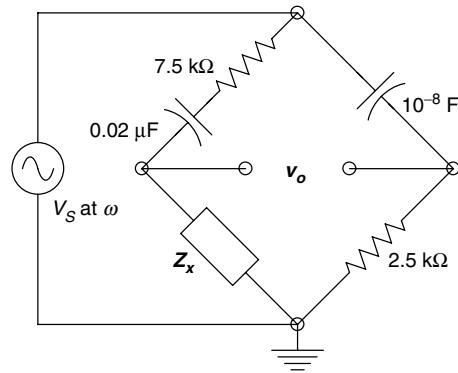


FIGURE P5.12

- 5.13 A four-arm ac bridge is shown in Figure P5.13. $f = 1000 \text{ Hz}$. Assuming null, find the constants of arm CD.

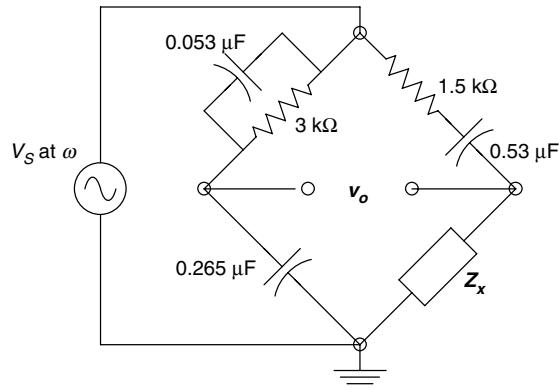


FIGURE P5.13

6

Survey of Sensor Input Mechanisms

6.1 Introduction

An input *sensor* or *transducer* is a device which permits the conversion of energy from one form to another. It is the first element in an instrumentation or measurement system. For example, it might convert temperature to voltage. A broader definition might substitute 'information' for 'energy'. Its linearity, range, noise and dynamic response largely determine the resolution, sensitivity and bandwidth of the overall system. A *transducer* is a device that, in some way, obeys reciprocity (e.g. the physical input quantity is converted to an output voltage and if a voltage is applied to the output terminals, the input quantity is generated at the input interface of the transducer). Examples of transducers include piezoelectric crystals used to sense force or pressure, and electrodynamic devices such as loudspeakers and D'Arsonval meter movements. Obviously, not all sensors are transducers.

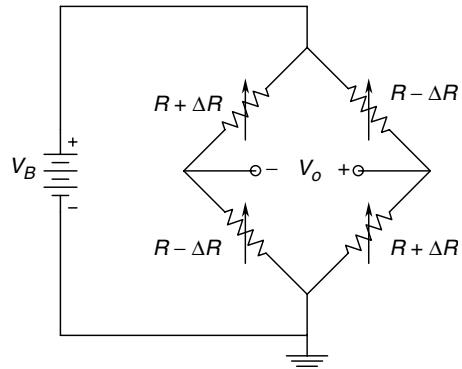
There are two approaches to categorizing sensors. One way, which we use in Chapter 7, is to group together all those different types of sensors used for a given application, such as the measurement of fluid pressure. Another method, which we use in this Chapter, is to group sensors on the basis of the mechanism by which they work, such as the generation of an open circuit voltage due to the input QUM, or a change in resistance proportional to the QUM.

6.2 Categories of Sensor Input Mechanisms

In this section, we present an outline of a comprehensive (yet incomplete) list of the categories of mechanisms by which sensors work. In the following sections, we elaborate on representative sensors, and their dynamic ranges, bandwidths and the auxiliary circuitry needed to produce a useful electrical (analog) output.

6.3 Resistive Sensors

As the input quantity varies, the resistance of practically all resistive sensors varies around some baseline or average value, R_o . Consequently, the most widely used means for converting the change in R due to the input to an output voltage is to include the transducer as an arm of a Wheatstone bridge (Section 4.2), or as part of an Anderson

**FIGURE 6.1**

Wheatstone bridge with four active arms. Bridge sensitivity is four times that of a bridge with a single variable resistance.

constant current loop circuit (Section 4.4). As long as the ratio $\Delta R/R_o$ is $\ll 1$, the bridge output voltage will be linear with ΔR . As $\Delta R/R_o \rightarrow 1$, the bridge output voltage vs $\Delta R/R_o$ tends to saturate, hence sensitivity and linearity are lost. In some resistive sensors, such as *unbonded strain gauge* force sensors, two resistors are available, whose values increase linearly ($R_o + \Delta R$) with the applied force, and two resistors whose values decrease with the input ($R_o - \Delta R$). These resistors can be assembled into a four-active arm Wheatstone bridge which has a linear output in ΔR (Figure 6.1).

6.3.1 Resistive Temperature Sensors

The electrical resistance of all metals and alloys increases with temperature. This increase can be modeled by a power series equation of the form:

$$R(T) = 25R + \alpha(T - 25) + \beta(T - 25)^2 \quad (6.1)$$

Here 25°C is taken as the reference temperature. The temperature coefficient (TEMPCO) of a resistive conductor is defined as α :

$$\alpha \equiv \frac{dR(T)/dT}{R(T)} \quad (6.2)$$

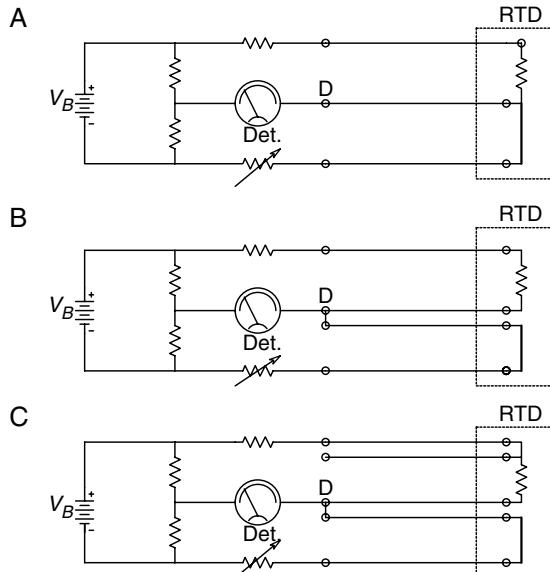
Table 6.1 gives tempcos and useful temperature ranges for various metals used as resistance thermometers, or *resistance temperature detectors* (RTDs).

Resistance temperature detectors (RTDs) are generally used in a Wheatstone bridge configuration. In order to compensate for lead lengths and thermoelectric EMFs at

TABLE 6.1

Properties of conductors used for RTDs

Material	Tempco (α)	Useful Range(°C)	Relative Resistivity (vs Cu)
Copper	0.00393	-200 to +260	1
Platinum	0.00390	-200 to +850	6.16
Nickel	0.0067	-80 to +320	4.4
Thermistors	-0.05	-100 to +300	-

**FIGURE 6.2**

Three methods for compensating for lead resistance in a Wheatstone bridge used with a remote RTD. The RTD is the resistance element in the rectangular block. (Source: Beckwith and Buck, 1961.)

junctions with the RTD material, extra leads are run to the measurement site, as shown in Figure 6.2.

Another type of RTD material which has a much higher tempco than pure metals or metallic alloys is the thermistor, which is made from amorphous semiconductor material, generally sintered mixtures of oxides, sulfides and silicates of elements such as Al, C, Co, Cu, Fe, Mg, Mn, Ni, Ti, U and Zn. The resistance of negative tempco (NTC) thermistors generally follows the rule:

$$R(T) = R_o \exp[\beta(1/T - 1/T_o)] \quad (6.3)$$

where T and T_o are kelvin temperatures and T_o is customarily taken to be 298°. The NTC thermistor tempco is easily calculated to be:

$$\alpha = \frac{dR/dT}{R} = -\beta \frac{1}{T^2} \quad (6.4)$$

The tempco evaluated at 298 K with $\beta = 4,000$ is -0.045 . When used in a Wheatstone bridge circuit, it is possible to resolve temperature changes of a milli °C.

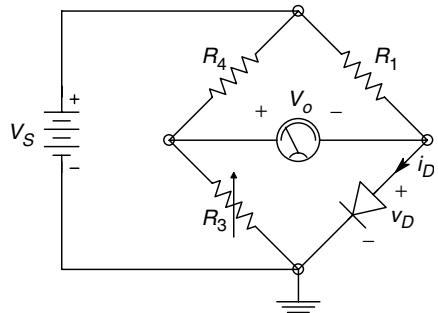
If the power dissipation of the thermistor is sufficient to cause it to warm to *above* ambient temperature, then fluid at ambient temperature moving past the thermistor will conduct heat away from it, cooling it and causing an increase in its resistance. Thus, this mode of operation may be used to measure fluid velocity when the fluid temperature is known.

Diodes, or *pn* junction devices can also be used as temperature sensors. For example, the current in a forward biased diode is given by the well known approximation:

$$i_D = I_{rs} [\exp(v_{Dq}/nkT) - 1] \quad (6.5)$$

where v_D is the voltage across the diode (positive for forward bias), n is a constant ranging from 1 to 2, k is the Boltzmann constant, T is the junction temperature in Kelvins, and I_{rs} is the *reverse saturation* current. Experimentally it has been observed that I_{rs} varies with temperature according to the relation [Millman, 1979]:

$$I_{rs}(T) = I_{rs}(T_o)2(T - T_o)/10 \quad (6.6)$$

**FIGURE 6.3**

A forward biased, Si, *pn* diode used as a nonlinear RTD in a Wheatstone bridge.

where $I_{rs}(T_o)$ is the reverse saturation current measured at temperature T_o . T_o is generally taken as 300 K. The reverse saturation current approximately doubles for each 10°C rise in temperature. In general, the current through a forward biased *pn* junction diode in a simple, series circuit with a resistor and a battery will increase, while the voltage across the diode will decrease with increasing temperature. Figure 6.3 illustrates a simple Wheatstone bridge circuit which can be used to measure temperature using a diode. In this case, the diode forward voltage decreases approximately by $2 \text{ mV}/^\circ\text{C}$.

6.3.2 Resistive Strain Gauges

Resistive strain gauges fall into two, broad categories—bonded and unbonded gauges. Bonded gauges consist of fine wires or conducting films which are cemented to some structural beam or machine part in which we wish to measure the strain. The minute elongation or compression of the beam, in response to a mechanical load, causes a length change in the bonded gauge conductors. This length change, in turn, causes a small change in the resistance of the conductors, which is usually sensed with a Wheatstone bridge circuit.

To examine how a strain gauge works, we consider a length of wire, L cm, with a circular area of cross-section $A \text{ cm}^2$, and a resistivity of $\rho \Omega\text{cm}$. The resistance of this wire (at a given temperature) is given by

$$R = \rho L / A \Omega \quad (6.7)$$

If the wire is stressed mechanically by a load W newtons (pounds), a stress of $s = W/A \text{ Pa}$ (psi) will occur. A is in m^2 for the SI (or MKS) units and in^2 for the English units. As a result of this stress, a strain, ε , will occur, given by:

$$\varepsilon = \Delta L / L = s / Y \text{ in/in} \quad (6.8)$$

where Y is the Young's modulus for the material of the wire, generally given in psi in the US. Using a Taylor's series expansion on R , we can write:

$$\Delta R = (-L/A^2)\Delta A + (L/A)\Delta\rho + (\rho/A)\Delta L \Omega \quad (6.9)$$

To obtain the fractional change in R , we divide the above equation with equation 6.7 for R :

$$\Delta R/R = -\Delta A/A + \Delta\rho/\rho + \Delta L/L \quad (6.10)$$

We now define the *Gauge Factor*, GF, as

$$GF = \frac{\Delta R/R}{\Delta L/L} = \frac{\Delta R/R}{\varepsilon} = 1 + \frac{\Delta \rho/\rho}{\Delta L/L} + \frac{-\Delta A/A}{\Delta L/L} \quad (6.11)$$

The right hand term in equation 6.11 is *Poisson's ratio*, μ . From equation 6.11, we can calculate the change in resistance, ΔR , for a given strain, ε . For example, if we load a 0.01" diameter steel wire with a one pound load, this causes a stress of $1/(\pi \times 0.0052) = 1.27 \times 10^4$ psi. The strain caused by this stress is $\varepsilon = s/Y = 1.27 \times 10^4 / 3 \times 10^7 = 4.23 \times 10^{-4}$ in/in or 423 microstrains. Assume the wire has an unstrained resistance of 220Ω , and the gauge factor, GF, is 3.2. ΔR is then given by:

$$\Delta R = GF\varepsilon R = 0.298 \Omega \quad (6.12)$$

The output voltage of an initially balanced, equal arm, 220Ω Wheatstone bridge with a 6 V excitation, V_s , which includes the wire described above as one arm would be:

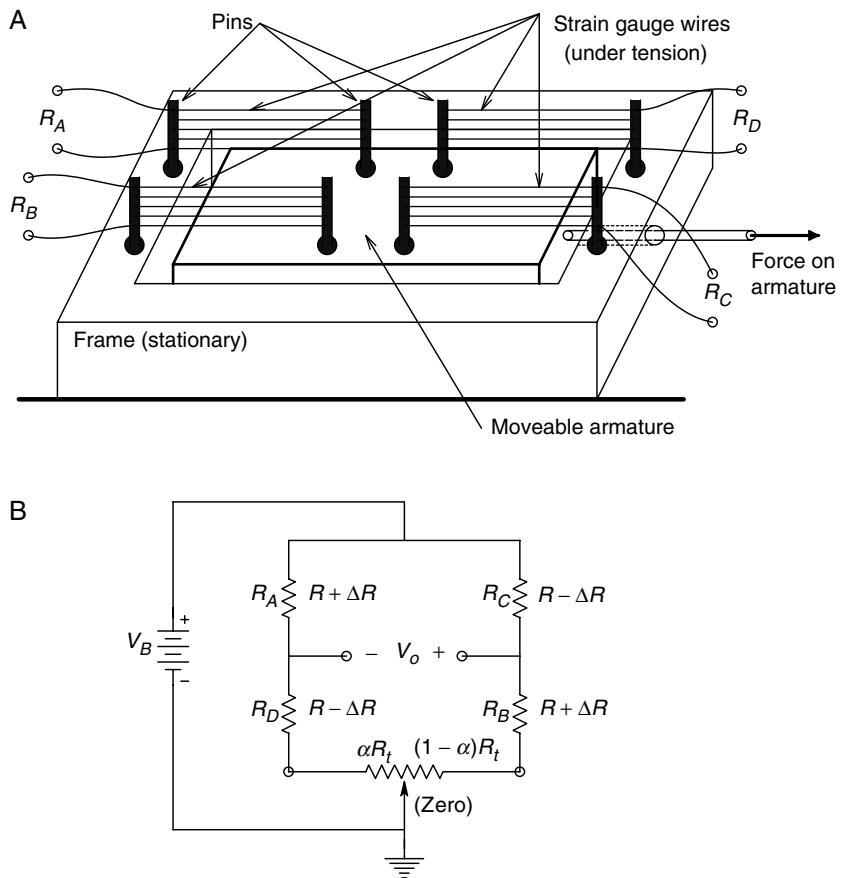
$$V_o = V_s \frac{\Delta R}{4R} = 6 \frac{0.298}{4(220)} = 2.03 \text{ mV} \quad (6.13)$$

The outputs of strain gauge bridges generally need amplification. Sensitivity to temperature changes can be reduced by using two matched gauges on one side of the bridge—the lower one is active, the upper one is unstrained and acts as a thermal compensator for the active bottom gauge. Both gauges are at the same temperature. See Beckwith and Buck (1961) for an excellent description of bonded strain gauge applications. A good discussion of the sources of error in bonded strain gauge systems can be found in Lion (1959).

The frequency response of bonded strain gauges largely depends on the mechanical properties of the structure to which they are bonded. It generally ranges from dc through to the audio range.

Unbonded strain gauges are used in several applications—the direct measurement of small forces and the measurement of pressure (pressure acts on a diaphragm or piston to produce a force). Figure 6.4A shows the innards of an unbonded force sensing strain gauge. There are two pairs of wires under tension. When force is applied in the direction shown, the lefthand pair is strained more and the right hand pair, less. Due to this asymmetry, the resistance of one pair of wires increases while that of the other pair decreases a like amount. The two pairs of wires are connected to form a Wheatstone bridge, having inherent temperature compensation, as shown in Figure 6.4B. Typically, the maximum $\Delta R/R$ attainable at full rated load is about 0.01. The bridge excitation voltage is practically limited, as in the case of bonded gauges, by power dissipation in the wires and the effects of heating on their resistances.

Unbonded strain gauges exhibit a mechanical resonance, which limits their high frequency response. The taut resistance wires and the moving armature form a spring and mass system which typically resonates at some audio frequency. Some unbonded strain gauge transducers are designed to have a full bridge output (i.e. two of their four wire elements are resistors which increase in like amounts, while two other wire elements decrease in resistance in like amounts due to the applied force). Zero output balance of a four-arm bridge at zero force is usually done with a potentiometer, connected as shown in Figure 6.4B.

**FIGURE 6.4**

A. A four-arm, unbonded strain gauge force sensor. The four resistance windings are under equal tension with zero applied force to the armature. When force is applied to the armature in the direction shown, strain increases in R_A and R_B , and decreases in R_C and R_D . B. Schematic diagram of the four-arm unbonded strain gauge force sensor. R_t is used to obtain an output null at zero or reference force.

6.3.3 Photoconductors

Photoconductors (PCs) are materials whose resistance decreases upon illumination with light. They should not be confused with photodiodes or solar cells, which can produce an EMF or short circuit current in response to the absorption of light quanta. Uses of photoconductors include exposure meters for cameras, light sensors in spectrophotometers, light sensors in a variety of counting systems where an object interrupts a light beam hitting the photoconductor, systems which sense a decrease of overall ambient illumination and turn on outside lighting (of course, the lighting must not be sensed by the photoconductor), oil burner safety systems and burglar alarms. PCs are also called *light dependent resistors* or *photoresistors*. Figure 6.5 illustrates the cross-sectional schematic of a typical PC device. Note that it is a thin wafer.

PCs can be made from a number of intrinsic as well as doped semiconductor materials. Each semiconductor has a distinct spectral response to light, ranging UV–FIR. Absorption of light quanta with a certain energy range by semiconductor substances produces electron hole pairs, which enter the conduction band and drift in an applied electric field, contributing to the current passing through the semiconductor. Greatest

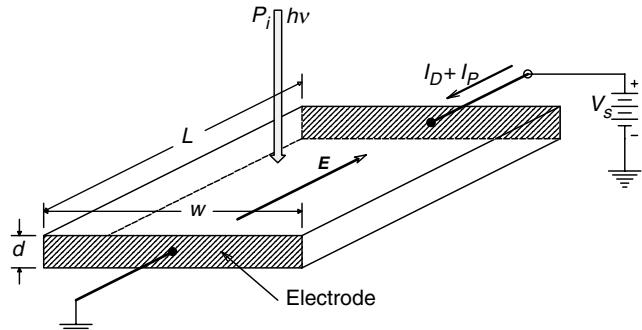


FIGURE 6.5
Diagram of a photoconductor slab.

photoconductor sensitivity occurs when the absorbed photons have energies equal to, or slightly greater than, the semiconductor's *bandgap energy*.

For example, cadmium sulphide (CdS), with an energy gap of 2.42 eV, works best in the visible range of wavelengths. Germanium and indium antimonide, with energy band gaps of 0.67 eV and 0.18 eV, respectively, are used in the near and far infrared wavelengths, respectively. EG&G Judson offers a J15D series of Mercury-Cadmium Telluride photoconductors which operate in the 2–22 μm band (FIR). These photoconductors are operated at 77 K and are used for infrared spectroscopy, thermography, CO₂ laser detection and missile guidance. Their response time constants are in the range 0.1–5 μs , depending on the lifetime of electrons in the crystal, which in turn depend on material composition and temperature. HgCdTe photoconductive detectors generate 1/ f noise, which limits their resolution below 1 kHz.

The CdS photoconductors are often made by depositing a thin layer of the material on an insulating (ceramic) substrate, often in a zigzag pattern to increase the effective length of the conductor's path and hence, its resistance.

The frequency response of photoconductors (in response to modulated light intensity) depends on, among other factors, the intensity and wavelength of the incident light, the kelvin temperature of the photoconductor, and the amount of metallic impurities in the semiconductor which can trap charge carriers and delay their spontaneous recombination. It can be shown that when the recombination time constant, τ_p , is large (corresponding to a low high frequency response), the gain of the photoconductor is high. A short time constant, giving good high frequency response, results in lower *gain* [Yang, 1988]. Gain is defined by Yang as the ratio of carriers collected by the ohmic contacts to the carriers generated by photons per unit time. The range of frequency response of photoconductors ranges from over 1 MHz to below the audio range. Germanium and lead sulphide have relatively broadband responses (tens of kHz), while materials such as selenium and cadmium sulphide have bandwidths in the hundreds of Hz.

Table 6.2 lists some of the materials used in PCs, their bandgap energies in eV and the wavelength of their peak spectral response.

In general, the total current in a PC can be written as:

$$I_{PC} = V_s[G_D + G_P] \quad (6.14)$$

where V_s is the bias voltage, G_D is the equivalent dark conductance and G_P is the photoconductance. G_P can be shown to be given by [Yang, 1988]:

$$G_P = \frac{I_P}{V_s} = \frac{q \eta \tau_p (\mu_p + \mu_n)}{L^2} \left[\frac{P_i \lambda}{hc} \right] S \quad (6.15)$$

TABLE 6.2
Properties of some photoconductors

PC Material (temp)	Bandgap Energy, eV	Wavelength of Peak Response, $\lambda_c \mu\text{m}$	Rise time/ Fall time
ZnS	3.60	0.345	
CdS	2.40	0.52	30 ms/10 ms
CdSe	1.80	0.69	15 ms/15 ms
CdTe	1.50	0.83	
Si (intrinsic)	1.12	1.10	1 μs /1 μs
Ge (intrinsic)	0.67	1.85	0.1 μs /0.1 μs
PbS	0.37	3.35	
InAs	0.35	3.54	$\approx 1 \text{ ns}$
Te	0.33	7.75	
PbTe	0.30	4.13	
PbSe	0.27	4.58	2 μs
HgCdTe (77 K)		5.0	5 μs
InSb (77 K)	0.18	6.90	
GeCu (4 K)		25	
GeBe (3 K)		55	

where τ_p is the *mean lifetime of holes*

$\mu_p = |v_p/E|$ = hole mobility in cm^2/Vs

$\mu_n = |v_n/E|$ = electron mobility in cm^2/Vs

v_p and v_n = mean drift velocities of holes and electrons, respectively

E = the uniform E -field in the semiconductor

q , η , P_i , λ , h & c have been defined above

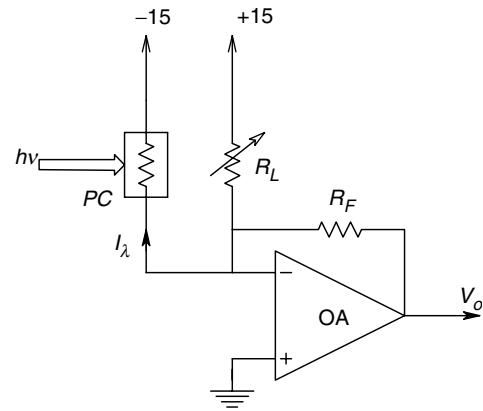
Note that in general, $\mu_n > \mu_p$. L is the distance between the electrodes. The expression for G_p is an approximation, valid up to the cut-off wavelength, λ_c . $[P_i \lambda/hc] = \Phi_i$, the incoming number of photons/s on area $wL \text{ m}^2$. The dark conductance of a Si PC can be found simply from the room temperature resistivity of Si, ρ , and the geometry of the PC. For example:

$$G_D = A/\rho L = wd/\rho L = 0.2 \text{ cm} \times 0.001 \text{ cm}/(2.3 \times 10^3 \Omega \text{ cm} \times 0.02 \text{ cm}) \\ = 4.348 \times 10^{-6} \text{ S} \quad (6.16)$$

The photoconductance for a Si PC, illuminated by $1 \mu\text{W}$ of 512 nm photons is, from equation 6.15:

$$G_p = \frac{1.6 \times 10^{-19} \times 0.8 \times 10^{-4}(1350 + 450)}{10^{-4}} \left[\frac{\frac{10^{-6} \times 512 \times 10^{-9}}{6.625 \times 10^{-34} \times 3 \times 10^8}}{\frac{\text{Js}}{\text{m/s}}} \right] \\ = 5.935 \times 10^{-4} \text{ S} \quad (6.17)$$

A great advantage of PCs is their unique ability to respond to MIR and FIR photons at wavelengths not sensed by PIN PDs or APDs. We note that for various materials,

**FIGURE 6.6**

An op-amp is used as an operational transresistor to condition the photocurrent from a photoconductor. The op-amp should have low dc drift, such as found in a chopper design.

as the bandgap energy decreases, λ_c increases and the response time constants of the PCs decrease.

Certain PCs are also unique in being able to convert incident X-ray photons to a conductance change. X-ray sensing PCs are made of *amorphous selenium* (*a*-Se) [Soltani *et al*, 1999]. An *a*-Se PC with $E = 0 \text{ V}/\mu\text{m}$ produces ca. 10^3 electron hole pairs per 50 keV X-ray photon. There is a 50% attenuation of a 50 keV electron beam in $d = 365 \mu\text{m}$ *a*-Se. Soltani *et al* describe a charge coupled, X-ray photon sensing array using *a*-Se sensors that have superior image resolution to phosphor X-ray sensors.

Figure 6.6 illustrates a simple op-amp circuit that gives $V_o \propto P_i$. Note that R_C is used to cancel the PC's dark current. A Wheatstone bridge can also be used to convert photoconductance to output voltage, albeit nonlinearly. Unlike PDs, PCs make thermal noise. It can be shown that the total MS noise current input to the op-amp's summing junction is:

$$\overline{i_{ntot}^2} = \{4kT[G_D + G_P + G_C + G_F] + i_{na}^2\}B \text{ msA} \quad (6.18)$$

Note that the MS noise increases with input light power because G_P increases with P_i . In addition to the noise current, the op-amp also has an equivalent short circuit input noise voltage, $e_{na}^2 B \text{ mV}$.

6.3.4 Conductive Relative Humidity Sensors

Humidity sensors respond to the amount of water vapor in the air or other gas. In this section, we will consider only those humidity sensors that change their conductance in response to the partial pressure of water vapor to which they are exposed. In one embodiment, the Dunmore sensor, two noble metal wires are wound in the form of two, non-touching helices on an insulating bobbin or form. A thin, hygroscopic coating, consisting of an ionizable salt, covers the wires. As the humidity increases, the salt in the coating reversibly absorbs water vapor and dissociates into ions. The ions conduct electric current, which is an increasing function of the humidity. This sensor's conductance increases with humidity. Since this sensor is an electrochemical system, the use of dc excitation for the bridge and sensor will lead to polarization, and possible electrolysis and chemical breakdown of the hydrated salt. Polarization due to ion migration causes the apparent resistance of the sensor to increase and hence, causes erroneous readings. The ionizable salt humidity sensors must be used with low voltage, low frequency ac excitation to prevent these problems.

Lithium chloride (LiCl), potassium dihydrogen phosphate (KH_2PO_4) and aluminum oxide (Al_2O_3) have been used in various conductive humidity sensor designs [Lion, 1959]. The sensors using only LiCl appear to respond to a limited range of humidity, which is dependent on the LiCl concentration. The sensor resistance is strongly dependent on temperature, which affects water absorption and dissociation of the salt. At constant temperature, the LiCl sensor resistance decreases logarithmically over a three or four decade range, in proportion to the relative humidity.

Figure 6.7A illustrates a wide range relative humidity sensor described in Lion (1959), attributed to Cutting, Jason & Wood (1955). An aluminum rod is coated at one end with an Al_2O_3 layer, followed by a porous conducting layer of graphite or vapor deposited metal. This sensor is used with ac excitation. Figure 6.7B [Lion, 1959], shows the that the resistance of the sensor varies from about $5 \text{ M}\Omega$ at 0% RH to about $2 \text{ k}\Omega$ at 100% RH. Of special interest is the fact that there is a corresponding increase in sensor capacitance

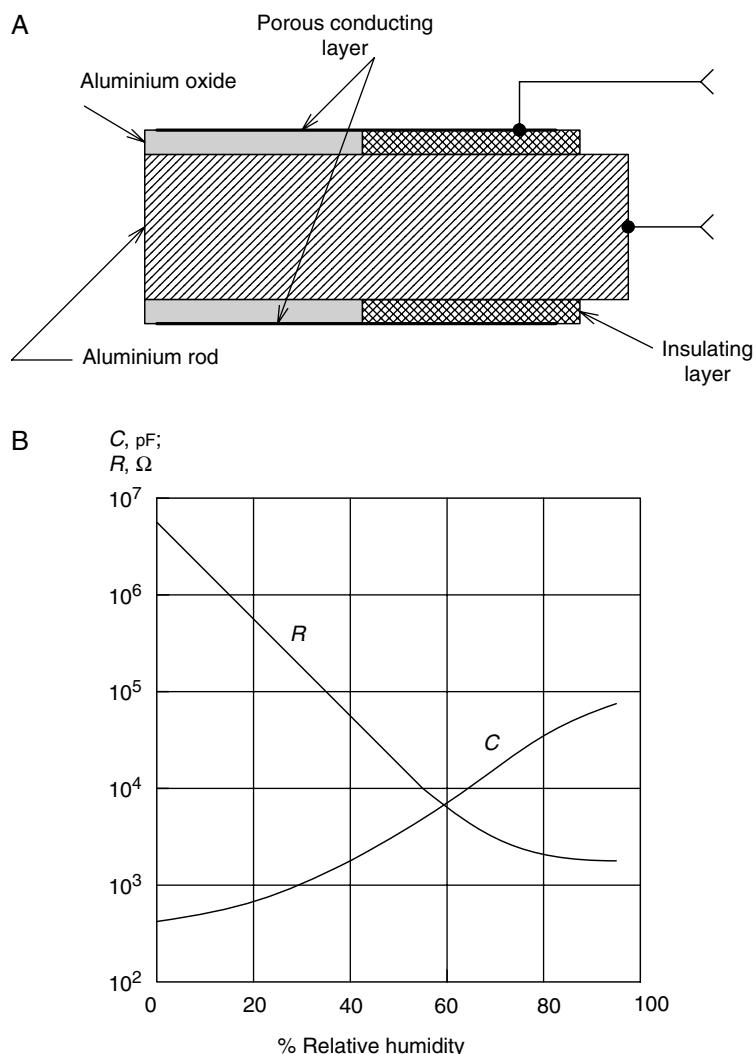


FIGURE 6.7

A. Cross-section through an aluminum oxide humidity sensor. (Source: Lion, 1959.) B. Plot of an aluminum oxide humidity sensor's equivalent parallel resistance and capacitance as a function of relative humidity. (Source: Lion, Figure (1-7)8, 1959.)

from about 500 pF at 0% RH to 105 pF at 100% RH. Hence, an ac capacitive sensing bridge can also be used to obtain sensor output, as well as an ac Wheatstone bridge. This humidity sensor is unique in that it is claimed to be temperature insensitive over a -15 to 80°C range.

Commercial humidity measurement systems and sensors are offered by several companies. Omega Corp. markets a number of different humidity and temperature sensing systems. However, they use capacitance humidity sensors, rather than conductive types. Ohmic Instruments Co. offers a series of conductive humidity sensors based on water vapor absorption by dissociable salts. They offer ten different, Dunmore type sensors, each with a narrow range of RH sensitivity. Sensor resistances vary in the range $>10\text{ M}\Omega - <1\text{ k}\Omega$ over their narrow ranges of RH sensitivity. A broad range instrument (model WHS 5-99L), covering 5–90% RH uses four overlapping, narrow range sensors in a weighted resistor network. A 60 Hz excitation is used.

Thunder Scientific Corp. has developed and marketed, what they call, the 'Brady Array' humidity sensor. From published technical specifications, this sensor is a precise array of semiconductor crystals and interstitial spaces. Physical adsorption of water molecules from water vapor somehow causes an increase in the conduction band electrons in the crystals, raising the conductivity. Unlike the Dunmore type sensors, which require water vapor adsorption and ionic dissociation, the Brady sensor operates on a purely physical basis and thus, has a very rapid response time to step changes in RH (in hundreds of milliseconds) *vs* minutes for Dunmore sensors. The simplest circuit for a Brady sensor consists of a 5 V RMS, 1000 Hz voltage source, in series with the sensor and a $1\text{ M}\Omega$ resistor. The voltage across the $1\text{ M}\Omega$ resistor is proportional to the RH. Brady sensors work over a range from nearly zero to 100% RH. They have been shown to work normally in atmospheres of water vapor in air, oxygen, nitrogen, CO_2 , hydrogen, natural gas, butane and vapors from various hydrocarbons such as acetone, trichloroethylene, jet fuel, gasoline, kerosene, oils, etc. Their working temperature range has been given -20 to 70°C . Typical accuracy is stated as $\pm 2\%$ of indicated RH and the resolution is 0.1% RH.

Temperature compensation of a Brady sensor is easily accomplished by taking a like sensor and hermetically sealing it so that it is at 0% RH, and then using the sealed sensor as an element in the reference side of a Wheatstone bridge (Figure 6.8). A temperature compensated Brady array is claimed to drift no more than $1\text{ mV}/^\circ\text{F}$ over the -20 to 70°C operating range. This is 0.02% of the full scale per $^\circ\text{F}$.

Calibration of resistive (and other kinds) of RH sensors is traditionally done by using saturated salt solutions at known temperatures. For example, at 25°C , the RH over a saturated aqueous $\text{LiCl}(\text{H}_2\text{O})$ solution is 11%, over $\text{MgCl}_2(6\text{H}_2\text{O})$ it is 33%, over

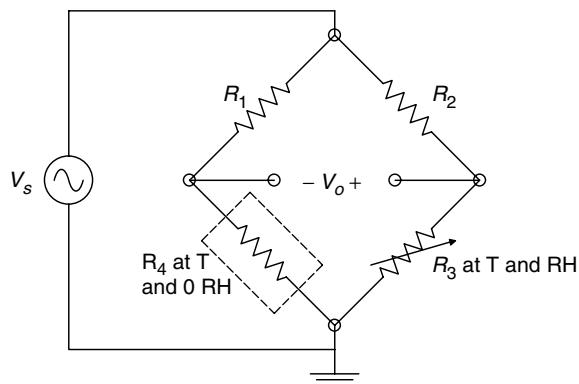


FIGURE 6.8

Two Brady array humidity sensors are used in a Wheatstone bridge to obtain temperature compensation over a wide range; sensor R_4 is sealed at 0% RH, and responds only to temperature changes.

$\text{Na}_2\text{Cr}_2\text{O}_7(2\text{H}_2\text{O})$ it is 54%, over NaCl it is 75%, over K_2CrO_4 it is 86%, over $(\text{NH}_4)\text{H}_2\text{PO}_4$ it is 93% and over K_2SO_4 it is 97%.

6.3.5 Use of Resistance Change to Sense Position or Angle

Direct measurement of shaft angle in applications such as electromechanical, position feedback control systems is often done by coupling the shaft either directly, or through gears, to a single- or multi-turn potentiometer. The potentiometer is one arm of a Wheatstone bridge, or is used as a voltage divider. Often, a feedback linearization circuit is used with a bridge, such as shown in Figure 6.9, to produce a voltage output which varies linearly with potentiometer resistance (shaft angle). The potentiometers used in such shaft angle transduction must be linear and have high resolution. The resolution of a standard wire wound, three-, ten- or fourteen-turn potentiometer is limited by the wiper connected to the shaft making contact with individual turns of the helically wound resistance wire. One way to avoid this problem is to use potentiometers which have a continuous, cermet or conductive plastic resistance materials.

Potentiometers used as angular position sensors are typically specified to have independent linearity over their total range of rotation. The Bourns 3400 series of ten-turn, wire wound potentiometers typically have $\pm 0.15\%$ independent linearity and are available with $\pm 0.05\%$ independent linearity. The Bourns 3400 series has a resistance tempco of $+0.007\text{%/}^{\circ}\text{C}$. Conductive plastic potentiometers, such as the Beckman single turn 6550 series, have independent linearities of $\pm 0.5\%$ and tempcos of -150 to $+300$ ppm. Cermet potentiometers, such as the single turn Beckman 6300 series, also have independent linearities of $\pm 0.5\%$, but have lower tempco ranges of ± 100 ppm.

In addition to precision rotational potentiometers for angle sensing, linear displacements can be measured with linear potentiometers (i.e. pots whose resistance varies linearly with the linear position of the wiper). The Beckman series 400 rectilinear potentiometer series uses an infinite resolution, cermet resistance element. Wiper travel lengths range from 1.5 cm to 15.5 cm in various 400 series models. Linearity ranges from $\pm 1.35\%$ for the 0.5 cm model to $\pm 0.25\%$ in the 15.5 cm model.

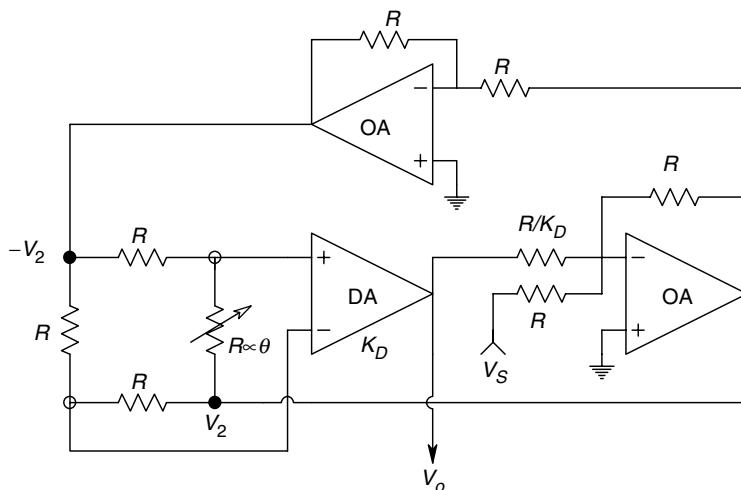


FIGURE 6.9

A Wheatstone bridge output linearization circuit, useful in applications where the variable resistance varies linearly over a wide range as a function of a parameter under measurement, such as shaft angle.

A disadvantage of potentiometers as position sensors is that they require some extra torque or force to move their wipers. Another disadvantage is that because of the wiping action, they eventually wear out. However, optical coding disks need no wipers and are limited in resolution by the number of output bits available. An LVDT can operate with negligible friction if a teflon lining is used in its core. Still, potentiometers offer an inexpensive means of sensing angular or linear position in a number of applications.

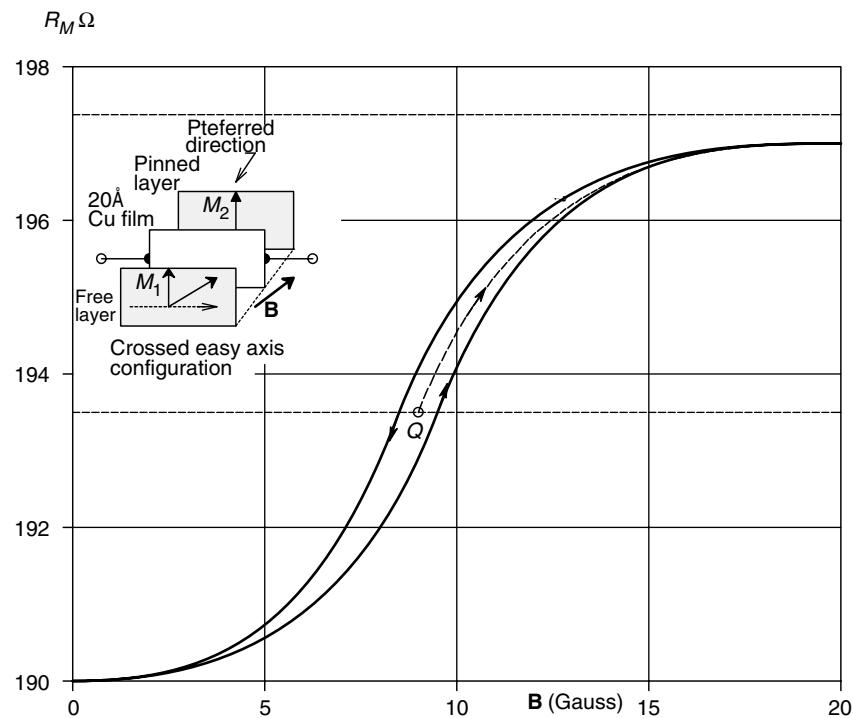
6.3.6 Giant Magnetoresistive Effect Based Sensors

The giant magnetoresistive (GMR) effect is a physical phenomenon in which a magnetic field changes the resistance of a layered, thin film sensor. GMR is a relatively new effect, discovered in the late 1980s independently by Peter Gruenberg of the KFA Research Institute in Julich, Germany and Albert Fert at the University of Paris-Sud. Researchers at IBM extended the initial work on GMR devices and developed sensitive GMR transducers based on sputtered thin film technology. IBM's interest in developing sensitive, miniature GMR devices has been to improve the data density stored on computer floppy and hard drive magnetic discs. IBM also calls their GMR sensor the *spin valve read-head*.

GMR sensors have a resistance that is modulated by an imposed magnetic flux density and thus, a Wheatstone bridge can be used to sense the $\Delta R/R_M$ and consequently, ΔB . Unfortunately, GMRs are not particularly linear, which makes them acceptable for binary applications, but presents problems for linear analog sensing, unless feedback is used.

A basic GMR sensor is a multiple layer, thin film device. In its simplest form, it has a non-magnetic, conductive spacing layer sandwiched between two ferromagnetic film layers. Usually, the magnetization in one ferromagnetic layer is fixed, or *pinned*, along a set direction lying in the plane of the layer. In the absence of an external \mathbf{B} field, the resistance of the conductive layer is maximum, R_m . The magnetization of the *free magnetic layer* is allowed to rotate in response to an externally applied magnetic field vector lying in the plane of the free layer. The lowest GMR sensor resistance occurs when the pinned and free layers are magnetically oriented in the same direction. There can be a 10–15% decrease in R_m due to the externally applied \mathbf{B} , before the GMR effect saturates at high $|\mathbf{B}|$. If the \mathbf{B} vector reverses sign, we observe basically the same negative ΔR , except for the small hysteresis present. This is because electrons in the conductor layer, with parallel spin directions, move freely in both films. Higher resistance occurs when the magnetic orientations in the two films oppose each other, since movement of electrons in the conductor of either spin direction is hampered by one or the other magnetic films. Depending on the strength and orientation of the applied external \mathbf{B} field, the pinning direction, the materials used and their thicknesses, the $R_M = f(B_x, \theta, \varphi)$ curve can have a variety of shapes [Wilson, 1996]. Note that ideally, for maximum sensitivity, $\mathbf{B} = \mathbf{B}_x$ (i.e. \mathbf{B} is parallel with the x axis; the GMR sensor lies in the xz plane). Using polar coordinates, φ is the angle \mathbf{B} makes with the y axis, which is perpendicular to the xz plane, and θ is the angle the projection of \mathbf{B} on the xz plane makes with the x axis. So, for maximum sensitivity, $\theta = 0$, and $\varphi = 90^\circ$.

In many cases, the GMR $R_M = f(B_x)$ curve exhibits both hysteresis and some dead zone as shown in Figure 6.10. This GMR sensor has the 'crossed easy axis' configuration, in which the applied \mathbf{B} is perpendicular to the free layer easy axis and the pinning layer magnetization direction is parallel to \mathbf{B} . The response is linear around a Q -point, Q , and is relatively hysteresis free [Wilson, 1996]. Due to its magnetic sensitivity, the GMR must

**FIGURE 6.10**

Resistance *vs* applied magnetic field in Gauss for a typical GMR magnetosensor.

be magnetically shielded from the Earth's field and the stray magnetic fields from manmade power wiring and machinery. To use such a sensor as a linear device, it must be magnetically biased around the Q -point.

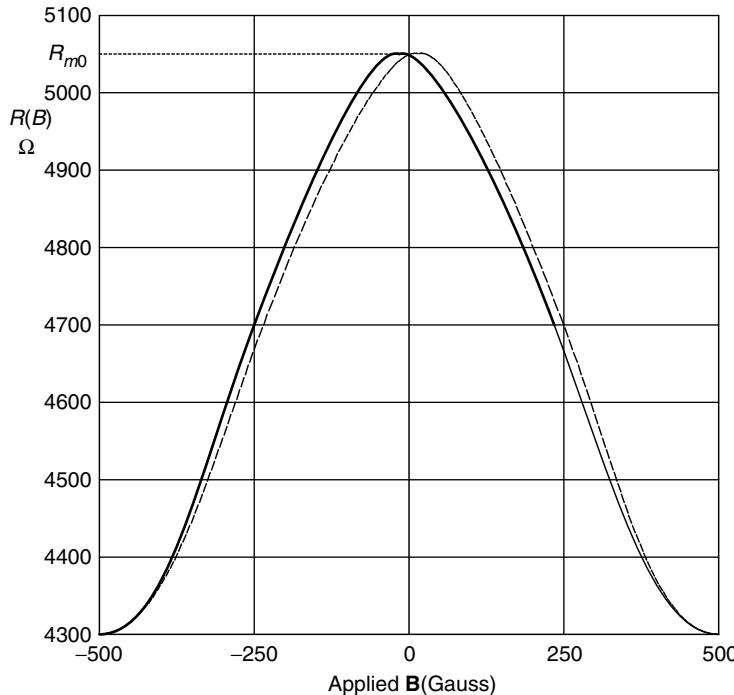
Much research is currently underway to improve the linearity and sensitivity of spin valves or GMR sensors. Attention is being paid to the use of multilayered thin films of various compositions, including two pinned layers; 6–12 layers are being studied by Fujitsu [Kanai *et al.*, 2001].

GMR sensors are available commercially. Infineon Technologies offers a 4-element GMR IC package in which the four GMRs are connected in a bridge. Their GMR B6 is intended as a rotary position sensor in that the bridge output is a sinusoidal function of the input magnet's B field angle. The magnetic field-free value of each resistor is ca. 700Ω . Rhopoint Components of Oxted, Surrey, UK offers a 12-sensor GMR device kit for experimenters and R&D use.

Nonvolatile Electronics Inc., Eden Prairie, MN (NVE) also markets GMR sensor assemblies based on anti-ferromagnetic coupling. The NVE GMR sensors have fairly linear $R(B)$ characteristics, showing maximum resistance at zero gauss (Figure 6.11). The NVE sensor was determined graphically to have a maximum $\Delta R/\Delta B = 12.9\%$, and fractional sensitivity, S_B , of:

$$S_B = \frac{\Delta R/R_{m0}}{\Delta B} = 3.575 \times 10^{-4} \text{ gauss}^{-1} \quad (6.19)$$

NVE claims that if one of their GMR sensors uses a high permeability external flux concentrator and 5 V excitation, 100 gauss will give a 350 mV voltage change. They compare that to a 5 mV output from a Hall sensor, given the same input.

**FIGURE 6.11**

Magnetoresistance *vs* Gauss characteristics of a commercial GMR sensor. Note curve is even in B and has a slight hysteresis.

NVE makes gradiometer GMR assemblies in which two matched GMR resistors are placed at each end of a rectangular bar, spaced at a distance of 0.5 or 1.0 mm. The four GMR sensors are connected in a Wheatstone bridge circuit. The gradiometer gives no output if the magnetic field is equal at each sensor. If a magnetic field density gradient exists over the separation length of the sensors, two at one end (S_1 and S_3) will be in a stronger field than the pair (S_2 and S_4) at the other end. Hence, the bridge is unbalanced and produces an output. Figure 6.12 illustrates the gradiometer chip and bridge circuit. If there is no gradient in \mathbf{B}_x , which means $d\mathbf{B}_x/dx = 0$, then assume that all four of the GMR sensors are subject to equal \mathbf{B}_x , all have resistance $R_o - \delta R$ and the bridge is balanced. Now let \mathbf{B}_x at the right end of the chip be larger by some amount, ΔB . This causes the resistances of sensors S_1 and S_3 to be smaller than those of S_2 and S_4 . In summary, with $\Delta B > 0$, $R_{s1} = R_o - \delta R_+ = R_{S3}$ and $R_{S2} = R_o - \delta R = R_{S4}$, where clearly, $\delta R_+ > \delta R$. Using the voltage divider relation, we can find the unbalanced bridge output.

$$\begin{aligned} v_o = v_1 - v_2 &= V_B \left[\frac{R_o - \delta R}{2R_o - (\delta R + \delta R_+)} - \frac{R_o - \delta R_+}{2R_o - (\delta R + \delta R_+)} \right] \\ &= V_B \left[\frac{(\delta R_+ - \delta R)}{2R_o - (\delta R + \delta R_+)} \right] > 0 \end{aligned} \quad (6.20)$$

If $(\delta R + \delta R_+) \ll 2R_o$, then:

$$v_o \cong \frac{V_B}{2R_o} (\delta R_+ - \delta R) = \frac{V_B}{2R_o} (\Delta \delta R) > 0 \quad (6.21)$$

Note that if the gradient in \mathbf{B}_x is reversed, then $v_o < 0$.

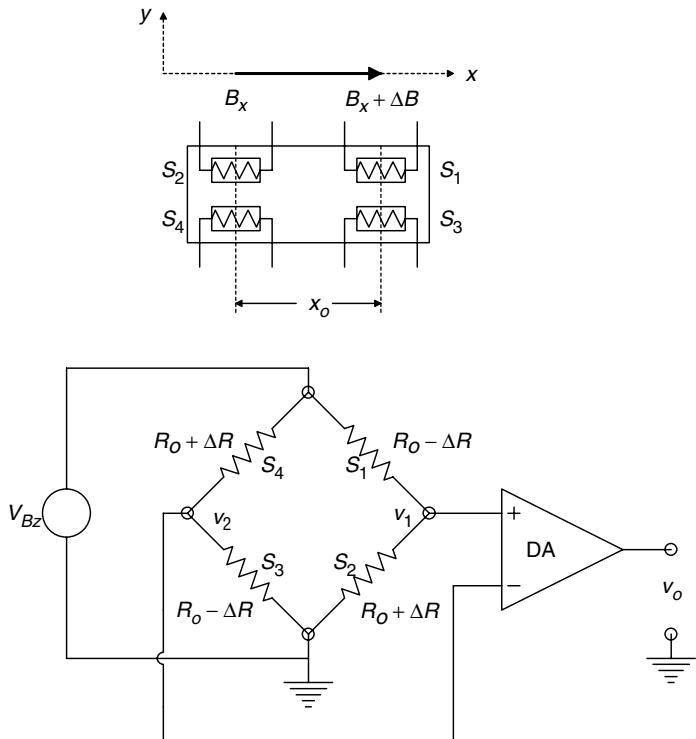


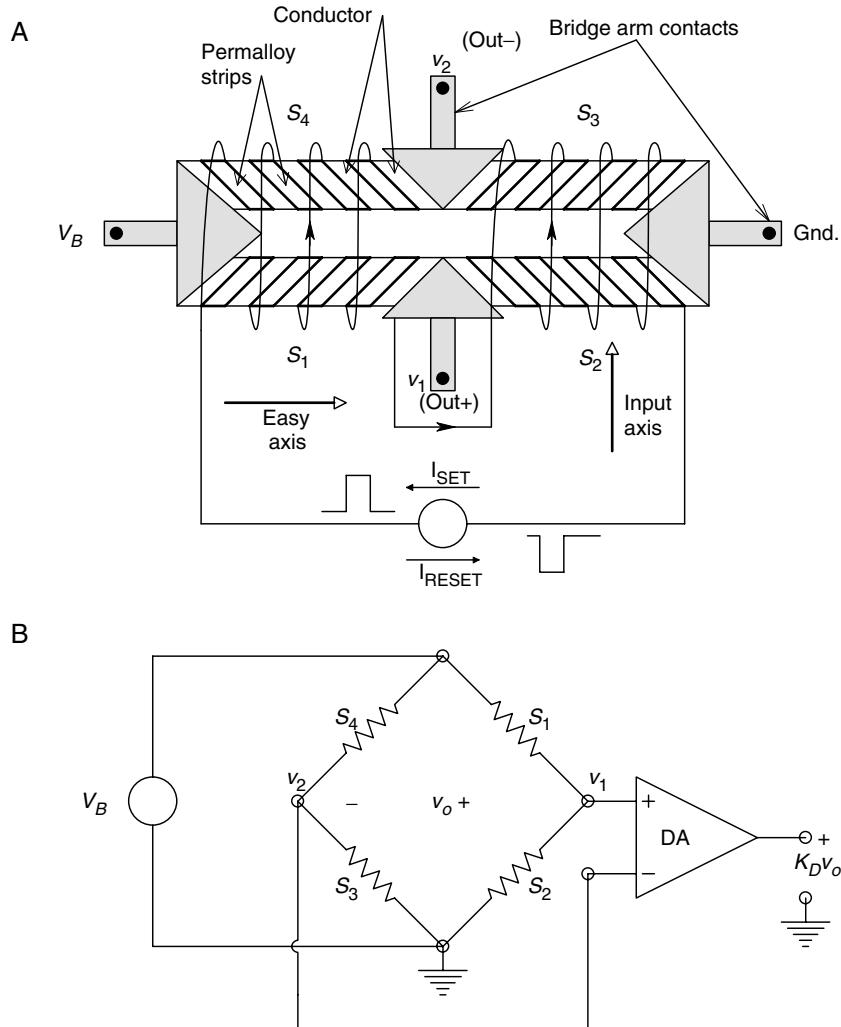
FIGURE 6.12
Circuit and diagram of a GMR gradiometer Wheatstone bridge.

6.3.7 Anisotropic Magnetoresistance (AMR)

The basic magnetoresistive effect was first described by Lord Kelvin (William Thompson) in 1856, in bulk ferromagnetic materials. Like many great discoveries in science, it took over 100 years for the development of thin film, lithographic, semiconductor technology to realize practical AMR devices. The basic AMR device consists of strips of a thin film of Permalloy (nickel-iron alloy) deposited on a silicon wafer and arranged in a 'barber pole' geometry [Caruso *et al*, 1999a]. Figure 6.13 illustrates four AMR sensors arranged in a Wheatstone bridge.

The phenomenon of anisotropic magnetoresistance can be observed in any ferromagnetic material. It is best described in terms of quantum mechanics and is due to spin-orbit interactions in the conduction band electrons in the metal. The anisotropy in AMR refers to the fact that the resistance of the sensor is a function of the angle between the current density vector, J_x , in the sensor and the internal magnetic flux density, B_i . B_i is the vector sum of the internal B_e (the QUM) and the internal, fixed B_m . That is, $B_i = B_e + B_m$. Figure 6.14 illustrates the \mathbf{B} vector summation at a point P in the AMR film. Note that the larger B_e is in the film, the lower the angle between the current density vector, J , and the net internal flux density vector, B_i . The AMR sensor resistance increases nonlinearly with decreasing angle, θ_i ; it is maximum when $\theta_i = 0$. The graph in Figure 6.14 illustrates how the vector addition of B_e and B_m affect AMR sensor resistance when the initial angle between B_m and J is 45° .

The sensor's $R_{AMR} = f(|\mathbf{B}_i|)$ can be linearized by tinkering with the sensor geometry and biasing [Ripka *et al*, 1998]. AMR device composition is generally proprietary, but thin films of permalloy are often used. AMR sensors made by Honeywell, NVE and Philips have integrated feedback coils and reset (flipping) coils used to saturate B_m of the sensor material in a direction orthogonal to the input axis (preferred direction), to avoid

**FIGURE 6.13**

A. Four anisotropic magnetoresistor devices arranged as a linearized magnetosensor. B. Wheatstone bridge circuit of the 4-element AMR sensor.

distortion of the sensor characteristic caused by perpendicular fields [Ripka *et al*, 1998]. Bridge output sensitivity is usually given in V/V/applied magnetic field. The first V is the bridge output voltage, the second V is the bridge excitation voltage, \$V_B\$, which is generally around 3–5 V. The magnetic field is given by some writers in Oersteds (cgs)—an Oe is a unit of *magnetizing force*. SI magnetizing force is \$\mathbf{H}\$ in A/m. Still others specify magnetic field, \$\mathbf{B}\$, in gauss (cgs). Note that \$10^4\$ gauss equals 1 tesla = 1 weber/meter² (SI units). The units gauss and tesla are units of *magnetic flux density*. In air, a uniform, 1 Oe magnetizing force creates a 1 gauss flux density, that is, \$B = \mu_0 H\$. Why cannot manufacturers of magnetic devices consistently use SI units? Complain!

Useful input flux density range for AMR bridge sensors is from ca. 5 pT–1 mT. By contrast, GMR sensors useful range is \$10\text{ }\mu\text{T}\$–\$10^4\$ T. A SQUID sensor (Chapter 8) works between 10 fT and 10 T (Wb/m²) [Caruso *et al*, 1999a]. The horizontal component of the Earth's magnetic field in Cambridge, MA, is ca. \$1.7 \times 10^{-5}\$ Wb/m² or 0.17 gauss, the

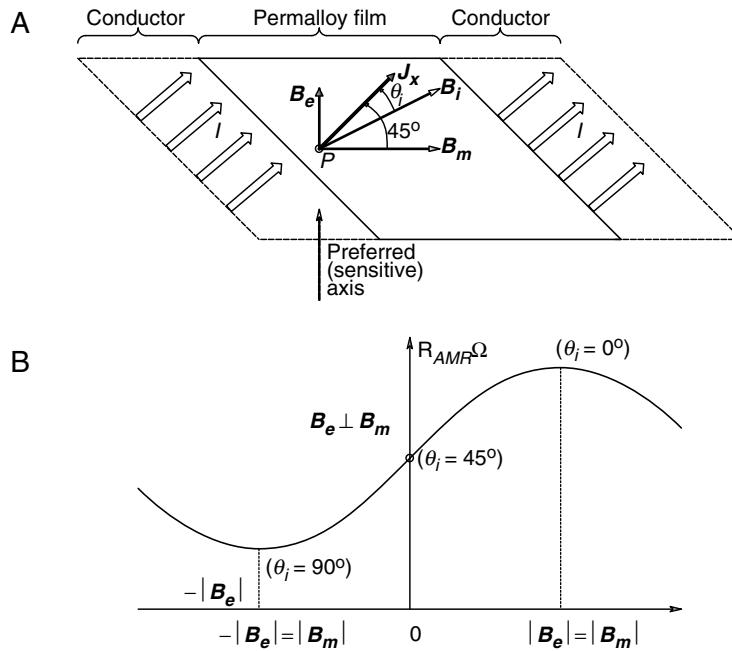


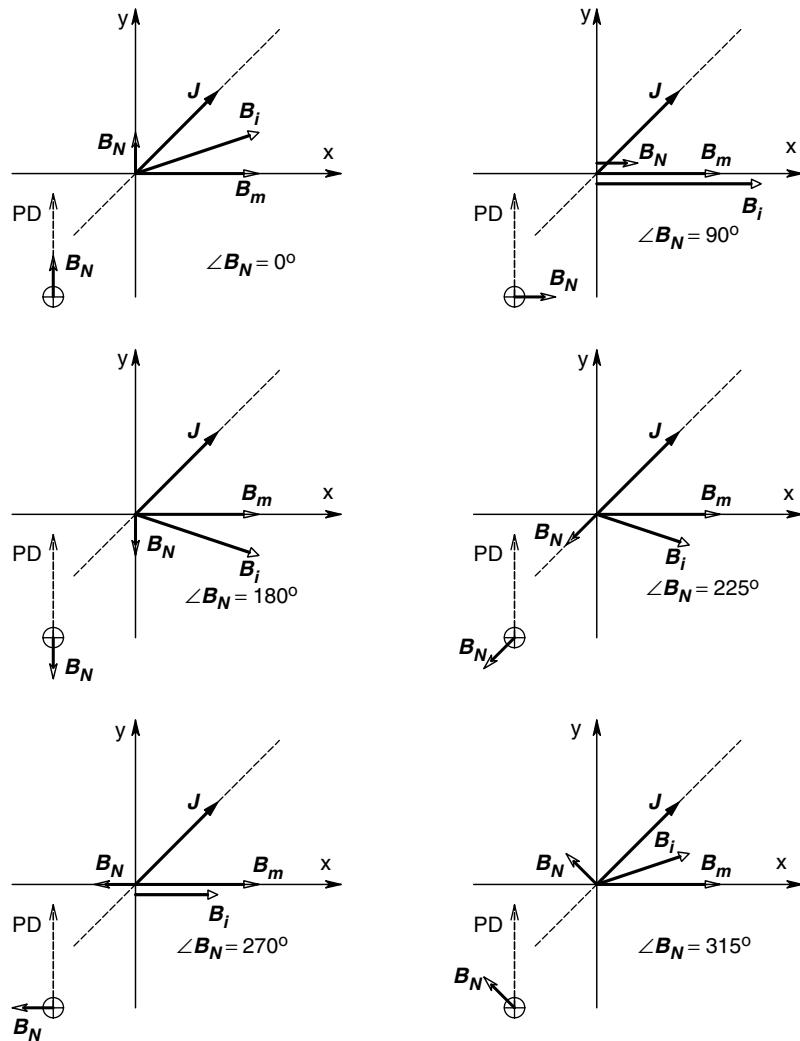
FIGURE 6.14

A. B vector summation at a point in an AMR film. B. R_{AMR} vs applied $|B_e|$ for the AMR sensor.

vertical component is ca. 0.55 gauss [Sears, 1953]. These values are well within the sensitivity range of AMR sensors.

Consider the example of designing a simple AMR sensor compass, designed for use on a level platform, tangential to the Earth's surface. A handheld AMR compass with a bubble level, used by hikers, and a gimbaled AMR compass in a motor vehicle are representative applications. An AMR compass to be used on a platform that can experience roll, pitch and yaw, such as a boat or aircraft, requires three AMR sensors with preferred directions along the x , y and z axes, as well as roll and pitch inputs from rate integrating gyros or clinometers [Caruso and Withanwasam, 1999b]. The stable, horizontal platform compass uses two AMR sensors with orthogonal preferred axes along the x and y axes defining the plane of the platform. Figure 6.15 illustrates the vector magnetic field conditions inside one sensor strip as a function of the direction of the external magnetic field vector, \mathbf{B}_N , with respect to the sensor's preferred input direction (PD). Many such sensor strips are incorporated in bridge sensor, S_1 . The second sensor bridge (S_2) is close to S_1 and has its PD rotated 90° clockwise from PD₁. Because of the vector summation of the internal magnetizing flux density, \mathbf{B}_m , and the horizontal component of the Earth's magnetic field, \mathbf{B}_N , the output of the S_1 bridge varies sinusoidally with the compass direction, θ . The θ in this example is measured as the angle between PD₁ and the north-pointing vector, \mathbf{B}_N . Thus, for example, if PD₁ is pointing toward magnetic East, $\theta = +90^\circ$. We assume the S_1 sensor bridge output voltage, V_{o1} , follows the magnetically induced resistance change and can be modeled by the relation,

$$V_{o1} = (V_{min} + \Delta V/2) + (\Delta V/2) \cos \theta \quad (6.22)$$

**FIGURE 6.15**

Six vector diagrams illustrating the conditions in an AMR sensor used to sense the Earth's magnetic field, \mathbf{B}_N . PD is the sensor's preferred direction (direction of maximum response). \mathbf{J} is the current density vector in the AMR film, \mathbf{B}_m is the internal magnetizing flux density and \mathbf{B}_i is the net internal flux density vector.

It is easily shown that the voltage output of the orthogonal S2 bridge follows the approximate relation:

$$V_{o2} = (V_{min} + \Delta V/2) - (\Delta V/2) \sin \theta \quad (6.23)$$

Equations 6.22 and 6.23 can be solved simultaneously to find an expression for θ :

$$\theta = \tan^{-1} \left[\frac{(V_{min} + \Delta V/2) - V_{o2}}{V_{o1} - (V_{min} + \Delta V/2)} \right] \quad (6.24)$$

A microcomputer can calculate θ from the variable inputs V_{o1} and V_{o2} , the constants V_{min} and ΔV , and display it digitally in degrees of heading magnetic. Note that because $|\mathbf{B}_N|$ remains constant and only its angle changes, V_{min} and ΔV remain constant.

Many other interesting applications of AMR sensor bridges exist. For example, counting moving vehicles based on their perturbation of the Earth's magnetic field or the field of manmade magnets, is possible using 2- or 3-axis AMR sensor arrays. Vehicle counting can also be used to ensure efficient operation of parking lots and garages. The direction of vehicle travel can also be determined. Another application of the AMR sensor bridge is galvanically isolated current measurements, from dc up to 10–15 MHz. The tangential component of the solenoidal B -field surrounding a current carrying wire at some point P at a radius R from the center of the wire is given by:

$$\mathbf{B}_T = \mu_0 I / (2\pi R) \text{ Wb/m}^2 \quad (6.25)$$

Here, we assume that the wire is of infinite length and its radius $r \ll R$ [Sears, 1953]. If lower currents are to be measured, the field at the center axis at the end of an N -turn

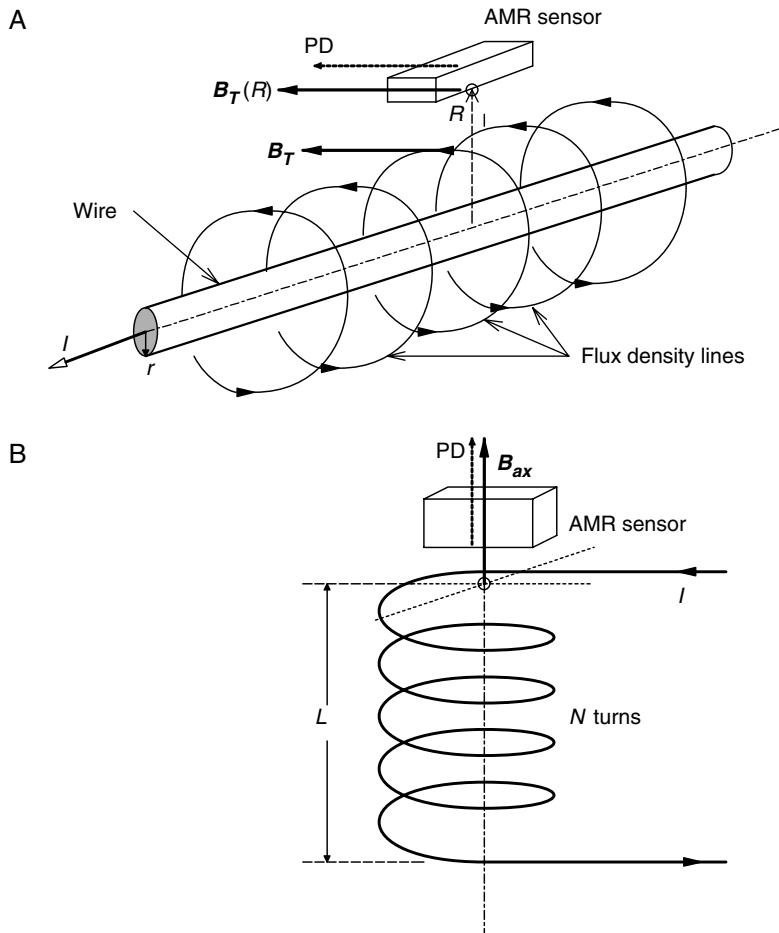


FIGURE 6.16

Galvanically isolated current measurement using AMR magnetosensors. A. The sensor responds to the solenoidal \mathbf{B} field surrounding a single, current carrying conductor. The tangential B field is $B = \mu_0 I / (2\pi R)$. The current I can be dc or powerline frequency ac. B. A solenoidal coil is used to increase sensitivity; it can be shown that at the end of the solenoid, $B = \mu_0 NI / (2L)$ (where μ_0 is the permeability of empty space = $4\pi \times 10^{-7}$ Ns 2 /C 2 , L is the solenoid's length in meters, N is the number of turns on the solenoid, R is the radial distance from the single conductor to the sensor, and $R \gg r$.

solenoid coil is:

$$B_{ax} = \mu_0 NI/(2L) \quad (6.26)$$

where L is the solenoid's length, and N is the number of turns in the coil (Figure 6.16). Curiously, the coil diameter drops out of the equation for B_{ax} . Figure 6.16 illustrates these two geometries for AMR current sensors.

AMR sensors also find application in counting paper currency that has magnetic ink or a magnetic stripe embedded in the paper. A biasing magnet's field is disturbed by the bill as it moves past the magnet. This perturbation is sensed by the AMR sensor. NVE (1998) reports that this technique can easily detect counterfeit bills and can be used to sort moving bills by denomination. This method is far less expensive than optical character reading (OCR) techniques.

Yet another application of AMR (or for that matter, GMR) sensors is a prototype isolation amplifier (IsoA) design by the author. The heart of this circuit is shown in Figure 6.17. An AMR sensor is connected as a Wheatstone bridge. The bridge can be powered with either dc or ac. If ac is used, a phase sensitive rectifier must be placed between DA₂ and the integrator, also, the excitation frequency, ω_c , must be much higher than the highest frequency in $(v_1 - v_2)$. The dc bridge excitation is shown. The voltage difference, under isolated measurement, is converted to a signal current i_s by a VCCS with transconductance, G_m . The current i_s through the left coil creates a field B_s , which lies along the AMR sensor's PD. B_s causes identical changes, ΔR , in the barber pole AMRs, which in turn unbalance the bridge. The bridge output, $(v_3 - v_4)$, is amplified by DA₂ and serves as the input, v_5 , to the integrator. The integrator output, v_f , causes a feedback current, i_f , to flow which forces $B_f = -B_s$. Due to the integrator, $v_5 = 0$ when $i_f = i_s$, $B_f = -B_s$ and the bridge is at null. The output voltage of the op-amp transresistor is:

$$v_o = i_f R_f = G_M v_f R_f \quad (6.27)$$

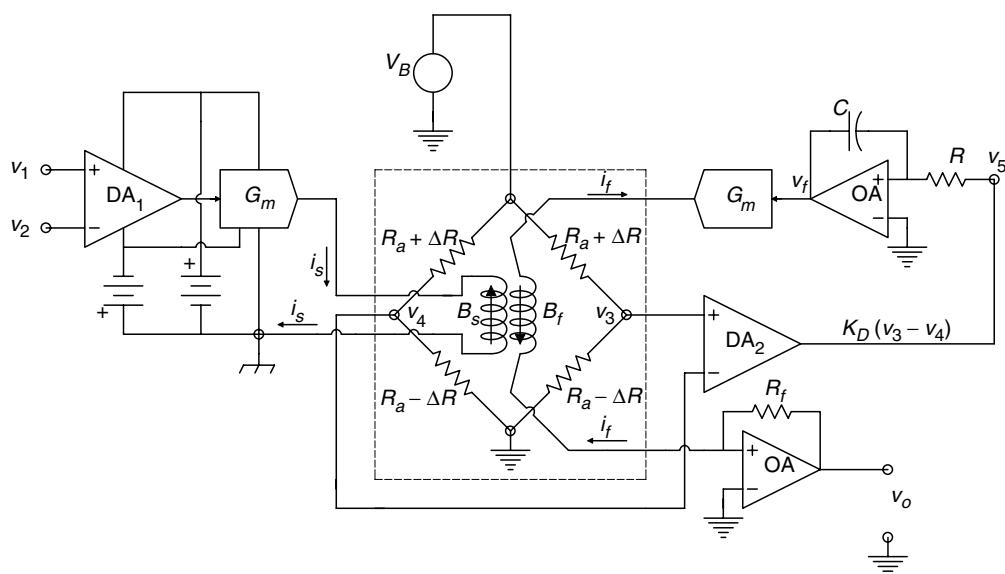


FIGURE 6.17

A prototype isolation amplifier designed by the author using an AMR bridge in a servo-nulling mode.

Thus, the isolated input voltage is:

$$i_f = v_o / R_f = i_s = (v_1 - v_2) K_{D1} G_m \quad (6.28)$$

↓

$$(v_1 - v_2) = \frac{v_o}{R_f K_{D1} G_m} \quad (6.29)$$

It is also possible to show that:

$$(v_1 - v_2) = v_f / K_{D1} \quad (6.30)$$

Galvanic isolation of the AMR bridge, IsoA, depends on the insulation between the chip containing the AMR bridge and the signal input coil. Both capacitance between the coils and the chip, and the dielectric breakdown of this insulation, are important for effective galvanic isolation.

6.4 Voltage Generating Sensors

A wide variety of physical systems generate an EMF in response to some input quantity. In most cases, the magnitude of the induced EMF is linearly proportional to the input quantity. Often, the induced voltage is very small and special types of signal conditioning must be used.

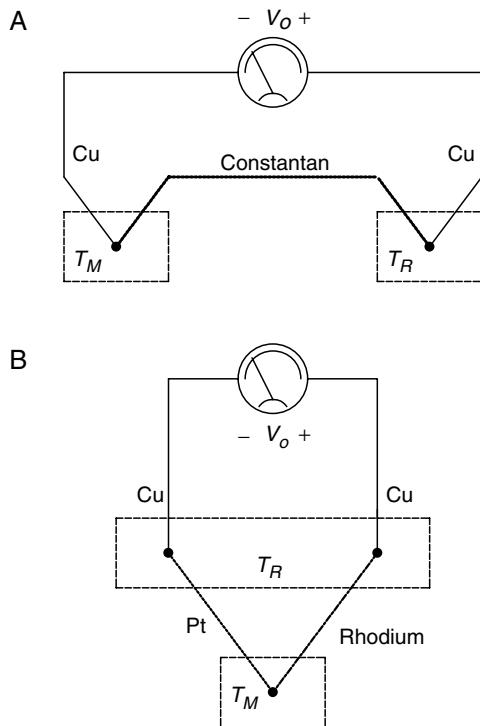
6.4.1 Thermocouples and Thermopiles

Thermocouples are traditionally used to measure temperatures, often at extremes. Thermopiles are close arrays of thermocouples, arranged in series to have great sensitivity and resolution. They are generally used to measure the energy of electromagnetic radiation at optical frequencies, such as from lasers. The total absorption of the radiation causes a minute temperature rise, which is measured by the series thermocouples.

A thermocouple consists of a junction, often spot welded, between two dissimilar metal wires. A basic thermocouple system must contain two couples (Figure 6.18A). However, thermocouple systems may contain three or more couples. In a two couple system, there is a net EMF in the loop, causing current to flow if the two couples are at different temperatures. In a simple, two couple system, one couple junction is maintained at a reference temperature, typically 0°C, by melting ice in water. Often the thermocouple system (measuring and reference junctions), consisting of two different metals must be joined to a third metal, such as copper, to connect to a millivoltmeter. The joining of the two thermocouple metals to copper produces one additional junction, as shown in Figure 6.18B. The addition of the third metal to the system will have no effect on the performance of the basic two-metal thermocouple system, as long as the two junctions with the third metal are at the same (reference) temperature. This property is called the *Law of Intermediate Metals* [Beckwith and Buck, 1961].

Table 6.3 gives some common thermocouple materials and the temperature ranges over which they are effective. The net EMF in a thermocouple system is not a linear function of the temperature difference between the reference and the measuring junction. In general, the net EMF can be expressed as a power series:

$$E_o = A(\Delta T) + [B(\Delta T)^2 / 2] + [C(\Delta T)^3 / 3] \quad (6.31)$$

**FIGURE 6.18**

A. A simple thermocouple circuit in which there are two couples (one for reference, one for measurement), and two metals. B. A thermocouple system in which there are three metals. The two, dissimilar junctions to copper at the reference temperature behave as the single reference junction in Figure 6.18A.

TABLE 6.3
Properties of common thermocouples

Material	ST vs Pt	Material	ST vs Pt
Bismuth	-72	Silver	6.5
Constantan	-35	Copper	6.5
Nickel	-15	Gold	6.5
Potassium	-9	Tungsten	7.5
Sodium	-2	Cadmium	7.5
Platinum	0	Iron	18.5
Mercury	0.6	Nichrome	25
Carbon	3	Antimony	47
Aluminum	3.5	Germanium	300
Lead	4	Silicon	440
Tantalum	4.5	Tellurium	500
Rhodium	6	Selenium	900

where ΔT is the temperature difference above the reference junction temperature, usually made 0°C . The thermoelectric sensitivity of a thermocouple pair is defined as $S_T = dE_o/dT = A + B(\Delta T) + C(\Delta T)^2$ and is usually expressed in $\mu\text{V}/^\circ\text{C}$. Lion (1959) gives a table of thermoelectric sensitivities of various metals *vs* platinum at temperatures near 0°C . These are presented in Table 6.4. Lion points out that the sensitivities of thermocouples made from any pair of metals can be found by subtracting the tabular values of S_T for platinum. For example, S_T for copper *vs* constantan is $[6.5 - (-35.5)] = 41.5 \mu\text{V}/^\circ\text{C}$. Of special note is the extremely high thermoelectric sensitivities of bulk semiconductor materials such as Ge and Si.

TABLE 6.4

Thermoelectric sensitivities of thermocouples made with materials listed vs Platinum, in $\mu\text{V}/^\circ\text{C}$, with reference junction at 0°C

Materials	Useful Temp Range, $^\circ\text{C}$	Max Temp for Short Periods	Sensitivity in $\mu\text{V}/^\circ\text{C}$ at $^\circ\text{C}$
Copper/Constantan	-300 to 350	600	15 @ -200 60 @ $+350$
Iron/Constantan	-200 to $+800$	1000	45 @ 0° 57 @ $+750$
Pt/Pt90Rh10	0 to 1450	1700	0 @ -138 5 @ 0 12 @ $+1500$
Iron/Copnic	-200 to $+860$	1000	60 @ 0
Chromel P/Alumel	-200 to 1200	350	40-55 between 250 and 1000°C
W95Re5/W26Re7	0 to 2316		

Thermocouple EMFs were traditionally measured using accurate potentiometers (calibrated with standard cells). Modern thermocouple measurement makes use of high input impedance, solid state, electronic microvoltmeters. Electronically regulated, solid state temperature references for thermocouple systems are also used, replacing melting ice and water in a thermos flask. Wahl Instruments, Inc. offers electronically regulated, thermal calibration sources and an ice point reference chamber.

A *vacuum thermocouple* is used in conjunction with a sensitive dc microammeter to measure the true RMS value of the current in its heater resistance element. A thermocouple junction is bonded to the heater wire and thus generates an EMF proportional to the temperature of the heater resistor. A reference junction is at room (ambient) temperature, so the net EMF of the thermocouple is proportional to ΔT , the temperature of the heater above ambient temperature. The temperature rise of the heater is given by:

$$\Delta T = \overline{i^2} R_H \Theta_o \quad (6.32)$$

where $\overline{i^2}$ is the mean squared heater current, R_H is the heater resistance (assumed constant) and Θ_o is the thermal resistance of the heater *in vacuo*. Its units are degrees Celsius/watt. The dc microammeter current is then given by Ohm's law as:

$$\begin{aligned} I_M &= E_o / (R_M + R_T) = A \Delta T / (R_M + RT) \\ &= \overline{K_T} i_H^2 R_H / (R_M + R_T) \end{aligned} \quad (6.33)$$

where R_M is the microammeter resistance, R_T is the thermocouple resistance and A is the thermocouple EMF constant at room temperature. Typical parameters for a Western Electric model 20D vacuum thermocouple are: $R_H = 35 \Omega$, $R_T = 12 \Omega$, $K_T \cong 2.9 \text{ V/W}$.

Note that vacuum thermocouples are used for true RMS ammeters or voltmeters by giving the associated dc microammeter a square root scale. Meter deflection is proportional to the power dissipated in R_H , hence i_H^2 . Thermocouple true RMS meters work at dc and at frequencies from the tens of Hz up to the tens of MHz. Due to their true MS deflection, they are widely used for noise voltage measurements.

6.4.2 Photovoltaic Cells

Photovoltaic cells are used to measure light intensity, as well as to generate dc electric power, when used as solar cells. Photovoltaic cells include barrier layer cells, and *pn* junction photodiodes. We first examine the photovoltaic effect in *pn* junction diodes. When light of energy $h\nu$ is absorbed by an open circuited *pn* junction device, the built-in electric field at the junction that results from the doping acts to concentrate the hole electron pairs generated by photon interaction with valence band electrons in such a manner that the holes are swept toward the *p*-side and the electrons toward the *n*-side of the junction. The net effect is the production of an open circuit EMF, given by the relation [Nanavati, 1975]:

$$E_{oc} = V_T \ln\{[qA(L_p + L_n)G(\lambda)/I_{rs}] + 1\} \quad (6.34)$$

where $v_T = nkT/q = 0.026$ V at 300 K, q is the magnitude of the electron charge in coulombs, A is the cross-sectional area of the junction, n is a constant ranging 1–2, L_n is the electron diffusion length, L_p is the hole diffusion length and $G(\lambda)$ is the hole electron pair generation rate per unit volume, which is also a function of the light wavelength and intensity. This relation describes a logarithmic increase of the open circuit photovoltaic voltage with light intensity, other factors being constant. Maximum E_{oc} in bright sunlight is typically 400–500 mV. Figure 6.19A illustrates the current voltage curves for a photodiode or solar cell both in the dark and when illuminated. Figure 6.19B shows typical V_{OC} and I_{SC} curves vs illumination for a photovoltaic device.

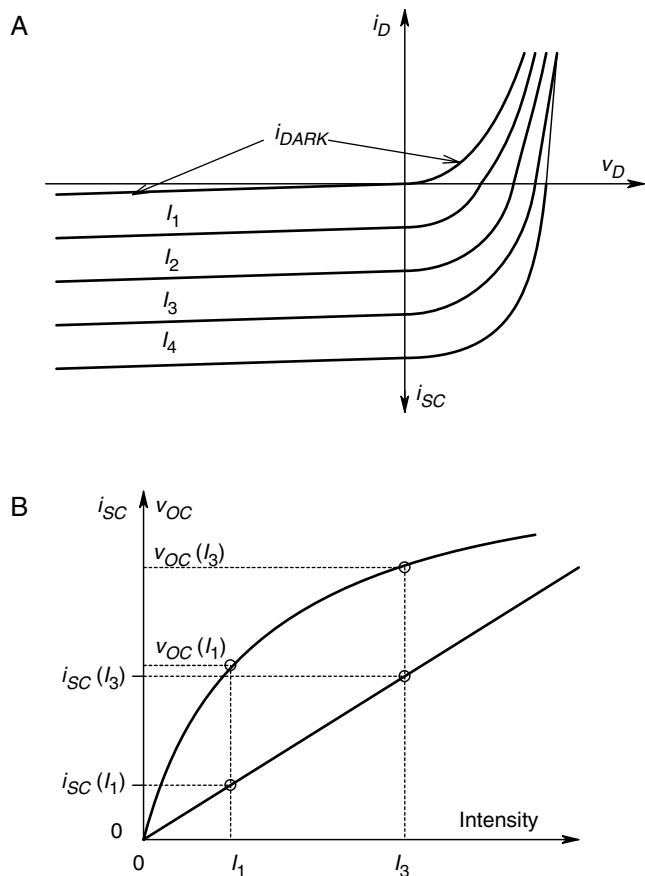


FIGURE 6.19

A. Current voltage curves for a photovoltaic (solar) cell or *pin* photodiode. This class of device is operated in the fourth quadrant. B. Plots of the open circuit voltage and short circuit current of a photovoltaic device. Note that I_{SC} is linear with illumination and V_{OC} varies logarithmically.

When illuminated, it can be shown [Nanavati, 1975] that the current curve is given by:

$$i_D = I_{rs}[\exp(v_D/v_T) - 1] - qA(L_p + L_n)G(\lambda) \quad (6.35)$$

(The parameters in equation 6.35 are defined in the previous equation.)

Photodiodes operated as photovoltaic cells are maximally sensitive to photons with energies near their band gap energy, ϕ_0 . If $h\nu$ is less than ϕ_0 , the photons will not generate hole electron pairs. On the other hand, if the photons have energies far in excess of ϕ_0 , they will be absorbed near the surface of the exposed semiconductor layer, where the recombination rate is high, thus decreasing efficiency. Semiconductor photodiodes made from mixed compounds (PbSe, InAs, InSb) with low band gap energies are used to create FIR detectors. A line of such detectors are offered by EG&G Judson.

Germanium photodiodes have maximum sensitivity at a wavelength of 155 nm. However, their sensitivity ranges from visible wavelengths to 180 nm. An alternate means of using *pn* photodiodes is shown in Figure 6.20. An op-amp is used as a current to voltage converter. Since the op-amp's summing junction is at virtual ground, the photodiode is operated in the short circuited mode, where $v_D = 0$ V. The op-amp output is thus given approximately by:

$$V_o = R_F qA(L_p + L_n) G(\lambda) V \quad (6.36)$$

Under zero bias conditions, the diode's junction capacitance is quite large and hence, its speed of response to changes in the light flux is not as fast as it is for a reverse biased photodiode operated as a photoresistor. EG&G Judson's J16 series of Germanium photodetectors are useable over an optical input intensity range of 2×10^{-12} to 10^{-2} W. Their -3 dB frequencies range widely, 0.07–250 MHz. The faster photodiodes have smaller active junction areas and shunt capacitances, and higher shunt resistances.

A *solar cell* is operated in the fourth quadrant of its current voltage curves (Figure 6.19A). The open circuit photovoltage causes a current to flow in a load and thus, the solar cell can deliver electrical power. A solar cell is made with a considerably expanded *pn* junction area, since it must do work. A typical solar cell of 2 cm^2 area might consist of a thin (0.5 mm), rectangular slab of *n*-type silicon, on top of which is diffused a thin, *p*-type (boron treated) surface layer, through which the light must pass. Such a cell can deliver about 10 mW in bright sunlight with about 12% conversion efficiency. Solar cells are designed to have maximum efficiency in delivering steady state, dc power to a load, which might be charging batteries used for a portable instrumentation module. To obtain maximum efficiency, there is a trade-off between a number of design factors. For example, the thickness of the top (*p*) layer of the solar cell, d , measured from

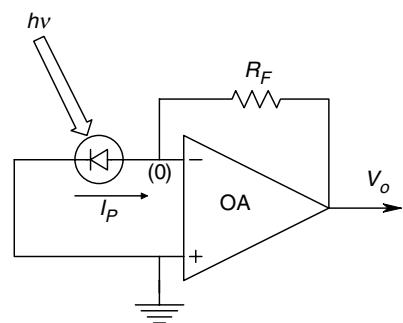


FIGURE 6.20

Op-amp transresistor circuit in which a photodiode is operated in the short circuit mode. Response speed is low due to large zero bias junction capacitance.

its surface to the center of the junction, must be less than L_n to allow electrons generated near the surface to diffuse to the junction before they recombine. Similarly, the thickness of the n region must be such that the electrons generated in the region can diffuse to the junction before they, too, recombine. This requirement implies a match between the hole diffusion length, L_p , the thickness of the n region and the optical penetration depth in the surface p layer. It is also desirable in solar cell design to have a large contact potential in order to realize a large photovoltage. This means that heavy doping must be used. The need for heavy doping must be compromised with the requirement for long carrier lifetimes, which are reduced by heavy doping. Efficient solar cells require a low internal (ohmic) resistance. There is no problem in making metallic electrode contact with the bottom (dark) layer of the solar cell, but low resistance also requires a large electrode contact area with the top layer. A large metallic area on the top semiconductor layer prevents light from reaching the cell, thus decreasing efficiency. The tops of solar cells are generally given an anti-reflection coating, similar to that used on camera lenses, to maximize light absorption.

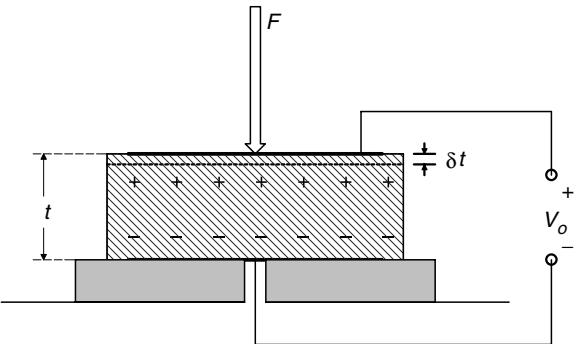
6.4.3 Piezoelectric Transducers

There is a class of piezoelectric materials which can serve, among other applications, as mechanical input transducers, enabling the measurement of pressure, force, displacement and other related physical phenomena. Piezoelectric materials, when mechanically strained in a preferred manner, generate an open circuit EMF. Actually, the mechanical strain causes a unidirectional separation of electric charges resident in the interior of the piezoelectric crystal structure. These displaced charges form an effective net charge on a capacitor formed by the bonding of metallic electrodes on the surface of the piezoelectric crystal. The transducer material itself, independent of piezoelectric activity, is an insulator having very low conductivity and high dielectric constant. Thus, the effectively separated charges produce an open circuit voltage on the capacitor equal to Q/C . In practice, the situation is not that simple. The piezoelectric transducer material is not a perfect dielectric and the charges leak through the exceedingly small volume conductance. Of course, any voltage measuring system attached to the transducer will have finite input impedance and perhaps, a dc bias current, further altering the EMF caused by the stress. In addition, the transducer has mechanical mass, stiffness and damping, giving it mechanical resonance properties which can further complicate the transducer's transfer function when the frequency of the mechanical input approaches the transducer's mechanical resonance frequency. In the discussion below, we assume that the piezoelectric transducers are operated at frequencies well below their mechanical resonant frequencies.

Piezoelectric materials include a number of natural crystals and manmade ceramic materials. Quartz, Rochelle Salt, and ammonium dihydrogen phosphate (ADP) are examples of naturally occurring piezomaterials. Barium titanate, lead zirconate-titanate and lead metaniobate are synthetic piezoceramics. All piezoelectric materials have a *Curie temperature*, above which piezoelectric activity ceases to exist.

In order to understand how piezoelectric input transducers work, we will examine an example of a crystal responding to thickness compression. Figure 6.21 shows a rectangular crystal block, with metallic electrodes vapor deposited on its top and bottom surfaces. The transducer has thickness t and top area A . The capacitance of the transducer is approximately:

$$C_X = \kappa\epsilon_0 A/tF \quad (6.37)$$

**FIGURE 6.21**

A piezoelectric transducer slab responding to thickness compression.

where κ is the dielectric constant of the piezoelectric material. A uniformly applied force, F , will cause the crystal to compress a minute amount, δt . The piezo material is assumed to have a Young's modulus, Y .

$$Y = \text{stress}/\text{strain} = (F/A)/(\delta t/t) \quad (6.38)$$

The net charge displacement for the transducer is given by:

$$Q = d \times F C \quad (6.39)$$

where d is the *charge sensitivity* of the material in $(C/m^2)/(N/m^2)$. Thus we can write the open circuit voltage across the transducer as:

$$E_o = Q/C = \frac{d \times F}{\kappa \epsilon_0 A/t} = g(F/A)t V \quad (6.40)$$

Here g is the *voltage sensitivity* of the transducer in $(V/m)/(N/m^2)$. F/A is the pressure on the transducer's active surface.

The axes along which a natural crystal of piezoelectric material is cut, in order to make a transducer, determine the type of mechanical input to which it will have maximum sensitivity. In addition to simple thickness compression, piezoelectric transducers may also be made to respond to length compression, thickness and face shear, as well as twisting and bending. Synthetic piezo materials may be cast in a variety of shapes not available with natural materials, such as thin walled cylinders, half cylinders, discs, hollow spheres and hemispheres. Table 6.5 summarizes some of the salient properties of certain piezo materials.

For mechanical input frequencies well below piezoelectric transducer mechanical resonance, the equivalent circuit shown in Figure 6.22 may be used to model the electrical behavior of the transducer. Although charge is displaced internally by strain, we must use a current source in the model. Since current is the rate of flow of charge, the equivalent current source is:

$$i_{eq} = dQ/dt = d \times (dF/dt) = d \times \dot{F} \quad (6.41)$$

Thus, from Ohm's law, the open circuit voltage across the transducer is given by the transfer function:

$$\frac{V_o}{F}(s) = \frac{s F d R_L}{s + s C_T/G_L} \quad (6.42)$$

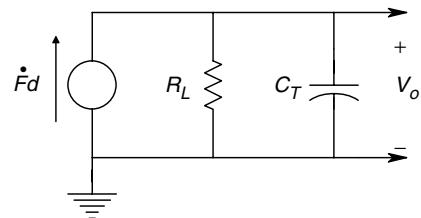
TABLE 6.5

Properties of piezoelectric materials

Material	Orientation	Charge Sensitivity, d in pC/N	Voltage Sensitivity, g	Curie Temp.
Quartz	X cut; length along Y , length longitudinal	2.25×10^{-12}	0.055	550°C
	X cut; thickness longitudinal	-2.04	-0.050	
	Y cut; thickness shear	4.4	0.108	
Rochelle Salt	X cut 45° length longitudinal	435	0.098	55°C
	Y cut 45°; length longitudinal	-78.4	-0.29	
Ammonium dihydrogen phosphate	Z cut 0° face shear	48	0.354	125°C
Barium Titanate	Z cut 45°; length longitudinal	24	0.177	125°C
	Parallel to polarization	86–160	0.016	
	Perpendicular to polarization	-56	0.005	
Lead Zirconate Titanate	Parallel to polarization	190–580	0.02–0.03	170–360°C
Lead Metaniobate	Parallel	80	0.036	>400°C

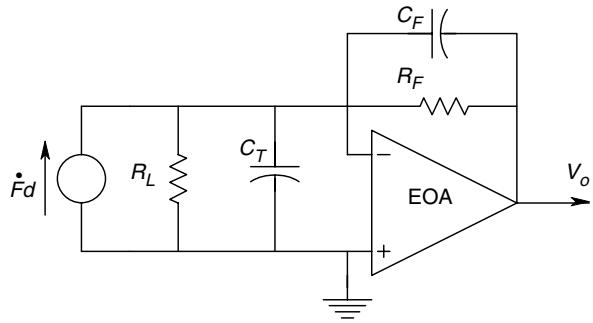
FIGURE 6.22

Electrical equivalent circuit of a piezoelectric transducer responding to applied force. Mechanical frequencies in the applied force are assumed to be well below the transducer's mechanical resonance frequency.



This transfer function is seen to be a simple high-pass filter. Thus, because the charge displaced by the strain caused by force F leaks off the crystal, true dc response of a piezoelectric transducer is impossible. G_L is the total leakage conductance of the piezoelectric crystal plus the input conductance of the meter or oscilloscope used to measure V_o , plus the leakage conductance of the cable used to connect the transducer to the measuring system. C_T is the capacitance of the transducer plus the input capacitance of the measuring device. If the transducer is connected to an oscilloscope with a $10\text{ M}\Omega$ input resistance, we can neglect the other leakage resistances because they are very large ($\approx 10^{13} \Omega$). C_T may be on the order of 1 nF for a piezoceramic transducer. Thus the time constant of the transducer/oscilloscope system is *ca.* 10 ms, and the response to a step of force input will have decayed to zero in about 30 ms.

To improve and control low frequency response of a piezoelectric transducer system, we wish to connect the transducer to an amplifier with an ultra high input impedance. Such amplifiers are called *electrometer amplifiers*, and they generally have input resistances of 10^{13} – $10^{15} \Omega$. Another common means of conditioning the output of a piezoelectric transducer is to use *charge amplifier* configuration, using an electrometer op-amp (charge amplifiers are introduced in Section 2.6.1 of this text). Figure 6.23 shows one such circuit for a charge amplifier. Here, we neglect any series resistance between the transducer and the op-amp's summing junction. If the op-amp is considered to be ideal, it is easy to see that the transducer current, F_d , only flows through the amplifier's

**FIGURE 6.23**

An electrometer op-amp is used as a charge amplifier to condition the output of a piezoelectric force (or pressure) transducer.

feedback conductance. Hence, the transfer function of the transducer and amplifier can be written as:

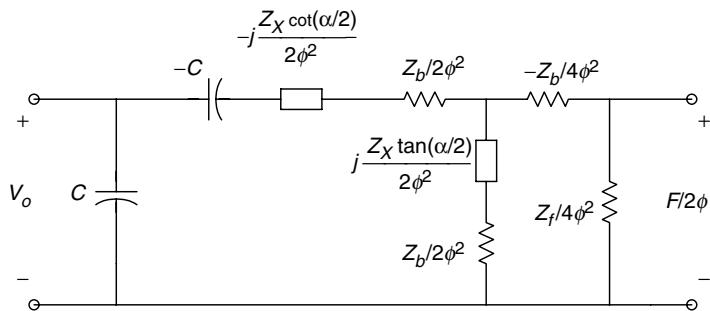
$$\frac{V_o}{F}(s) = \frac{-d R_{FS}}{s + C_F R_F} \quad (6.43)$$

We still have a high-pass system with no dc response, but now, the effect of C_T and G_T are negligible, and the low frequency pole is controlled with the known feedback elements, R_F and C_F . For example, let R_F be $10^{10} \Omega$ and C_F be 5 nF . The time constant is then 50 s and the break frequency is 3.18 mHz . The high frequency gain is then $d/C_F \text{ V/N}$. For a quartz transducer, d is about $2.3 \times 10^{-12} \text{ Cb/N}$, thus the mid-band gain would be $4.6 \times 10^{-4} \text{ V/N}$. The use of piezoceramics with larger ds obviously leads to greater sensitivities when used in the charge amplifier configuration. Due to their greater stability and higher Curie temperature, quartz pressure transducers are used in applications such as the measurement of internal combustion engine combustion chamber pressures.

Other applications of piezoelectric input transducers include measurement of minute, dynamic displacements (microphones, phonograph cartridges to translate the grooves in vinyl phonograph records into sound, measurement of minute pulsations on body surfaces, measurement of surface roughness, measurement of acceleration (with an attached mass) etc.).

Piezoelectric transducers used as sonar receivers, or in ultrasonic applications, such as medical imaging, are generally operated at their mechanical resonant frequencies for maximum sensitivity. At resonance, the electrical equivalent circuit of a transducer changes markedly. Figure 6.24 illustrates the Mason electrical model for a piezoelectric input transducer valid for frequencies near the mechanical resonance frequency, f_o , of the transducer [Fox and Donnelly, 1978]. η is the turns ratio of the ideal transformer. It can be shown that $\eta = A e_{33} t$, where A is the effective surface area of the transducer in m^2 , e_{33} is the piezo electric stress constant in the material in Cb/m^2 and t is the transducer thickness in m .

Here, we assume a transducer responding in the thickness mode. In addition, the characteristic impedance of the piezoelectric material is given by $Z_x = 2\rho A f_o$, where ρ is the density of the piezo material in kg/m^3 and f_o is the mechanical resonant frequency of the transducer in Hz. The parameter $\alpha = \pi f/f_o$. f is the frequency of the applied mechanical force. When the transducer is driven at exactly its mechanical resonant frequency, its electrical equivalent circuit (Thevenin source impedance) reduces to a series $R\text{-}C$ circuit with capacitance C_o and a real part, $4\eta^2/[(2\pi f_o C_o)2(Z_f + Z_b)] \Omega$. Here, C_o is the transducer's clamped capacitance, and Z_f and Z_b are the acoustic source impedances faced by the front and back faces of the transducer, respectively. As can

**FIGURE 6.24**

Electrical equivalent circuit of a piezoelectric transducer near its mechanical resonance frequency.

be seen from the transducer model of Figure 6.24, operation of the transducer even slightly off the resonant frequency leads to a far more complex model for the Thevenin source impedance of the transducer's electrical output. In general, the real part of the Thevenin impedance of the transducer's output is numerically small, typically ranging from the tens of ohms to the hundreds of ohms. Contrast this with the extremely high resistances presented by piezoelectric transducers, operated well below their resonant frequencies.

As a consequence of the low, real source impedance presented by a piezoelectric input transducer operated at its mechanical resonance frequency, a step-up transformer can be used to obtain optimum SNR when measuring threshold signals, as is done in sonar systems. Transformer optimization of output SNR is discussed in detail in Section 3.5 of this text.

6.4.4 Pyroelectric Sensors

Pyroelectric materials (PyMs) are crystalline or polymer substances that generate internal, electrical charge transfer, in response to *internal heating*, caused by the absorption of long wave, infrared radiation (LIR). The charge transfer can be sensed as a current or voltage change, depending on the kind of electronic signal conditioning associated with the PyM. In general, PyM materials are also piezoelectric (i.e. they also respond to applied mechanical stress by internal charge transfer). PyM sensors include the polymers polyvinylidene fluoride (PVDF) and polyvinyl fluoride (PVF); and the crystalline substances lithium tantalate (LiTaO_3), strontium and barium niobate, triglycine sulfate (TGS), Rochelle salt, KDP (KH_2PO_4), ADP ($\text{NH}_4\text{H}_2\text{PO}_4$), barium titanate (BaTiO_3), LZT, etc.

PyM sensors are fabricated by taking a thin rectangle or disc of the PyM material and coating both sides with a very thin layer of vapor deposited metal such as gold, silver or aluminum. Electrical contact is made with silver epoxied wires or pressure contacts. The side of the sensor that is to receive radiation is often given an extra, thin, rough, heat absorbing coating such as platinum black. This coating maximizes the ratio of absorptivity to reflectivity for the PyM sensor. As IR radiation sensors, PyM sensors are used in the LIR wavelength range of $2\text{--}25 \mu\text{m}$ ($5000 \text{ cm}^{-1}\text{--}400 \text{ cm}^{-1}$ wavenumber).

Note that PyM sensors respond only to *changes* in temperature. In the thermal steady state, there is no net internal charge transfer and no voltage across their electrodes. Thus, if a constant input radiation power, P_i , is applied, the sensor's temperature rises to an equilibrium value, $T'_a > T_a$, where radiation and conduction heat losses equal the

input power. In general, we can write a heat balance differential equation for the PyM sensor as:

$$P_i(t) = C_T \frac{d(T'_a - T_a)}{dt} + \frac{(T'_a - T_a)}{\Theta} = C_T \frac{d\Delta T}{dt} + \frac{\Delta T}{\Theta} \quad (6.44)$$

where C_T is the PyM material's *heat capacity* in J/K, Θ is its *thermal resistance* in K/W, and $\Delta T = T'_a - T_a$. T_a is the starting steady state, ambient temperature of the PyM and T'_a is the temperature it rises (or falls) to as a result of absorbing (radiating) P_i over some time T_s . Θ depends on the PyM material used, its configuration and even how it is mounted (Θ can be reduced by direct thermal conduction (heatsinking) and by air convection). C_T is given by:

$$C_T = cAh \quad (6.45)$$

where c is the PyM's specific heat in J/(cm³ K), A is the absorbing surface area in cm² and h is the PyM thickness in cm.

If $\Delta T/T$ is small, we can assume that P_i remains constant over T_s . The differential equation, 6.44, can be Laplace transformed and written as a transfer function:

$$\frac{\Delta T}{P_i}(s) = \frac{\Theta}{s\Theta C_T + 1} \quad (6.46)$$

Now the short circuit current from the irradiated PyM is given by:

$$i_p(t) = K_p A \Delta \dot{T} \quad (6.47)$$

Laplace transforming:

$$I_p(s) = K_p A s \Delta T(s) \quad (6.48)$$

When equation 6.48 for $I_p(s)$ is substituted into equation 6.46, we can finally write the transfer function:

$$\frac{I_p}{P_i}(s) = \frac{s K_p A \Theta}{s\Theta C_T + 1} \quad (6.49)$$

K_p is the PyM's *pyroelectric constant* in C/(m² K). Table 6.6 gives the important physical constants of certain common PyMs (note units).

TABLE 6.6

Physical properties of certain PyMs

Pyroelectric Material	Pyroelectric Coefficient K_p in $\mu\text{C}/(\text{m}^2 \text{ K})$	Dielectric constant $\kappa = \epsilon/\epsilon_0$	Thermal Resistance $\Theta \text{ K/W}$	Specific Heat c in $\text{J}/(\text{cm}^3 \text{ K})$
Triglycine sulfate (TGS)	350	3.5	2.5×10^{-3}	2.5
Lithium Tantalate (LiTaO ₃)	200	46	2.38×10^{-4}	3.19
Barium titanate (BaTiO ₃)	400	500	3.33×10^{-4}	2.34
PVDF film	40	12	7.69×10^{-3}	2.4

Source: Data from Fraden (1993b) and Pállas-Areny and Webster (1991).

Assume the PyM sensor is at thermal equilibrium at temperature T_a . A radiation blocking shutter is opened, permitting a step of radiation from a warm object, at constant temperature T_b , to reach the sensor. The *short circuit current* is given by equation 6.22:

$$I_p(s) = \frac{P_{io}}{s} \frac{sK_p A \Theta}{(s\Theta C_T + 1)} = P_{io} \frac{K_p A / C_T}{(s + 1/(\Theta C_T))} \quad (6.50)$$

In the time domain, this is simply an exponential decay waveform:

$$i_p(t) = (P_{io} K_p A / C_T) \exp[-t/(\Theta C_T)] \quad (6.51)$$

ΘC_T is the sensor's *thermal time constant*, which is material and dimensionally dependent. P_{io} is assumed constant and is given by:

$$P_{io} = A \Delta H = K_s (T_b^4 - T_a^4) \quad (6.52)$$

$$K_s = A \sigma / (1/e_a + 1/e_b - 1) \quad (6.53)$$

Now examine the circuit in Figure 6.25. The PyM sensor is connected to an electrometer op-amp, as a transimpedance. Note that the equivalent circuit for the PyM can be configured as an *ideal current source*, $i_p(t)$, in parallel with the sensor's very low electrical leakage conductance, G_p , and its electrical self-capacitance, C_p . The $i_p(t)$ is given by equation 6.51 above for a step input of LIR power. Let us first neglect the op-amp's feedback capacitor, C_F . The output voltage of the op-amp is given by (note the direction of i_p):

$$v_o(t) = R_F i_p(t) = R_F (P_{io} K_p A / C_T) \exp[-t/(\Theta C_T)] \geq 0 \quad (6.54)$$

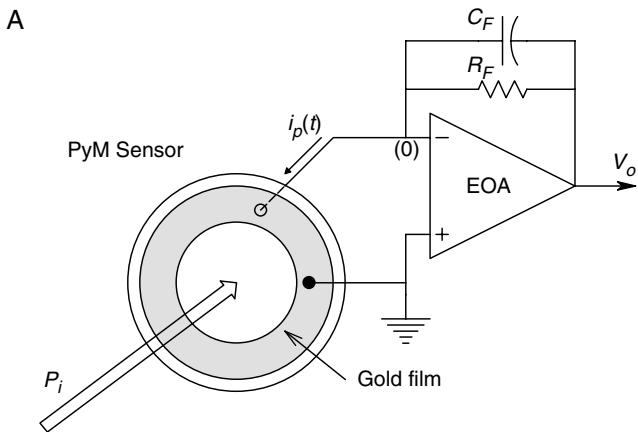


FIGURE 6.25

A. An electrometer charge amplifier is used to condition the output of a pyroelectric IR sensor. B. Equivalent circuit of the PyM sensor away from its mechanical resonance.

The peak $v_o(t)$ is $V_{opk} = R_F(P_{io}K_pA/C_T)V$, hence we can calculate the temperature, T_b , of the warm object from V_{opk} and a knowledge of the system's constants.

$$T_b = \left[T_a^4 + \frac{V_{opk}C_T}{R_FK_pAK_s} \right]^{1/4} = [T_a^4 + V_{opk}K_{sys}]^{1/4} \quad (6.55a)$$

The problem with this approach is that, while V_{opk} is measured fairly accurately, the system's constants and T_a are generally not that accurately known. Hence a known blackbody temperature source at T_{cal} can be used to find the lumped constant K_{sys} .

$$K_{sys} = (T_{cal}^4 - T_a^4)/V_{ocal} \quad (6.55b)$$

Such a blackbody reference source can be built into the transient radiation thermometer, or be an external blackbody source, such as the Mikron™ Model 310 blackbody radiation calibration source made by Kernco Instruments, Inc.

Another approach to self-calibration is to keep the PyM at a known T_a , then expose it to T_{cal} and measure the peak V_{ocal} , then expose it to T_b and measure V_{opk} . The computer subtracts V_{opk} from V_{ocal} to form ΔV_o :

$$\Delta V_o = (1/K_{sys})[T_{cal}^4 - T_b^4] \quad (6.56)$$

Solving for T_b :

$$T_b = [T_{cal}^4 - K_{sys}\Delta V_o]^{1/4} \quad (6.57)$$

Thus, we see that the calculated T_b of the body relies on measurement of ΔV_o , K_{sys} and the known T_{cal} .

To counteract high frequency measurement noise, a low-pass filter can be added to the current-to-voltage conversion op-amp in Figure 6.25, by placing capacitor C_F in the feedback path. With C_F in place, the system response to a step input of LIR irradiative power, P_i , is:

$$V_o(s) = P_{io} \frac{K_pA/(C_F C_T)}{[s + 1/(C_F R_F)] [s + 1/(C_T \Theta)]} \quad (6.58)$$

The inverse Laplace transform of equation 6.58 can be shown to be:

$$v_o(t) = \frac{P_{io}K_pA R_F \Theta}{(C_F R_F - C_T \Theta)} \{\exp[-t/(C_F R_F)] - \exp[-t/(C_T \Theta)]\} V \quad (6.59)$$

This is a positive waveform that rises with an initial time constant $C_T \Theta$, and falls more slowly with time constant, $R_F C_F$. Its peak is proportional to P_{io} . The measurement noise comes from:

1. Unwanted mechanical vibration of the PyM sensor which is piezoelectric
2. Electronic noise from the op-amp
3. Thermal (Johnson) noise from the PyM's Norton conductance and from R_F

Applications of PyM sensors include intruder sensors and burglar alarms, automatic patio light switches, sensors in various LIR spectrometers used in analytical chemistry, etc. They are also used in the line of Thermoscan™ fever thermometers that quickly and non-invasively measure the temperature of the eardrum [Northrop, 2002]. Commercial PyM sensors come in a variety of packages.

Some manufacturers (e.g. Molelectron) package PyM sensors alone, and with an op-amps or a JFET for 1st stage signal conditioning. Manufacturers such as GEC-Marconi list PyM sensor *Responsivity* (n V/W (input)), *noise voltage* ($\mu\text{VRMS}/\sqrt{\text{Hz}}$), *noise equivalent power* (NEP) ($\text{nW}/\sqrt{\text{Hz}}$), *noise equivalent irradiance* (NIR) ($10^{-4}\text{W}/\sqrt{\text{Hz}}\text{m}^2$) *specific detectivity* (D^*) ($10^6\text{m}/\sqrt{\text{Hz}}/\text{W}$). For example, at 1.0 Hz, the GEC-Marconi PLT533 lithium tantalate PyM sensor has responsivity = 1800 V/W, noise = $1.1\ \mu\text{V}/\sqrt{\text{Hz}}$, NEP = $0.6\ \text{nW}/\sqrt{\text{Hz}}$, NEI = $0.7 \times 10^{-4}\ \text{W}/(\text{m}^2\sqrt{\text{Hz}})$, $D^* = 5 \times 10^6\ \text{m}/\sqrt{\text{Hz}}/\text{W}$.

6.4.5 Sensors Whose Voltage Output is Proportional to $d\Phi/dt$

It is well known from basic physics that an EMF will be induced in a coil of N turns surrounding a magnetic flux, Φ , when the magnetic flux changes in time. This effect may be simply stated mathematically as:

$$E_o = Nd\Phi/dt = L di/dt \quad (6.60)$$

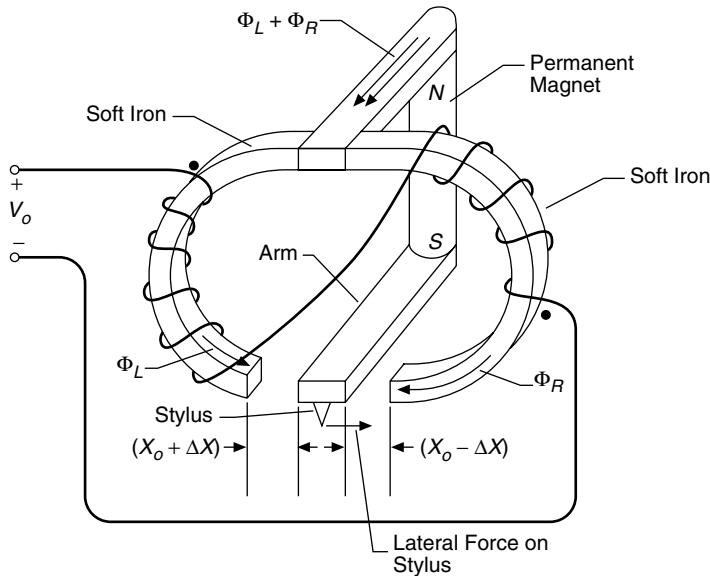
Many input transducers make use of this principle, including variable reluctance phonograph pickups, dynamic microphones, accelerometers, tachometers, etc. The source of the magnetic flux may be either a permanent magnet, or a dc or ac excited electromagnet (solenoid). The $d\Phi/dt$ may arise from a number of means, including modulation of the reluctance of the magnetic path by changing the size of an airgap in the flux path. Moving a coil perpendicular to the flux density also gives a $d\Phi/dt$, as does moving a permanent magnet in relation to a fixed coil.

6.4.5.1 The Variable Reluctance Phonograph Pickup

Magnetic Reluctance, \mathfrak{R} , is defined as the ratio of magnetomotive force to the magnetic flux, Φ , existing in a closed, magnetic 'circuit'. It may be thought of as being analogous to electrical resistance in an electrical circuit, where the MMF is analogous to voltage and the flux, Φ , is analogous to current. A simple variable reluctance phonograph pickup transducer, used to play antique 33 1/3 rpm vinyl phonograph records, is shown in Figure 6.26. The source of MMF is a permanent magnet (PM). Flux from the PM is split and passed through a right and left branch magnetic path so that the total flux in the magnet is the sum of left and right path fluxes. Thus:

$$\Phi = \Phi_R + \Phi_L \quad (6.61)$$

The reluctances of the paths are modulated by the lengths of the air gaps, caused to vary by the small lateral displacement of the moveable arm in the air gap, δx . The arm is displaced by a lateral force, F , acting on the stylus, S , protruding downward from the arm. In playing an old fashioned, non-stereo phonograph record, the stylus tip follows the helical groove of the record. Audio information is recorded on the record as lateral, side-to-side displacements of the groove. Due to its high inertia, the tone arm and cartridge do not respond to the small, audio frequency lateral forces the groove places on the stylus. Instead, the moveable (stylus) arm of the cartridge follows the lateral

**FIGURE 6.26**

Schematic of a monophonic, variable reluctance phonograph pickup.

modulations of the groove. The reluctance of the right and left hand magnetic paths may be written:

$$\mathfrak{R}_R = \mathfrak{R}_o - \delta x / \mu_o A \quad (6.62)$$

$$\mathfrak{R}_L = \mathfrak{R}_o + \delta x / \mu_o A \quad (6.63)$$

The total magnetic flux passing through the magnet is then:

$$\Phi = \text{MMF} / (\mathfrak{R}_R + \mathfrak{R}_L) = \text{MMF} / 2\mathfrak{R}_o \quad (\text{constant}) \quad (6.64)$$

An N -turn coil is wound with the dot polarity shown on the left arm of the magnetic circuit. This coil is connected in series with an N -turn coil on the right arm having the opposite dot polarity. The magnetic flux in the left arm is given by:

$$\Phi_L = \frac{\text{MMF}/\mathfrak{R}_o}{(1 + \delta x / \mu_o A \mathfrak{R}_o)} \cong \frac{\text{MMF}}{\mathfrak{R}_o} (1 - \delta x / \mu_o A \mathfrak{R}_o) \quad (6.65)$$

Similarly, the flux in the right arm can be approximated by:

$$\Phi_R \cong \frac{\text{MMF}}{\mathfrak{R}_o} (1 + \delta x / \mu_o A \mathfrak{R}_o) \quad (6.66)$$

The net EMF induced in the two series connected coils is found by taking the time derivative of the flux in each arm, multiplying by the number of turns in each coil and adding the EMFs algebraically:

$$E_o = \frac{2N(\text{MMF})}{\mathfrak{R}_o^2 \mu_o A} [d(\delta x) / dt] \quad (6.67)$$

Thus we see that the output EMF is proportional to the stylus velocity. The variable reluctance transducer does not respond to constant (dc) displacements.

6.4.5.2 Electrodynamic Accelerometer

Figure 6.27 illustrates a linear accelerometer made from a moving, cylindrical mass (which is also a permanent magnet), which moves inside a coil fixed to the accelerometer's case. The mass is constrained by a linear spring with stiffness k_s and is surrounded by oil, which gives viscous damping to the motion of the mass. The case of the accelerometer is attached to some structure whose linear acceleration we wish to measure. The output voltage of the accelerometer is proportional to the relative velocity between the case (and structure) and the moving mass. In mathematical terms:

$$V_o = K_V(\dot{x}_c - \dot{x}_m) \quad (6.68)$$

The motion of the mass can be described by simple Newtonian mechanics:

$$M\ddot{x}_m + B(\dot{x}_m - \dot{x}_c) + k_s(x_m - x_c) = 0 \quad (6.69)$$

The acceleration, \ddot{x}_c , is considered to be the system input. If we separate terms of the force equation and take their Laplace transforms, we may write the transfer function as:

$$\frac{X_m(s)}{X_c(s)} = \frac{sB + k_s}{s^2M + sB + k_s} = \frac{sB/ks + 1}{s^2M/ks + sB/ks + 1} \quad (6.70)$$

Now the relation for $\dot{X}_m(s) = sX_m(s)$ is substituted into the equation for the output voltage. After some algebra this yields:

$$V_o(s) = \dot{x}_c(s) \frac{sM/ks}{s^2M/ks + sB/ks + 1} \quad (6.71)$$

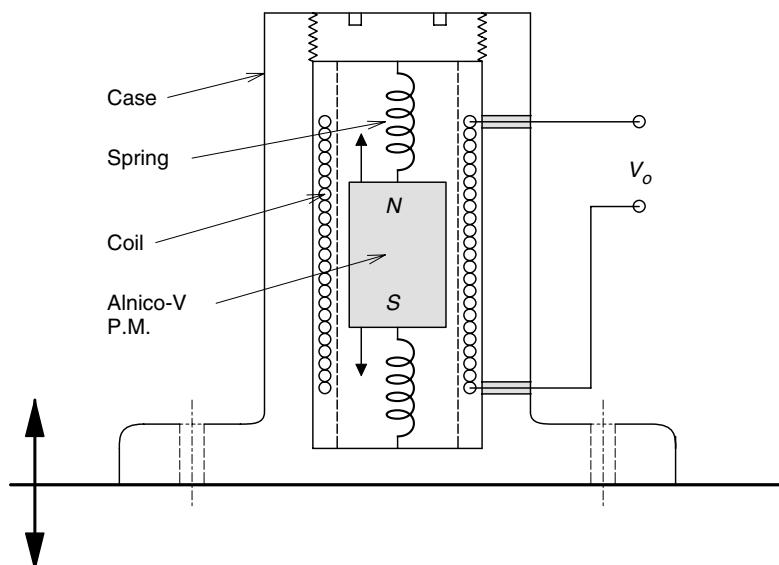
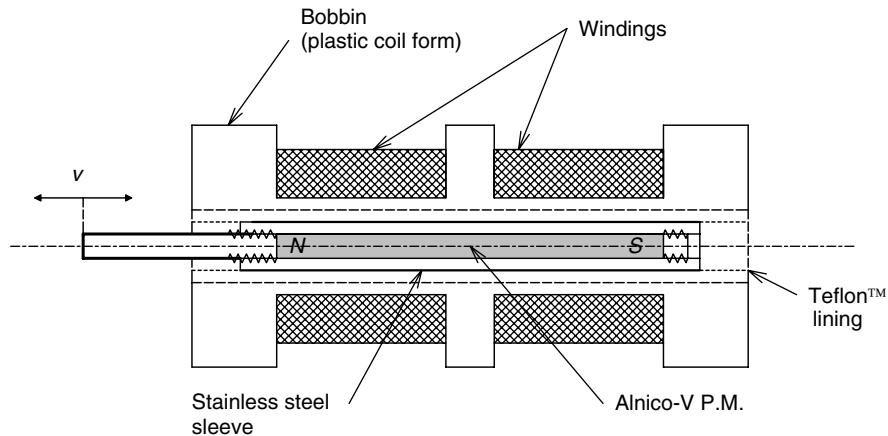


FIGURE 6.27

Cross-section of an electrodynamic accelerometer.

**FIGURE 6.28**

Cross-section of a linear velocity sensor (LVS).

Thus we see that this transducer's response at low mechanical input frequencies is proportional to *jerk* (i.e. the rate of change of acceleration of the structure to which the case is attached). There is no output in response to constant acceleration, such as due to the Earth's gravitational field.

6.4.5.3 Linear Velocity Sensors

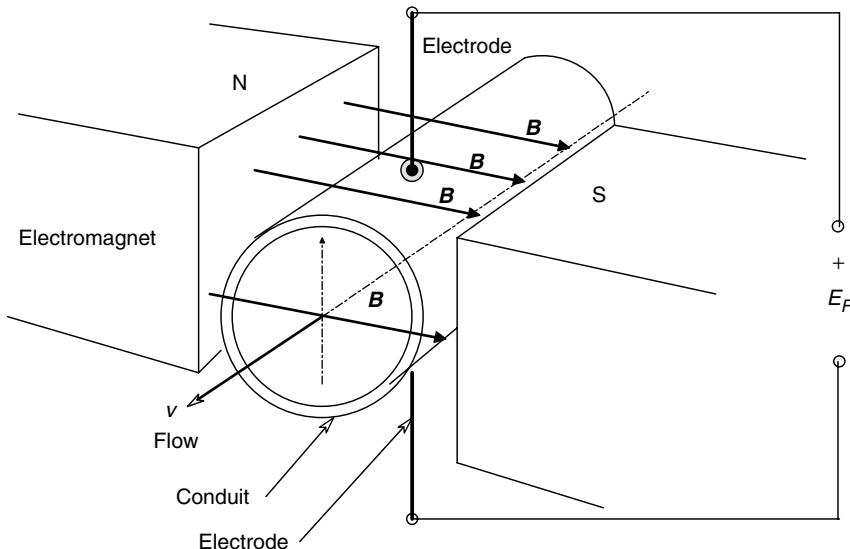
A linear velocity sensor (LVS), as the name suggests, produces an output voltage which is directly proportional to the rate of linear displacement of an object attached to its core. Figure 6.28 illustrates a section through an LVS. Note that the two coils are connected in series opposing configuration to improve linearity over the full displacement range. Schaevitz Engineering offers LVSs covering ranges of motion, spanning 12.5–500 mm. Sensitivities are in the range 4.8–26 mV/mm/s in various models using Alnico V magnetic cores. Linearity is $\pm 1\%$ of output over the nominal linear range. Robinson-Halpern also offers LVSs with working ranges of 12.5–225 mm; typical sensitivities are about 20 mV/mm/sec. LVS case lengths are roughly twice the linear range of motion.

6.4.6 Sensors Whose Output EMF Depends on the Interaction of a Magnetic Field with Moving Charges

The fluid version of this class of generating sensor is generally used with a constant magnetic (B) field, to measure the average velocity of a fluid flowing in an insulating pipe or conduit, including blood vessels. The solid state version of this class of sensor is the well-known Hall-effect device, used to measure magnetic fields, or to compute average power as a wattmeter.

6.4.6.1 Faraday Effect Flowmeters

Figure 6.29 illustrates the basic geometry of a Faraday flowmeter. Two electrodes on opposite sides of the circular cross-section pipe make ohmic contact with the conductive fluid in the pipe moving with average velocity, \bar{v} . A uniform magnetic field of B Tesla is perpendicular to the v vector and the L vector between the centers of the

**FIGURE 6.29**

Schematic diagram of a Faraday-effect liquid velocimeter. Note flow is orthogonal to \mathbf{B} .

electrodes. Faraday's law of induction predicts that the EMF between the electrodes is given by:

$$E_F = \int_0^d (v \times B) \cdot dL \quad (6.72)$$

For orthogonal vector components, equation 6.72 reduces to $E_F = BL\bar{v}$. For effective measurement, the conductivity of the fluid should be greater than $10^{-5} \Omega^{-1} \text{ cm}^{-1}$. If a circular pipe of diameter d is used for the conduit, and d is also the separation of the electrodes, then the average volume flow, Q , of the conducting fluid may be shown to be given by [Lion, 1959]:

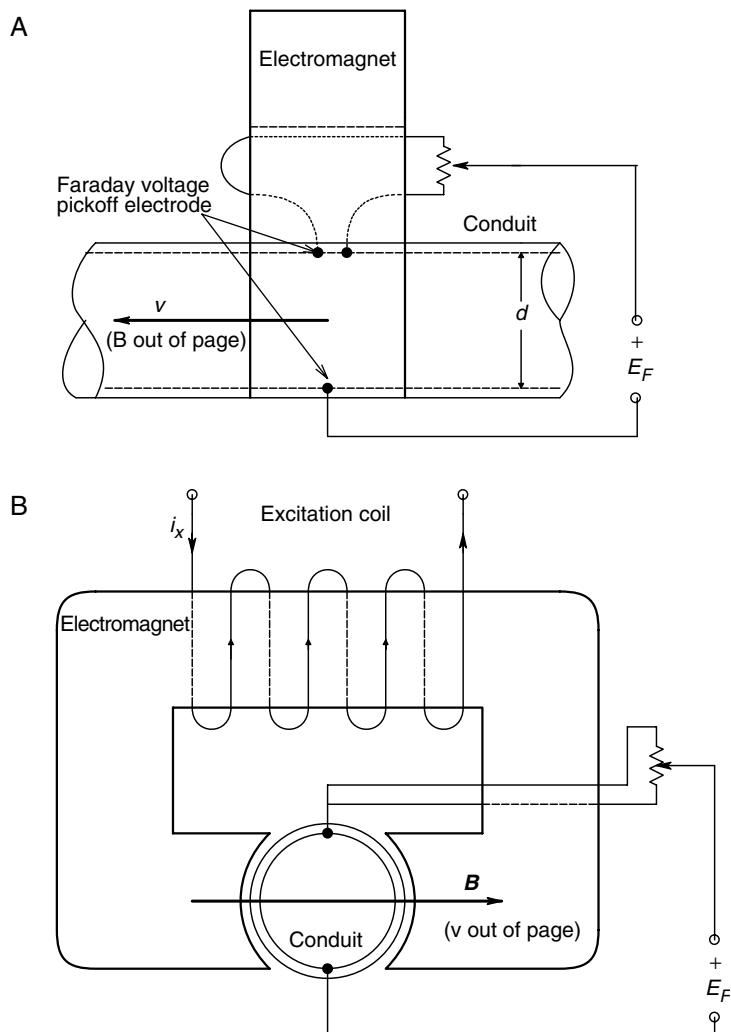
$$Q = E_F d\pi / (4 \times 10^{-8} B) \text{ ml/s} \quad (6.73)$$

Here, d is measured in cm, B in gauss and E_F in volts. Often the area of the conduit is not precisely known, hence flow estimates using the Faraday method are not accurate, since E_F is proportional to the average velocity of the fluid.

In the Faraday flowmeter, the magnetic field is generally made to be a low frequency sinusoid in order that very small EFs can be better resolved. An ac E_o can generally be amplified above the amplifier's $1/f$ noise frequency band.

The Faraday flowmeter has been used to measure blood velocity and estimate blood flow. In one version, the electrodes make contact with the surface of the blood vessel, which is conductive itself. Due to geometrical errors in setting up the orthogonality between B , \bar{v} and L *in vivo*, a net time varying magnetic flux passes through the one turn coil formed by the electrodes used to measure the Faraday EMF.

Thus, the total output EMF seen across the electrodes will be the sum of the Faraday EMF and $d\Phi/dt = AdB/dt$ V, which cuts the effective loop of area A formed by the E_F measurement circuit.

**FIGURE 6.30**

A. Longitudinal section (side view) of a Faraday velocimeter system using three sensing electrodes. The upper two electrodes permit cancellation of the unwanted $d\Phi/dt$ (quadrature) term in the output voltage. B. Axial (end-on) view of the three-electrode Faraday system.

Webster (1978) described several methods of canceling the unwanted $d\Phi/dt$ term. One method makes use of a cancellation circuit formed by making one of the electrodes a twin electrode, as shown schematically in Figure 6.30A. Each circuit will have a transformer error voltage, one of the form $E_{T1} = k_1 d\Phi/dt$, the other, $E_{T2} = -k_2 d\Phi/dt$. Thus, for some potentiometer wiper position, the two transformer EMFs will cancel, leaving the pure Faraday voltage. Another means of eliminating the effect of the transformer voltage is based on the fact that the transformer voltage is 90° out of phase (in quadrature) with the Faraday voltage (assuming sinusoidal excitation). Thus if we use a phase sensitive, sampling demodulator, we can sample the peak values of the Faraday voltage sinewave at the times the $d\Phi/dt$ sinewave is going through zero. The sampled Faraday voltage sinewave is held and low-passed filtered to obtain a dc signal proportional to \bar{v} . A third means of suppressing the unwanted transformer voltage makes use of a feedback system, shown in Figure 6.31. This system makes use of two phase sensitive

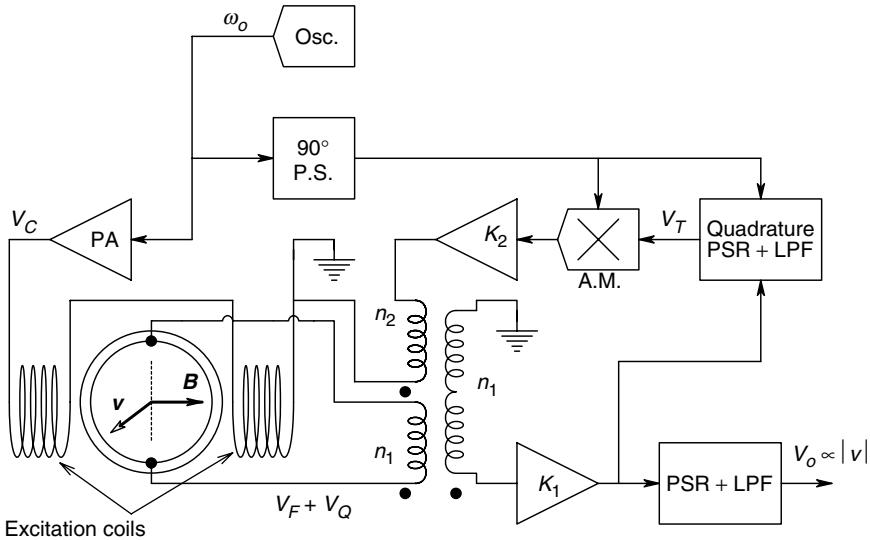


FIGURE 6.31

System to automatically null the unwanted quadrature output voltage from a Faraday liquid velocity sensor.

rectifiers and low-pass filters. One detects the desired Faraday voltage and the other, using a quadrature reference, detects the unwanted transformer voltage. A dc signal, V_T , proportional to the transformer voltage, acts through an analog multiplier to vary the level of a quadrature signal, V_{QF} , which is fed back through an auxiliary transformer winding, n_2 , to cancel the quadrature component, V_Q , which accompanies V_F . The quadrature detector loop acts to reduce the amount of transformer voltage contaminating the desired V_F signal, as shown in equation 6.74.

$$V_{OD} = V_Q / [1 + (n_1/n_2)K_1 K_{DQ} K_2(10)] \quad (6.74)$$

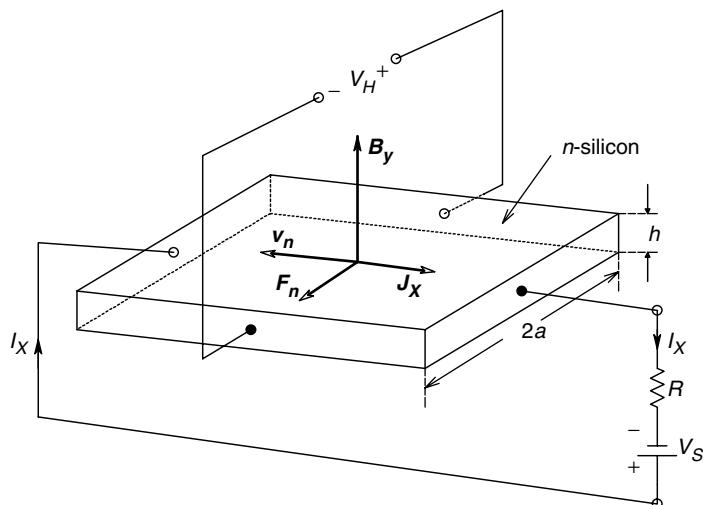
If an integrator is put between the quadrature PSD/LPF and the analog multiplier, it is easy to show that there will be complete cancellation of V_{DQ} . Webster (1978) describes several other configurations of electromagnetic blood velocimeters using the Faraday streaming effect.

6.4.6.2 Hall Effect Sensors

Hall effect sensors, unlike Faraday effect devices, use no moving, conductive fluid. Instead, the majority charge carriers (electrons or holes) have some average drift velocity as they traverse a bar of doped semiconductor. Figure 6.32 illustrates the geometry of a typical Hall sensor. Note that, as in the case of the Faraday transducer, the magnetic field \mathbf{B} , the drift velocity \mathbf{v} and the voltage measuring axis are ideally orthogonal. Each hole or electron with velocity \mathbf{v} in the semiconductor will, in general, be subject to a Lorentz force given by the vector equation:

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) = \{q|\mathbf{v}| |\mathbf{B}| \sin(\theta)\}\mathbf{u} \mathbf{N} \quad (6.75)$$

This force follows a right-hand screw rule in which the direction of the force is found by rotating \mathbf{v} into \mathbf{B} and observing the unit vector direction, \mathbf{u} , that a screw with right hand (normal) threads would advance. The q is the charge of an electron (or hole).

**FIGURE 6.32**

Schematic of a Hall sensor. An n -doped Si slab is used. Note orthogonality between J_x , B_y and F_n .

If electrons are the carriers, then we use a minus sign with q . As we see from Figure 6.32, the Lorentz force causes the moving electrons to crowd toward the front surface of the n semiconductor bar. An equilibrium density of electrons is reached when the average Lorentz force is balanced by the electrostatic force on the electron due to the electric field set up by their crowding. In other words:

$$-qE_z = -qv_nB_y \quad \text{or} \quad E_z = v_nB_y \quad (6.76)$$

It is well known that in an n semiconductor, the average drift velocity of electrons is given by:

$$v_n = -J_x/qn \quad (6.77)$$

where J_x is the current density in the semiconductor bar in the $+x$ direction and n is the electron doping density. This relation for v_n can be substituted in equation 6.77 above, giving:

$$E_z = -J_xB_y/qn \text{ V/m} \quad (6.78)$$

Now, if we consider the geometry of the semiconductor bar, $J_x = I/2ah$ and the Hall voltage developed across the two metallized faces of the bar is just $V_H = E_z 2aV$. Substitution of these relations into equation 6.78 yields the well-known expression for the Hall voltage.

$$V_H = -IB_y/(qn)h = R_H IB_y/h \quad (6.79)$$

where h is the height of the semiconductor bar (in the y direction) and R_H is the Hall coefficient—the $R_H = -1/qn$ for n -doped semiconductor and $R_H = +1/qn$ for p semicon. Since the Lorentz force has the opposite sign for p semiconductors, given that all other vectors have the same direction, the Hall voltage for p material is $V_H = +IB_y/(qn)h$.

Hall sensors have many uses, most of which can be subdivided into either analog or switching applications. As analog sensors, Hall devices can be used to produce an output

proportional to $|B| \cos(\theta)$, where θ is the angle of the \mathbf{B} vector with the transducer's y axis (axis of maximum sensitivity). Some commercial IC analog Hall sensors have built-in dc amplifiers to condition the V_H signal. Analog Hall sensors listed by Sprague Electric Co. have sensitivities of about 1.4 mV/gauss with high frequency bandwidths of 20–25 kHz (on B). Other Hall sensors listed by F. W. Bell do not contain built-in amplifiers and claim sensitivities in the range 10–55 mV/kG. Bell also offers a Model BH-850 high sensitivity Hall sensor that uses a 9 inch, high permeability bar to concentrate the magnetic flux. This sensor claims a sensitivity of 18 mV/G, which is adequate sensitivity to measure the Earth's magnetic field and thus has applications as a solid state compass.

One obvious application of analog Hall sensors is the measurement of magnetic fields. They can also be used to measure other physical or electrical parameters which can be made proportional to the magnetic fields (ac or dc). A less obvious application of an analog Hall sensor is as an audio frequency wattmeter. The current through a reactive load is passed through a solenoid coil to produce a B_y proportional to I . The voltage across the load is conditioned and used to generate a current through the Hall sensor proportional to V . The phase between the I and V sinusoids is θ . It is easy to show that the dc component of the Hall voltage is proportional to the average power in the load; that is, it is given by:

$$\overline{V_H} = kP_{AVE} = kVI \cos(\theta) \quad (6.80)$$

Here, k is a proportionality constant, V is the RMS voltage across the load, I is the RMS current through the load and $\cos(\theta)$ is the power factor.

Switching Hall sensors have many applications in counting and proximity (of a permanent magnet) detection. Their output is generally a TTL logic signal. An IC dc amplifier, Schmitt trigger and open collector logic output are mounted in the same package with the Hall chip. Hall switches have uses in automobile electronic ignitions, brushless dc motors (rotor position sensing), tachometers, keyboards, thermostats, pressure and temperature alarms, burglar alarms, etc.

6.5 Sensors Based on Variable Magnetic Coupling

This class of sensors operates with ac excitation. Their outputs vary according to the degree of magnetic coupling between the excitation coil(s) and the output winding(s). Most sensors which operate on the principle of variable magnetic coupling between coils are mechanotransducers (i.e. they can be used to measure or generate linear or rotational displacements). Since many other physical phenomena can be converted to small linear rotations or displacements, this class of transducer is also the basis for certain pressure, temperature and acceleration sensors. Lion (1959) Section 1.22 gives a good summary of this class of transducer, which includes, among others, the linear variable differential transformer (LVDT) and the synchro transformer.

6.5.1 The LVDT

The linear variable differential transformer is one of the most widely used mechanical input sensors in modern instrumentation practice. LVDTs are used to measure linear

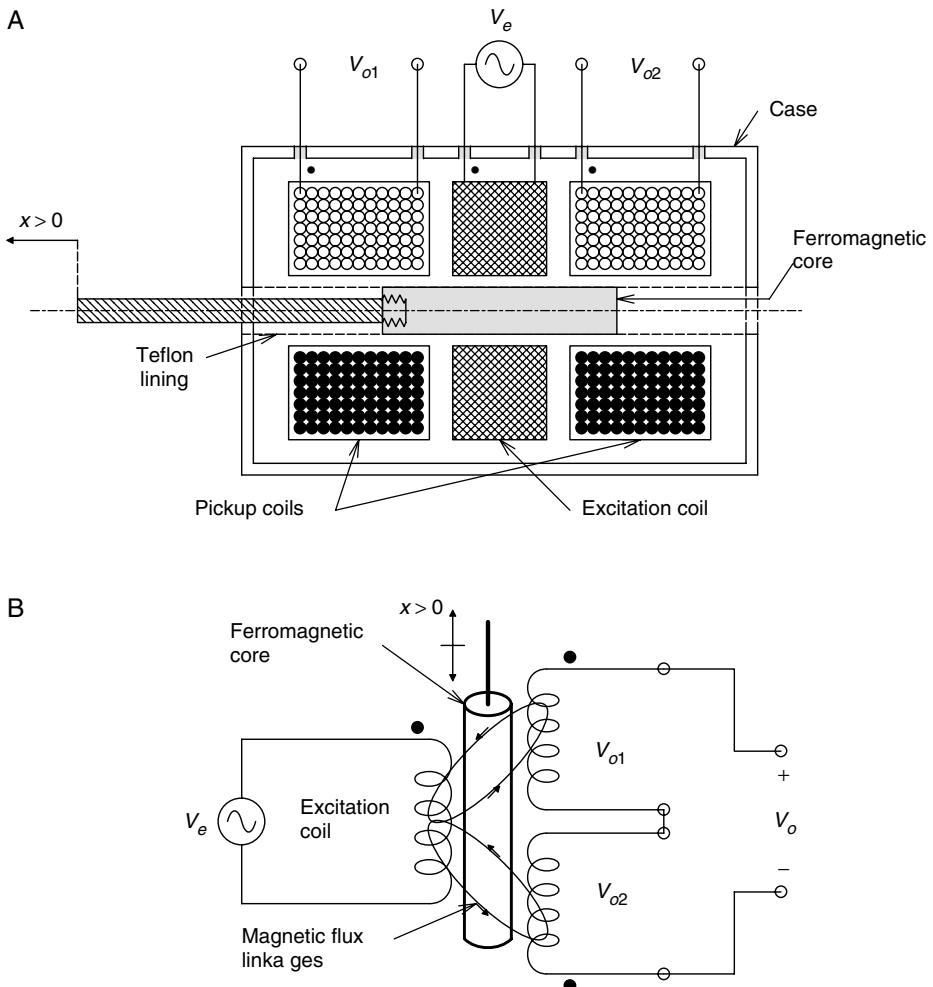
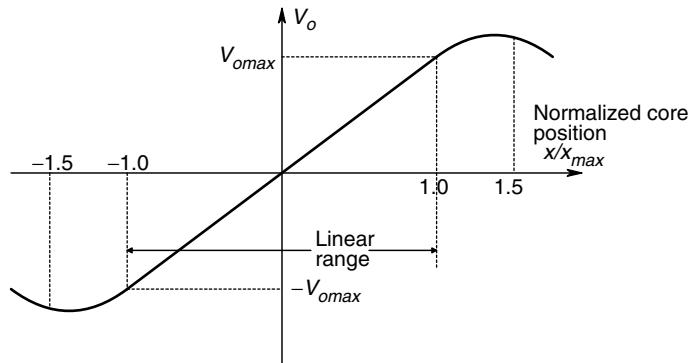


FIGURE 6.33
A. Longitudinal cross-section through a typical LVDT. B. LVDT schematic. Note dot notation on the output coils.

mechanical displacement or position in control systems and in precision manufacturing gauging, and can be used indirectly to measure force, pressure, acceleration, etc., or any quantity that can cause a linear displacement. Figure 6.33A illustrates a section through an LVDT showing the central excitation coil, two, symmetrically located pick-off coils and the hollow, cylindrical core tube through which slides the high permeability magnetic core. The position of the core in the center tube, relative to the excitation coil and the two pick-off coils, changes the mutual inductance between the excitation coil and the pick-off coils, varying the induced voltages in the two pick-off coils. As one voltage drops, the other rises, etc.

A schematic representation of an LVDT is shown in Figure 6.33B. Note that the two series pick-off coils have opposite dot polarity, so that, when the moveable magnetic core is at magnetic center ($x=0$), the induced EMFs in each coil sum to zero at the output. The ac excitation of an LVDT can range in frequency of about 50 Hz–20 kHz. Peak output response for a given displacement, $\delta x > 0$ is maximum at some frequency which may be in the range 400 Hz–2 kHz. Some LVDTs contain built-in oscillators and phase sensitive rectifiers, so one has only to supply a dc power source (e.g. 24 V), and observe a dc output which has a voltage/core position relationship similar to that shown in

**FIGURE 6.34**

Typical output voltage vs linear core displacement for an LVDT.

Figure 6.34. Several electronic IC manufacturers offer oscillator/phase sensitive rectifier chips (Signetics' NE5520, Analog Devices' AD630) to facilitate the use of LVDTs which do not contain this signal processing circuit internally.

LVDTs are not zero force sensors. If their cores experience a lateral force, there will be increased friction with the inside of the core tube. In addition to friction forces, the solenoidal magnetic field from the excitation coil acts to pull the core toward the center ($x=0$) position. The solenoid or axial force is proportional to the square of the excitation current. Typical peak axial force is less than one gram, which is negligible in most position sensing applications.

Specifications for LVDTs vary widely according to application. Those used for gauging applications (e.g. measuring the contour of a cam) have infinite resolution, linearity of $\pm 0.2\%$ of the full range and repeatability as small as $0.1016 \mu\text{m}$, [in SI]. Gauge LVDTs generally have some sort of special tip to contact the work, such as a hardened tungsten carbide ball, as well as a helical spring to maintain a constant static contact force with the work. Other versions of LVDT gauges use air pressure to obtain a variable gauge contact force which can be set for delicate (soft) materials. The Schaeftz PCA-230 series precision, pneumatic, LVDT gauges have gauging ranges from $\pm 0.005''$ to $\pm 0.100''$. Other models have ranges up to $\pm 1.000''$ and some have built-in oscillator/demodulator ICs for dc operation.

6.5.2 Synchros and Resolvers

We first consider synchros, also called *selsyns* or *autosyns*. These are angular position sensors that work on the principle of variable mutual inductance. They generally have multiple windings. There are three main classes of synchros—transmitters, repeaters and control transformers. Figure 6.35A is a schematic diagram of a synchro. Rotor excitation is generally at 60 or 400 Hz. The three stator windings are arranged at 120° spacings, so the induced stator voltages may be written:

$$v_1(t) = V_{pk} \sin(2\pi ft) \sin(\theta) \quad (6.81a)$$

$$v_2(t) = V_{pk} \sin(2\pi ft) \sin(180^\circ + \theta) \quad (6.81b)$$

$$v_3(t) = V_{pk} \sin(2\pi ft) \sin(240^\circ + \theta) \quad (6.81c)$$

Synchros were first widely used in the 1940s in military systems as angular position sensors and components in servomechanisms used to train guns or gun turrets, sonar

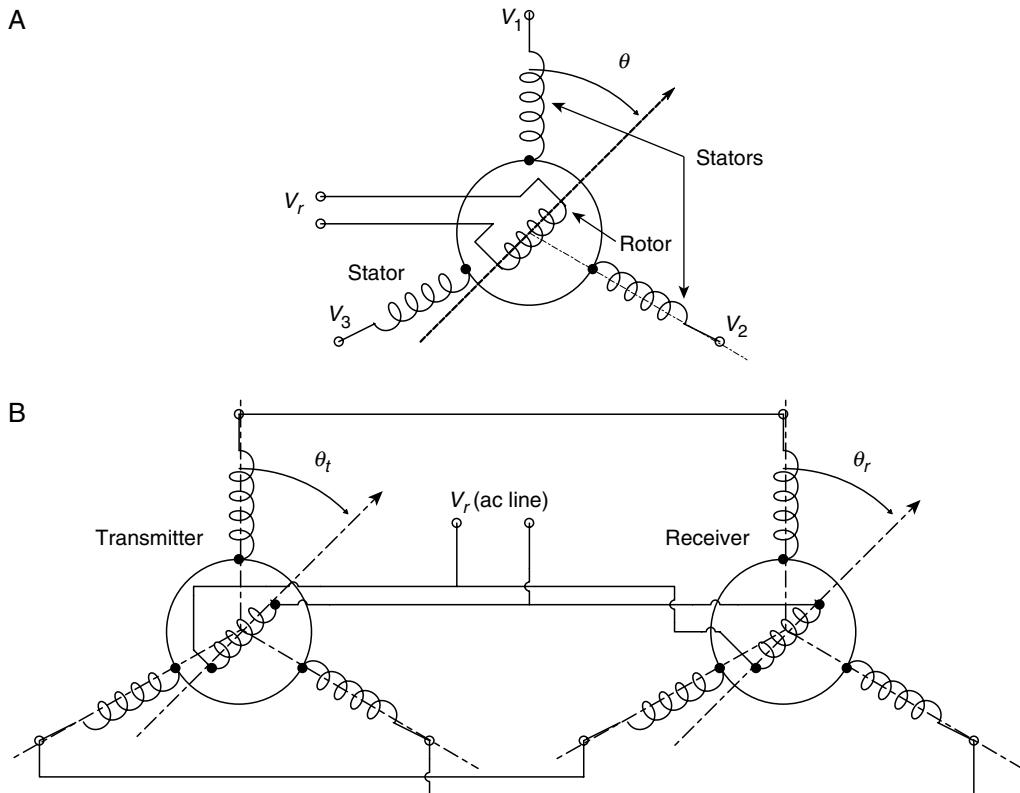


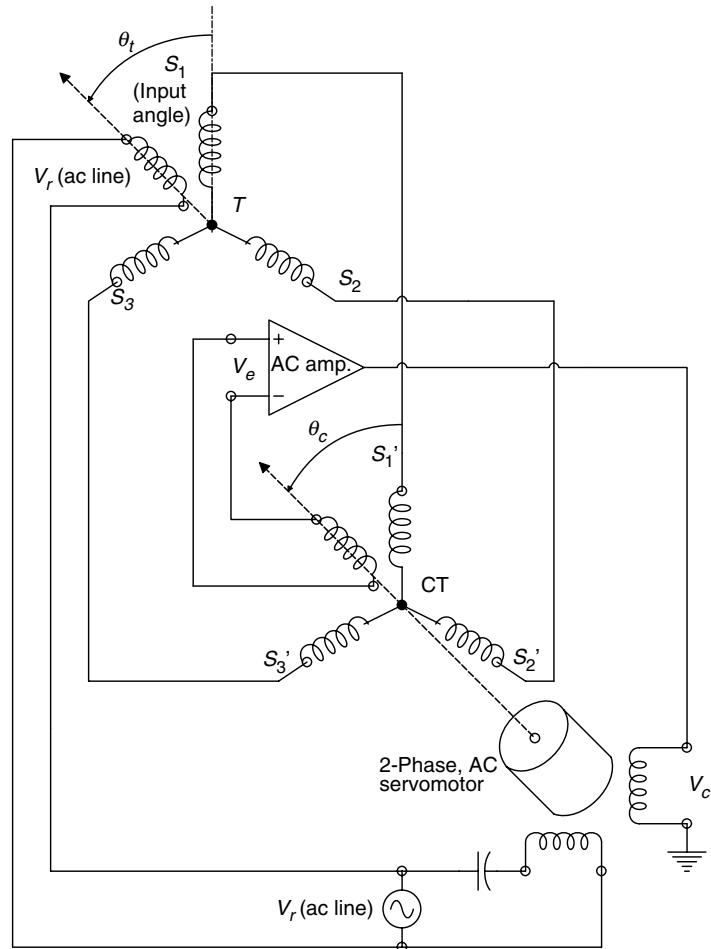
FIGURE 6.35

A. Schematic diagram of a synchro transmitter. B. Electrical connection between a synchro transmitter (T) and repeater (R) enables the repeater rotor to follow the angle of the transmitter's rotor.

arrays, or radar antennas. Synchros come in all sizes, ranging from the massive Navy 7G, developed during WW II, to modern, miniature synchros measuring 1" in diameter by 1.8'0 in length. Synchro accuracies range from ± 0.1 to about ± 1.0 .

The simplest application of two synchros is signaling a shaft rotation to a remote location. In this application, two synchros are used, a transmitter and a repeater, shown in Figure 6.35B. The rotors of the transmitter and repeater are connected in parallel to the ac line and the corresponding stator lines are connected. The rotor of the transmitter is at some angle θ_t . The rotor of the repeater is initially at some angle θ_r , and is free to rotate. The repeater rotor experiences a torque, causing it to rotate so that $\theta_r = \theta_t$. The repeater rotor experiences zero torque when $\theta_r = \theta_t$. The repeater rotor generally has a light external inertial load, such as a pointer and some viscous damping to prevent overshoots and oscillations when following sudden changes in the transmitter's θ_t .

Another major application of synchros is error signal generation in carrier operated, angular position control systems (Figure 6.36). A synchro transmitter and a synchro control transformer are used to generate a carrier frequency signal whose amplitude is proportional to $\sin(\theta_t - \theta_c)$, where θ_t is the rotor angle of the transmitter and θ_c is the rotor angle of the control transformer. If we consider θ_t to be the input, reference, or command signal and θ_c to be the output of a servomechanism designed to follow θ_t , then the output of the synchro transformer may be considered to be the servo system error for $\theta_e = (\theta_t - \theta_c) < 15^\circ$, where $\sin(x) \approx x$ in radians. The synchro transformer output can be shown to be a double-side-band, suppressed carrier modulated signal. When demodulated by a phase sensitive rectifier, the output will be a voltage proportional

**FIGURE 6.36**

Schematic of a simple position control system (servo) which uses synchros. A synchro transmitter (T) and a synchro control transformer (CT) are used.

to $\sin(\theta_t - \theta_c)$. Obviously, the synchro transmitter and transformer can be located at some distance from one another, which is a distinct advantage.

Modern practice, however, is to replace the electromechanical synchro transformer with a commercially available solid state circuit system into which the input angle is read as a binary number and the output is the same as would be seen from the rotor of a conventional synchro transformer. In spite of such solid state conveniences, synchro systems are difficult to work with. Modern position control system design generally makes use of optical coding disks to sense angular position and directly codes it into digital form for transmission to the control computer.

Resolvers are another rotational coordinate sensor. As shown schematically in Figure 6.37, resolvers differ from synchros in that they have two orthogonal stator windings and two orthogonal rotor windings. The stator windings are generally energized so that:

$$v_{s1}(t) = V_{s1} \sin(\omega t) \quad (6.82a)$$

$$v_{s2}(t) = V_{s2} \sin(\omega t) \quad (6.82b)$$

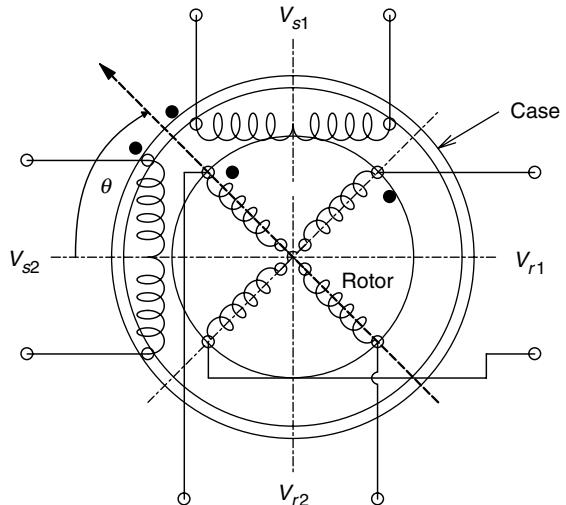


FIGURE 6.37
Schematic of a resolver.

Noting the dot convention shown in Figure 6.37, the induced output (rotor) voltages are:

$$v_{r1}(t) = v_{s1}(t) \cos(\theta) - v_{s2}(t) \sin(\theta) \quad (6.83a)$$

$$v_{r2}(t) = -v_{s1}(t) \sin(\theta) + v_{s2}(t) \cos(\theta) \quad (6.83b)$$

where θ is the rotor angle with respect to the reference shown. If just S_1 is excited, the resolver may be used to do polar to rectangular coordinate transformation in an electromechanical system. In modern instrumentation systems, such transformations are better done by software in a computer.

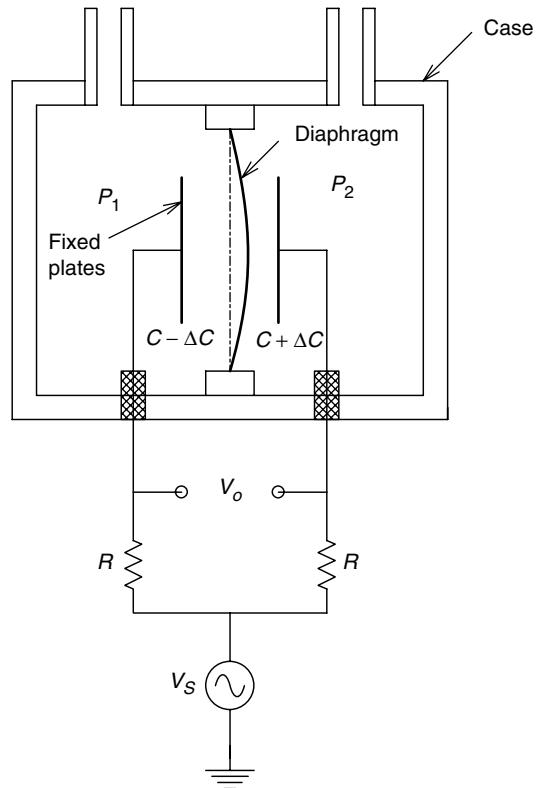
6.6 Variable Capacitance Sensors

Many mechanical input sensors, as well as some sensors which measure humidity or temperature, operate by the transduced quantity causing a change in capacitance, which in turn, is converted to an analog output voltage by an ac bridge circuit or other electronic system. Capacitive sensors have been designed to measure force by means of displacement of one or two capacitive electrodes. They can also be used to measure acceleration, thickness, depth of a dielectric liquid and pressure. In its simplest form, the mechanical input quantity causes a change in the separation between two (or more) capacitor plates, as described below.

The capacitance between two parallel, conducting plates, each of area A , separated by a dielectric of thickness d is given by (neglecting fringing effects):

$$C = \frac{\kappa \epsilon_0 A}{d} \quad (6.84)$$

where κ is the dielectric constant of the material separating the plates and ϵ_0 is the permittivity of free space ($8.85 \times 10^{-12} \text{ F/m}$). Small variations in C , ΔC , which may be measured by several means, may occur if the plate separation is caused to change

**FIGURE 6.38**

Section through a differential pressure sensor which uses the small change in capacitance caused by the deflection of its diaphragm to produce an ac output signal proportional to $(P_1 - P_2)$. A De Sauty bridge can be used.

by some physical quantity, or if the dielectric constant, κ , changes due to pressure, temperature, mechanical strain, humidity, etc. The overall dielectric constant may also be modulated by sliding a substance with a dielectric constant greater than one between the capacitor plates so that some fraction, α , of A has a dielectric constant of $\kappa = \kappa_1$ and the remaining area between the plates, $(1 - \alpha)A$, has a dielectric constant $\kappa = 1$. The area of the capacitor plates may be effectively changed by sliding one plate over the other at constant d . An example of this latter means of effecting a ΔC is found in parallel plate, rotary tuning capacitors used in radios.

The means of measurement of capacitance changes include various capacitance bridges or the use of the capacitor in a tuned $L-C$ circuit determining the frequency of an oscillator—the oscillator's frequency is measured to obtain a signal proportional to $1/\sqrt{LC}$.

One of the most sensitive means of producing an output voltage proportional to ΔC is through the use of a De Sauty bridge, shown in Figure 5.7 and discussed in Section 5.4.1.4 of this text. A capacitive pressure transducer in which a diaphragm deflects as the result of the pressure differential across it is shown in Figure 6.38. The diaphragm is grounded, and thus forms a common plate with two capacitors such that $C_1 = C_0 - \Delta C$, $C_2 = C_0 + \Delta C$ and $\Delta C = k(P_1 - P_2)$. If a De Sauty bridge is operated at a frequency $f_o = 1/2\pi RC_0$ Hz, with peak excitation voltage V_s , it is easy to show that:

$$\frac{V_o}{V_s}(s) = \frac{1/s(C_0 - \Delta C)}{1/s(C_0 - \Delta C) + R} - \frac{1/s(C_0 + \Delta C)}{1/s(C_0 + \Delta C) + R} \quad (6.85)$$

After some algebra, we can write:

$$\frac{V_o}{V_s}(s) = \frac{s^2 R \Delta C}{(1 + sRC_o)^2 - s^2 R^2 C^2 (\Delta C/C_o)^2} \quad (6.86)$$

Letting $s = j\omega = j/RC_o$ and noting that $(\Delta C/C_o)^2 \ll 1$, we finally obtain:

$$\frac{V_o}{V_s}(j\omega) = \frac{\Delta C}{C_o} \quad (6.87)$$

or,

$$v_o(t) = V_s \left(\frac{\Delta C}{C_o} \right) \sin(t/RC_o) \quad (6.88)$$

The bridge output, V_o , in this example is a double sideband, suppressed carrier signal whose amplitude is $V_o = V_s(\Delta C/C_o)$. Since it is possible to resolve a V_o as small as $1 \mu\text{V}$, when V_s is typically 5 V pk , and C_o is 100 pF , ΔC can be as small as $2 \times 10^{-4} \text{ pF}$!

Pressure measurements can also be made using a capacitor with fixed geometry but with a dielectric constant which changes with pressure. Certain nonpolar liquids, gases and piezoelectric materials can be used for this purpose. Lion (1959) reports that the dielectric constant of air at 19°C varies from 1.0006 at 1 atm to 1.0548 at 100 atm. Dielectric constants also change with temperature, so a null transducer can be used in a bridge circuit to obtain temperature compensation.

Dranetz, Howatt and Crownover, cited by Lion, developed a titanate ceramic called Thermacon, in which the dielectric constant decreased linearly over the temperature range of -40 to $+160^\circ\text{C}$. One specimen was described, which had a dielectric constant of 490 at room temperature and a dielectric constant tempco of $-0.003/\text{ }^\circ\text{C}$. However, the tempco increases dramatically as the Curie temperature is approached.

6.7 Fiber Optic Sensors

Fiber optic (FO) dielectric waveguides were developed primarily for broadband, long distance communications links. Their main advantages are low cost (silicon vs copper), broad signal bandwidths and immunity from interference caused by electromagnetic radiation, such as radio waves, pulses (EMPs) generated by lightning, or perish the thought, nuclear explosions. They may also be used in harsh environments, remote from their electro-optic signal conditioning systems.

Interestingly, fiber optic cables are also useful in a variety of sensor applications [Wolfbeis, 1991]. They have been used in the measurement of electric current, magnetic fields, temperature, force, pressure, strain, acceleration and pH. Several of their physical/optical properties have been used in realizing these applications. For example, (micro)bending of an optical fiber will result in the loss of light energy from the core into the cladding at the bend, decreasing the output light intensity. This phenomenon is used in sensing force, pressure, strain, etc. [Hochburg, 1986]. The pH may be sensed by measuring the emission of a fluorescent dye coating the end of the fiber—the emission characteristic of the dye changes with pH. The pH has also been measured spectrophotometrically with fiber optics—a pH sensitive indicator dye such as phenol red

is immobilized inside a cuprophan membrane at the end of a pair of FO cables. Light at two wavelengths is transmitted to the dye. Reflectance at one wavelength is pH dependent, reflectance at the other wavelength is independent of pH and is used as a reference. Reflected light at both wavelengths is collected by the output FO cable and processed. System output is linear with pH over a limited range [Peterson and Goldstein, 1982].

Another effect that is used in FO based sensors is the modulation of the fiber's optical polarization properties by mechanical strain, temperature changes, or magnetic fields. Fiber optic cables are also used in Sagnac effect, FO laser gyroscopes. Their use in FO gyros is, however, expeditious, and it should be noted that the Sagnac effect will occur whether the light travels in a closed path in air directed by mirrors, or in a multturn FO coil. Sagnac effect gyros are discussed in detail in Section 7.1.2.

6.7.1 Magneto-Optic Current Sensors

One of the major problems in measuring the electrical current in high voltage, high power transmission lines, is that of measurement circuit isolation. One means of measuring heavy 60 Hz currents in a conductor is through the use of a current transformer. This transformer consists of a number of turns of wire wound on a high permeability, toroidal magnetic core. The current carrying conductor is passed through the 'hole' in the toroid and its magnetic field induces a voltage in the toroid's winding, which is proportional to the current in the conductor. This induced voltage is measured with an ac millivoltmeter. This means of measuring heavy ac currents works fine at lower voltages on the conductor. However, when the voltage on the conductor reaches values too high to safely isolate the toroid's winding (and meter circuit) from the conductor, some other means must be used to ensure safety.

The *Faraday Magneto-optic Effect* offers a means to measure high currents on conductors at extremely high voltages above ground (50 kV and higher), with reasonable accuracy and excellent isolation. Magneto-Optic Current Sensors (MOCSSs), which make use of the Faraday magneto-optic effect, use either solid glass pathways or fiber optic waveguides. When linearly polarized light is passed through a transparent, diamagnetic material through which a magnetic field is also passed in the same direction, there will be, in general, a rotation of the polarization vector of the emergent ray. It may be shown that the polarization rotation angle is given by:

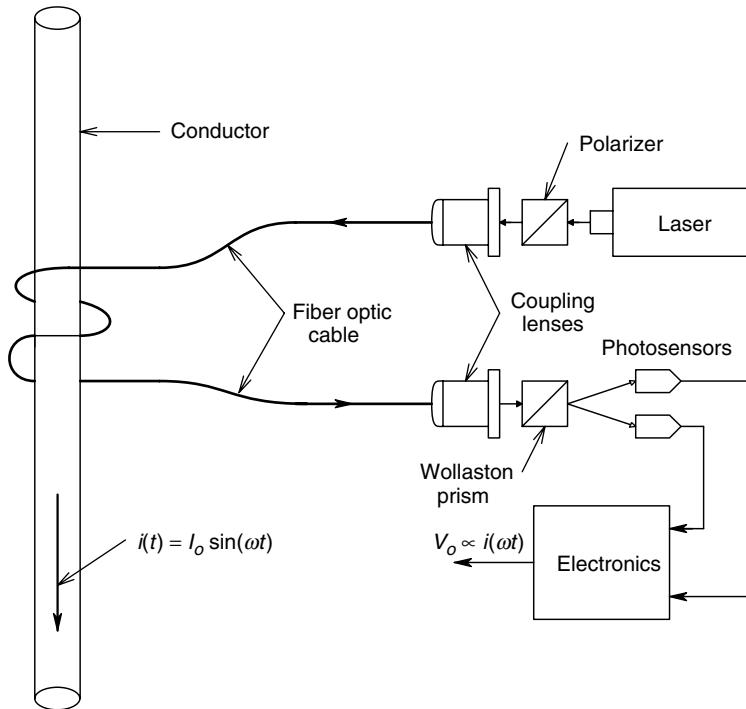
$$\alpha = V \oint \mathbf{H} \cdot d\mathbf{l} \quad (6.89)$$

The rotation, α , is proportional to the line integral of the magnetic field intensity vector, \mathbf{H} , along the light propagation path, \mathbf{l} , which encloses the magnetic field. V is the Verdet constant for the material used. Note that by the Ampere circuital law,

$$\oint \mathbf{H} \cdot d\mathbf{l} = I \quad (6.90)$$

where I is the current enclosed by one optical path loop. Hence, if N loops of a fiber optic cable are wound on a form so that all turns are aligned and have the same area, we may express the Faraday magneto-optical polarization rotation as:

$$\alpha = VNI \quad (6.91)$$

**FIGURE 6.39**

A fiber optic, electrical current sensor using the Faraday magneto-optical effect. The Wollaston prism separates the magneto-optically rotated E vector of the emergent light into two orthogonal components. The two orthogonal components are detected by photodiodes and their signals are processed by a ratio detector.

Figure 6.39 illustrates one version of a Faraday current transducer consisting of N turns of single mode optical fiber wound around a current carrying conductor. The optical rotation of the linearly polarized light entering the fiber is given by equation 6.90. There are several means of sensing small changes in the optical rotation of linearly polarized light, which are discussed below. The optical rotation, in theory, will remain unchanged regardless of the position of the current carrying conductor in the FO coil. Also, by using a diamagnetic material such as glass fiber, the system has substantial temperature independence [Rogers, 1973].

Another version of the Faraday MOCS, developed by Cease and Johnston (1990) is shown in Figure 6.40. Here, a single block of Double Extra Dense Flint (DEDF) Glass is used. It acts both as the Faraday medium and as an electrical insulator for the bus bar. For this geometry, the optical rotation angle is given by:

$$\alpha = V'BL \quad (6.92)$$

where V' is the Verdet constant, B is the magnetic flux density in teslas and L is the optical path length over which the Faraday effect occurs. Möller (1988) gives a table of Verdet constants. From this table we see that V' for ethanol at 25°C is 11.12×10^{-3} and V' for dense lead glass at 16°C is 77.9×10^{-3} minutes arc/gauss cm.

The magnetic field around a current carrying conductor at a distance R , where $R \gg r$, and r is the conductor's radius, may be shown by the Biot-Savart law to be given by:

$$B = \mu_0 I / 2\pi R \text{ Wb/m}^2 \quad (\text{in MKS units}) \quad (6.93)$$

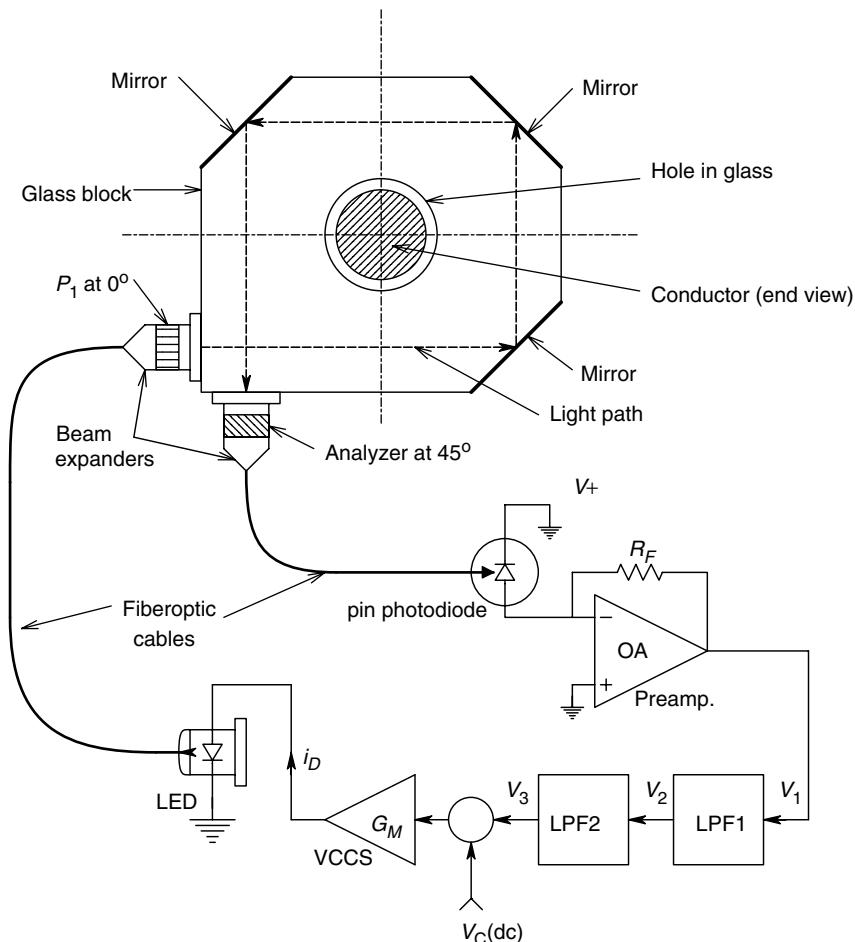
**FIGURE 6.40**

Diagram of another Faraday magneto-optical current sensor (MOCS) devised by Cease and Johnston (1990).

B in gauss equals 10^4 Wb/m^2 . Everywhere, B is tangential to a circle of radius R centered on the conductor. Such a magnetic field is called solenoidal and if the light path consists of N turns of an optical fiber wound on a circular form of radius R , concentric with the conductor, then the net Faraday rotation is given by:

$$\begin{aligned}\alpha &= V'(\mu_0 I / 2\pi R)(2\pi RN)(10^6) \\ &= V'\mu_0 I N(10^6) \text{ minutes of arc}\end{aligned}\quad (6.94)$$

For example, a MOCS has 100 turns of a lead glass fiber with a Verdet constant $V' = 80 \times 10^{-3}$ minutes arc/gauss cm, a current of 1 A will produce a B field which gives a Faraday rotation of 10.055 minutes, or $0.1676^\circ/\text{A}$. The constant, 10^6 , converts the Verdet constant, V' , in minutes arc/gauss cm to the MKS V in units of minutes arc/Wb/m²; $\mu_0 = 1.257 \times 10^{-6}$ Wb/Am.

As you will see in the next section, it is entirely possible to measure small angles of optical rotation to better than $\pm 0.001^\circ$. This polarimeter resolution sets the ultimate resolution of a FO MOCS.

6.7.2 Measurement of the Optical Rotation of the Linearly Polarized Light Output of Certain Optical Sensors

In order to consider the measurement of the rotation angle of linearly polarized light, we must first review the mathematical description of light as a propagating transverse electromagnetic wave in a dielectric medium. As is well-known, the propagation of light can be described by Maxwell's equations. In the simplest case, we will assume a plane wave propagating in the z direction in a source free, isotropic medium with permittivity, ϵ , and permeability, μ . From the Maxwell equations, we have

$$\nabla \times \mathbf{E} = -dB/dt \quad (6.95a)$$

$$\nabla \times \mathbf{B} = (\mu\epsilon)dE/dt \quad (6.95b)$$

$$\nabla \cdot \mathbf{E} = 0 \quad (6.95c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (6.95d)$$

The second order wave equation may be derived from the Maxwell vector equations, 6.95:

$$\nabla^2 \mathbf{E} - (\mu\epsilon)d^2 E/dt^2 = 0 \quad (6.96)$$

Assuming sinusoidal steady state conditions, we can write:

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0 \quad (6.97)$$

where k is the wavenumber defined as:

$$k \equiv 2\pi f/c = 2\pi/\lambda = 2\pi f(\mu\epsilon)^{1/2} = \omega/c \quad (6.98)$$

Note that in an anisotropic medium, k can be expressed as a vector:

$$\mathbf{k} = \mathbf{a}_x k_x + \mathbf{a}_y k_y + \mathbf{a}_z k_z \quad (6.99)$$

Here, we use \mathbf{a}_x as a unit vector pointing along the positive x axis and k_x is the wavenumber in the x direction, etc. The anisotropy may be considered to be due to different propagation velocities in the x , y and z directions (e.g. $c_x = 1/(\epsilon_x \mu_x)^{1/2}$).

We note that the \mathbf{B} and \mathbf{E} field vectors are mutually perpendicular to each other, and to the direction of wave propagation, which we will take as the positive z axis.

The polarization of an electromagnetic wave is defined in accordance with the IEEE standard [Balanis, 1989]. The polarization of a radiated EM wave is defined as 'that property of a radiated electromagnetic wave describing the time varying direction and relative magnitude of the electric field vector; specifically, the figure traced as a function of time by the extremity of the vector at a fixed location in space, and the sense in which it is traced, as observed along the direction of propagation.' There are three categories of EM wave polarization—linear, circular and elliptical. If the \mathbf{E} vector at a point in space as a function of time is always directed along a line which is normal to the direction of wave propagation, the field is said to be linearly polarized. To illustrate this

property mathematically, we assume that the \mathbf{E} vector has x and y components and may be written as:

$$\begin{aligned}\mathbf{E} &= \mathbf{a}_x \mathbf{E}_x + \mathbf{a}_y \mathbf{E}_y = \operatorname{Re}[\mathbf{a}_x \mathbf{E}_x^+ \exp(j\{\omega t - kz\}) + \mathbf{a}_y \mathbf{E}_y^+ \exp(j\{\omega t - kz\})] \\ &= \mathbf{a}_x E_{xr} \cos(\omega t - kz + \phi_x) + \mathbf{a}_y E_{yr} \cos(\omega t - kz + \phi_y)\end{aligned}\quad (6.100)$$

where \mathbf{E}_x^+ and \mathbf{E}_y^+ are complex, and E_{xr} and E_{yr} are real. Let us consider two cases of linear polarization. First, let $E_{yr} \equiv 0$, and $z = 0$ for convenience. This means that the locus of the instantaneous \mathbf{E} field vector is a straight line which is directed along the x axis. The \mathbf{E} field is given by:

$$\mathbf{E} = \mathbf{a}_x E_{xr} \cos(\omega t + \phi_x) \quad (6.101)$$

In a more general case of linear polarization, we let $\phi_x = \phi_y = \phi$, $z = 0$ and we find that the \mathbf{E} vector is given by:

$$\mathbf{E} = \left[\sqrt{(E_{xr})^2 + (E_{yr})^2} \right] \cos(\omega t + \phi) \quad (6.102)$$

This gives a straight line at an angle α with the x axis for all time, as shown in Figure 6.41. The polarization angle α can be written:

$$\alpha = \tan^{-1}(E_{yr}/E_{xr}) \quad (6.103)$$

Circular polarization is said to occur when the tip of the \mathbf{E} vector traces a circular locus in the xy plane when viewed by an observer standing at the origin, looking in the positive z direction (the EM wave is assumed to propagate in the $+z$ direction). Circular polarization can be either clockwise (right handed) or counter clockwise (left handed), as viewed looking from the rear of the EM wave in the $+z$ direction. It may be shown that circular polarization will occur when the \mathbf{E} field has two orthogonal, linearly polarized components with equal magnitudes and the two orthogonal components must have a time phase difference of odd multiples of 90° . As an example of circular polarization, let us assume in equation 6.100 that $\phi_x = 0$, $\phi_y = -\pi/2$, $z = 0$,

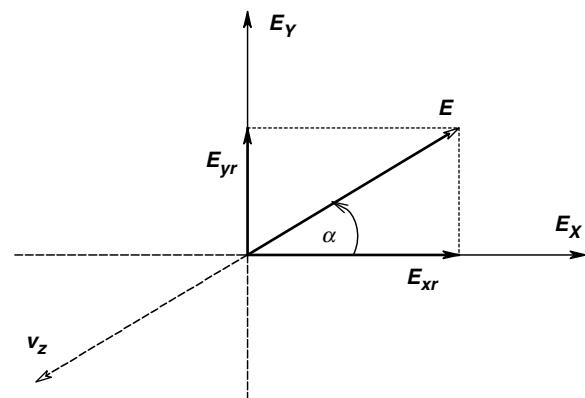


FIGURE 6.41

\mathbf{E} vector of light linearly polarized along the α -direction.

and $E_{xr} = E_{yr} = E_{CW}$. Thus:

$$\mathbf{E}_x = \mathbf{a}_x E_{CW} \cos(\omega t) \quad (6.104)$$

$$\mathbf{E}_y = \mathbf{a}_y E_{CW} \cos(\omega t - \pi/2) = \mathbf{a}_y E_{CW} \sin(\omega t) \quad (6.105)$$

The locus of the amplitude of the \mathbf{E} vector in the xy plane is just

$$|\mathbf{E}| = \sqrt{\mathbf{E}_x^2 + \mathbf{E}_y^2} = \sqrt{E_{CW}^2[\cos^2(\omega t) + \sin^2(\omega t)]} = E_{CW} \quad (6.106)$$

At any instant of time, \mathbf{E} is directed along a line making an angle α with the x axis, so:

$$\alpha = \tan^{-1}\left(\frac{E_{CW} \sin(\omega t)}{E_{CW} \cos(\omega t)}\right) = \tan^{-1}[\tan(\omega t)] = \omega t \quad (6.107)$$

Thus the angle α rotates clockwise with an angular velocity ω . The right hand, circularly polarized \mathbf{E} vector can also be written [Balanis, 1989]:

$$\mathbf{E} = E_{CW} \operatorname{Re}\{(a_x - ja_y) \exp[j(\omega t - kz)]\} \quad (6.108)$$

Using the development above, it is easy to demonstrate that a counterclockwise (left hand) rotating circular polarization vector will occur when $\phi_x = 0, \phi_y = \pi/2$ and $E_{xr} = E_{yr} = E_{CCW}$. In this case, $E = E_{CCW}$ and $\alpha = -\omega t$. The \mathbf{E} vector may be written as:

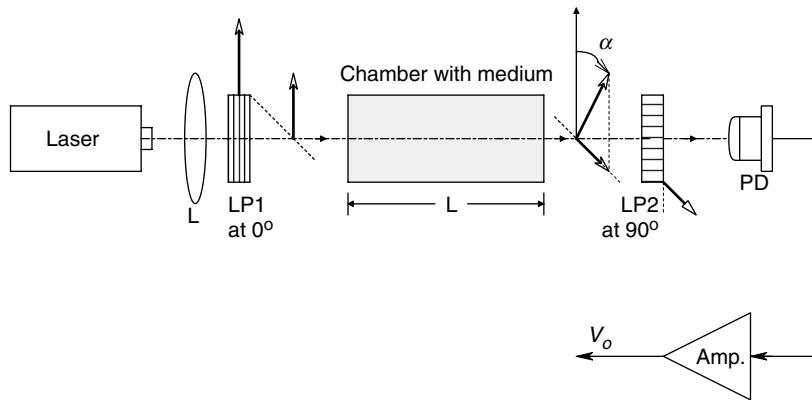
$$\mathbf{E} = E_{CCW} \operatorname{Re}\{(a_x + ja_y) \exp[j(\omega t - kz)]\} \quad (6.109)$$

It may be shown that linearly polarized light can be resolved into right and left hand circularly polarized components. This can be demonstrated physically by passing a beam of linearly polarized light through a quarter-wave plate optical device. In an isotropic medium, both right and left hand circular components of linear polarized light rotate at the same angular velocity. In an optically active, anisotropic medium, the velocity of light (in the z direction) is different for left and right hand circularly polarized waves. This difference in the propagation velocity of the L - and R -hand circularly polarized waves is referred to as circular birefringence. Coté *et al* (1990) shows that in an optically active, circularly birefringent medium, the net rotation of linearly polarized light exiting the medium is:

$$\alpha = (\pi z_o / \lambda) (n_R - n_L) \text{ rad} \quad (6.110)$$

where z_o is the distance the linearly polarized light propagates along the z axis through the circularly birefringent medium, and n_R and n_L are the indices of refraction in the medium for right and left hand circularly polarized light, respectively. (The index of refraction of a medium is defined as the ratio of the speed of light in vacuum to the speed of light in the medium; it is generally greater than unity.)

An EM wave is elliptically polarized if the tip of the \mathbf{E} vector traces out an ellipse in the xy plane as the wave propagates in the $+z$ direction. As in the case of circular polarization, ellipses traced clockwise are right handed, while left handed ellipses are traversed counterclockwise. The treatment of elliptical polarization is beyond the scope

**FIGURE 6.42**

A basic optical polarimeter. The sample chamber length L is typically 10 cm.

of this text. The interested reader who wishes to pursue this topic should consult Section 4.4 in Balanis (1989).

Polarimeters are optoelectronic systems designed to measure the optical rotation produced by an active medium. As we have seen above, the medium can be a piece of glass, an optical fiber, or a solution containing optically active molecules (such as D-glucose).

The first and most basic type of electronic polarimeter design is shown in Figure 6.42. Monochromatic light from a source, S , is collimated and then linearly polarized at a reference angle of 0° by an input polarizer. The linearly polarized beam is then passed through the optically active medium, M , of length L . The E vector of the emergent beam is still linearly polarized, but is rotated some angle α . The emergent beam is then passed through an analyzer polarizer whose pass axis is at 90° to the input polarizer. If $\alpha = 0$, then the detector, D , will sense zero light intensity. If $\alpha > 0$, the light intensity will increase according to the relation:

$$I = (I_{\text{MAX}}/2)[1 - \cos(2\alpha)] \quad (6.111)$$

Thus, $\alpha = 90^\circ$ will result in maximum light transmission through the system to the detector. Obviously, the $I = f(\alpha)$ function is nonlinear. This nonlinearity is true even for $\alpha < 7^\circ$ where the intensity is given by $I \approx I_{\text{MAX}}(\alpha^2)$, where α is in radians. To eliminate system nonlinearity for small rotation angles, we can orient the polarization axis of the analyzer polarizer to 45° . Now the light intensity at the detector will be maximum for $\alpha = 45^\circ$ and a constant level will be detected for $\alpha = 0$. The intensity at the detector for a 45° analyzer can be written as:

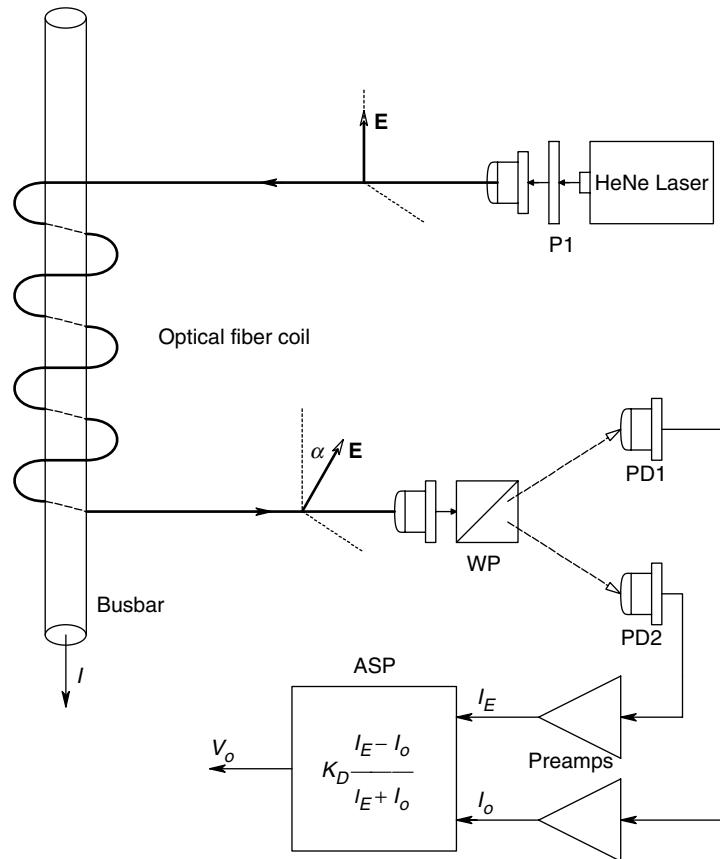
$$I = (I_{\text{MAX}}/2)[1 + \sin(2\alpha)] \quad (6.112)$$

Now, for $\alpha < 7^\circ$, I may be approximated by:

$$I \approx (I_{\text{MAX}}/2)[1 + 2\alpha] \quad (6.113)$$

where α is in radians. Thus the detector output will be linear in α and the dc bias may be subtracted from the voltage, which is proportional to the rotation. Limitations to the resolution and precision of the basic polarimeter are set by the noise in I_{MAX} and the noise in the electronic photodetector.

A second type of polarimeter, described by Rogers (1973), was used in the design of a fiber optic, Faraday magneto-optic, electric current sensor. The Faraday magneto-optic

**FIGURE 6.43**

Rogers' (1973) polarimeter used in the MOCS system shown in Figure 6.39.

effect, described in the preceding section, causes a rotation in the polarization angle of linearly polarized light traversing a single mode FO cable with turns around a current carrying bus bar. The emergent beam is collimated and passed through a Wollaston prism, WP, as shown in Figure 6.43. Two beams of light emerge from the prism and are directed to photodetectors PD1 and PD2. The Wollaston prism is made from two sections of calcite, a birefringent mineral having different refractive indices in two different directions. As light travels through the first prism, both the ordinary and extraordinary rays travel colinearly with different refractive indices. At the junction of the two prisms, the beams are interchanged so that the ordinary ray propagates with a lower refractive index and is refracted away from the normal to the interface, while the extraordinary ray is acted on by a higher refractive index and is refracted towards the normal. The divergence angle between the two beams is increased upon exiting the prism, which depends on the wavelength of the light used (the divergence angle is about 22° for HeNe laser light at 633 nm). If the input beam to the Wollaston prism is linearly polarized at an angle α with the x axis, the x and y components of the entering ray's \mathbf{E} vector will be resolved into the emergent rays: The upper (ordinary) ray will have its \mathbf{E} vector directed in the x direction, and be proportional to E_{xr} of the input ray, and the lower (extraordinary) ray will have its \mathbf{E} vector directed in the y direction, and be proportional to E_{yr} of the input ray. The detector outputs thus will be proportional to E_{xr}^2 and E_{yr}^2 . In the Rogers detector, the output is formed by analog processing the detector outputs

as described by the following equation:

$$V_o = K_D \frac{I_E - I_O}{I_E + I_O} \quad (6.114)$$

We note that if the analyzing Wollaston prism is rotated 45° clockwise with respect to the x axis of the polarized input ray to the FO coil, the intensity component of the extraordinary ray from the analyzer prism is given by:

$$I_E = KE^2 \cos^2(45^\circ - \alpha) \quad (6.115)$$

And the intensity of the ordinary emergent ray is:

$$I_O = KE^2 \sin^2(45^\circ - \alpha) \quad (6.116)$$

When these intensity expressions are substituted into equation 6.114, it is easy to show by trigonometric identities that the detector output is given by:

$$V_o = K_D \sin(2\alpha) \quad (6.117)$$

Equation 6.117 reduces to $V_o = K_D 2\alpha$, α being in radians, for $\alpha < 7^\circ$. Thus the Rogers ratio detector is substantially independent of noise affecting the \mathbf{E} , but is still affected by detector shot noise and noise in the ratio circuit. It should be noted that ac current in the bus bar will create a power line frequency \mathbf{B} field, which in turn, will give a sinusoidally varying α . Thus, we will see for V_o that,

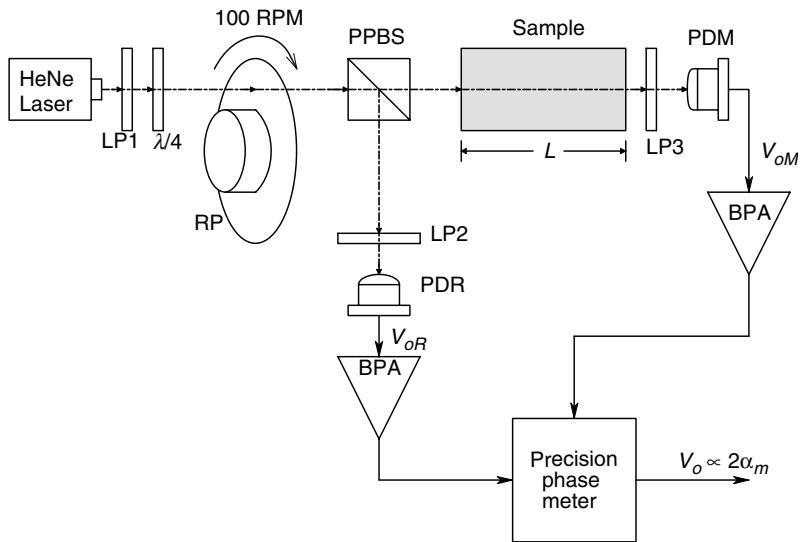
$$V_o(t) = K_D 2\alpha_o \sin(2\pi ft) \quad (6.118)$$

It is this signal which is proportional to $I(t)$ in the busbar.

A third means of resolving very small optical rotation angles was developed by Coté, et al (1992). The Coté system, which is a true angle measuring system, is shown in Figure 6.44. A collimated beam of laser light is passed through a linear polarizer, then through a quarter-wave plate to produce circularly polarized light. The circularly polarized ray is then acted on by a rotating, linear polarizer (RP). The emergent beam is linearly polarized with the angle of its \mathbf{E} vector rotating through 360°, with the angular velocity of the rotating polarizer, $2\pi f_r$ r/s. The emergent beam is split into a reference beam and a beam which is passed through the optically active medium, M, in which we wish to measure the optical rotation, α_m . The reference beam and the measurement beam from the medium are then each passed through analyzer polarizers to two photodetectors, PDR and PDM. The voltage output of each photodetector is proportional to the magnitude squared of the \mathbf{E} vectors emerging from the analyzer polarizers. In mathematical terms:

$$V_{oR} = \frac{K_D |\mathbf{E}_R|^2}{2} [1 + \cos(4\pi f_r t)] \quad (6.119a)$$

$$V_{oM} = \frac{K_D |\mathbf{E}_M|^2}{2} [1 + \cos(4\pi f_r t + 2\alpha_m)] \quad (6.119b)$$

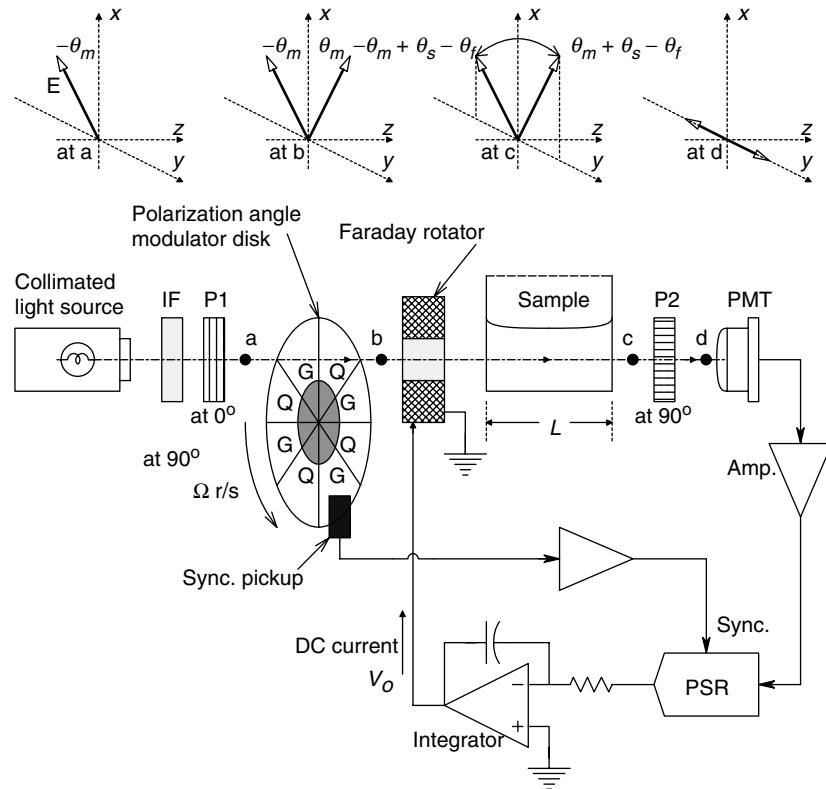
**FIGURE 6.44**

The millidegree polarimeter devised by Coté (1990). Note: LP = linear polarizer, $\lambda/4$ = quarter-wave plate, SM = synchronous motor, RP = rotating linear polarizer, PDR = reference photodetector, PDM = measurement photodetector, BPA = band pass amplifier, tuned to $2\omega_r$.

$V_{oM}(t)$ and $V_{oR}(t)$ are passed through bandpass filters tuned to $2f_r$ in order to eliminate their dc components and restrict the noise bandwidth. A precision phasemeter is used to directly measure the double optical rotation angle, $2\alpha_m$, from the cosine terms in equations 6.119 a and b. The magnitudes of E_R and E_M , and the noise associated with them, have little effect on the measurement of $2\alpha_m$. System resolution is better than $\pm 0.0005^\circ$ [Coté *et al.*, 1990]. Note that the rotating polarizer (RP) is effectively the 'chopper' in the Coté *et al.* system and effectively sets the system bandwidth by the Nyquist criterion. It is apparent that the Coté system will not work when measuring α_m varying at power line frequencies, unless the optical modulation frequency is made 600 to 1000 Hz. This system is best suited to measure small, static optical rotation angles such as those found when measuring glucose concentrations.

A fourth type of polarimeter, having an interesting closed loop design, was described by Gillham (1957). It had a noise limited resolution of about $\pm 0.0006^\circ$. A schematic diagram of the Gillham design is shown in Figure 6.45. Full-scale sensitivity was 0.02° for the open-loop version of the Gillham system, which makes use of a double side-band, suppressed carrier detection scheme. Referring to Figure 6.45, we see that monochromatic light passing through polarizer LP1 is linearly polarized at an angle $-\theta_m$ (clockwise) with respect to the x (vertical) axis. A polarizer chopper was made by cementing four equally spaced, quartz polarizers to a glass disc which was rotated at 1000 RPM. When the beam emerging from LP1 passes through the disk, the polarization vector is either unchanged as it passes through the glass, or is rotated $+2\theta_m$ (counter clockwise) when it passes through the quartz. Thus the beam emerging from the chopper wheel is alternately polarized at $-\theta_m$ and $+\theta_m$ at 66.7 cycles/s. with a square wave waveform.

The polarization modulated beam next passes through a Faraday rotator (FR), whose dc coil current causes an additional rotation of the chopped beam so that the beam exiting the FR has the angles, $(-\theta_m + \alpha_F)$ and $(+\theta_m + \alpha_F)$. The beam next interacts with the sample, S , whose optical rotation, α_S , we wish to measure. The ray emerging from

**FIGURE 6.45**

The polarimeter of Gilham (1957). A dc optical rotation of α_s is produced by the sample. The chopper disk rotates at 1000 RPM. It has alternating segments of plain glass, and linear polarizers which rotates the beam's E vector by 2θ degrees. The integrator forces a zero net dc optical rotation at (c), which gives zero dc light flux at (d). Note: P_1 and P_2 = orthogonal linear polarizers, PMT = photomultiplier photodetector, IF = interference filter monochromator.

the sample thus has alternate polarization vectors of $(-\theta_m + \alpha_S + \alpha_F)$ and $(+\theta_m + \alpha_S + \alpha_F)$. The beam next passes through polarizer LP2 on its way to the photodetector. LP2's axis is put at -90° to the input x axis, so that the photodetector's output is a 66.7 Hz square wave with peak-to-peak amplitude given by:

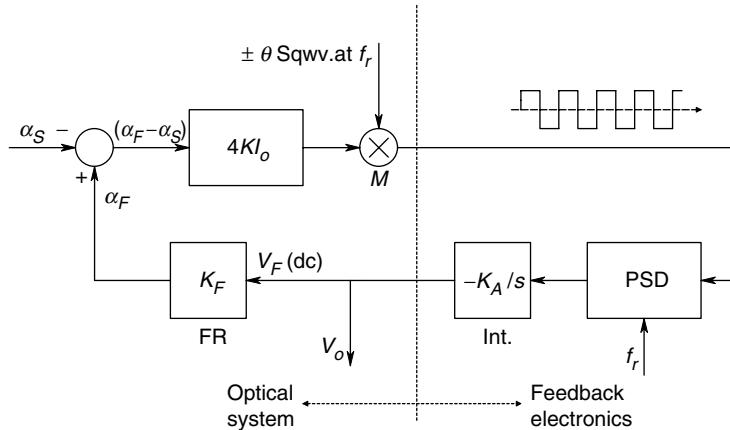
$$V_{d(PPK)} = (KI_o/2) \{ [1 - \cos(2(\theta_m + \alpha_S + \alpha_F))] \\ - [1 - \cos(2(-\theta_m + \alpha_S + \alpha_F))] \} \quad (6.120)$$

which reduces to:

$$V_{d(PPK)} = (KI_o/2) \{ \cos(2(\theta_m + \alpha_F - \alpha_S)) - \cos(2(\theta_m - \alpha_F + \alpha_S)) \} \quad (6.121)$$

By trigonometric identity, we simplify the equation 6.121 to find:

$$V_d(PPK) = (KI_o) \sin(2\theta_m) \sin(2\alpha_S - 2\alpha_F) \quad (6.122)$$

**FIGURE 6.46**

Systems block diagram of the feedback system used by Gilham (1957) to measure the optical rotation of the sample, α_S . The output of the phase sensitive demodulator, PSD, is integrated and the integrator output, V_o , is used to drive the Faraday rotator so its rotation is $\alpha_F = -\alpha_S$ in the steady state.

Assuming that the magnitudes of the arguments of the sine functions are less than 7° , we may finally write, noting that the angle values are in radians:

$$V_{d(PPK)} = 4K I_o \theta_m (\alpha_S - \alpha_F) \quad (6.123)$$

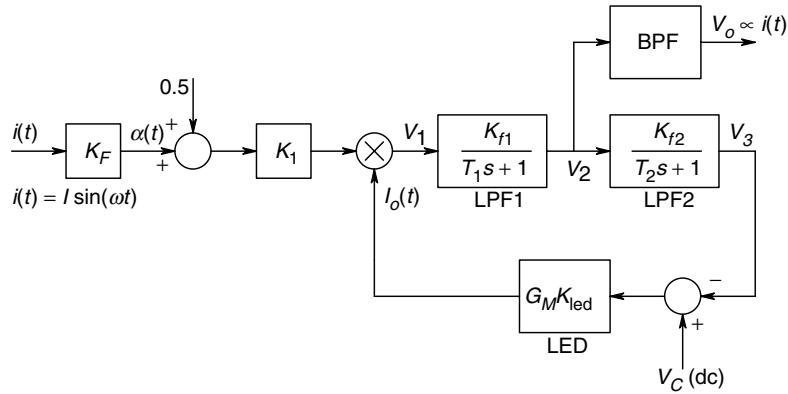
From equation 6.123, we see that if the Faraday rotator coil current is zero, the waveform observed at the detector is a true square wave, double sideband, suppressed carrier modulated signal. Gillham used $\theta_m = 3^\circ$. If the Faraday rotator is included in a feedback loop, as shown in the block diagram of Figure 6.46, it is easy to show that the steady state output of the integrator, V_o , is exactly:

$$V_o = \alpha_S / K_F \quad (6.124)$$

where K_F is the Faraday rotator constant in radians/volt. This servo loop generates a Faraday rotation angle in the steady state which is equal and opposite to the sample angle. Due to its relatively low modulation frequency (66.7 pps), the Gillham system is not suitable to measure 50 or 60 Hz power line currents with Faraday magneto-optical current sensors, rather it is intended to measure small ($< 0.03^\circ$), static optical rotation angles.

A fifth type of polarimeter which also uses feedback was described by Cease and Johnston (1990). Their system is shown schematically in Figure 6.40. Instead of using multiple turns of FO cable wrapped solenoidally around the current carrying busbar, they used a single block of quartz glass as a single turn optical pathway through which the busbar passes. The ac current in the busbar causes a proportional optical rotation, α , by the Faraday magneto-optical effect (positive α is taken here as counterclockwise). An analyzer polarizer, LP2, is assumed to have its axis at $+45^\circ$ (counterclockwise) to the vertical (0°) polarization axis of the input polarizer, LP1. Thus, the detector output voltage, V_1 , may be expressed as:

$$V_1 = (K_1 I_o / 2)[1 + \sin(2\alpha)] \quad (6.125)$$

**FIGURE 6.47**

Block diagram of the polarimeter used in the MOCS described by Cease and Johnston (1990).

For $\alpha < 7^\circ$, equation 6.125 reduces to:

$$V_1 \cong K_1 I_o [0.5 + \alpha] \quad (6.126)$$

where, as expected, α is in radians. As is seen in the system block diagram, Figure 6.47, the photodetector output, V_1 , is conditioned by two, simple lowpass filters. The output of the second filter, V_3 , is subtracted from a dc level, V_c , and this difference is amplified and used to control the average light intensity, I_o , of the LED light source for the system. Thus I_o is given by:

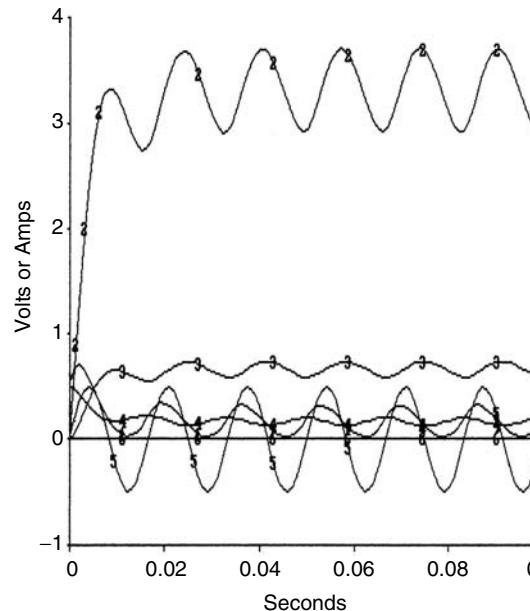
$$I_o = G_m K_{\text{LED}} (V_c - V_3) \quad (6.127)$$

It is seen that this feedback system is nonlinear, and that a simple pencil and paper analysis of the transfer function for V_2 vs α is out of the question. Accordingly, we may simulate the system's behaviour using Simnon™, a computer program specialized for the simulation of complex nonlinear systems. The short Simnon program is:

```

CONTINUOUS SYSTEM moct1
TIME t
STATE V2 V3
DER dV2 dV3
dV2=Kf1*V1/T1 - V2/T1    "Output of first LPF.
dV3=Kf2*V2/T2 - V3/T2    "Output of second LPF.
ALPHA=Kf*I      "Optical rotation angle from Faraday
"effect.
V1=0.5*Io*K1*(1 + sin(2*ALPHA))  "Output of PIN photodiode amp.
Io=GmKled*(Vc - V3)        "Output intensity of LED.
I=Ipk*sin(377*t)      "AC current in busbar.
"CONSTANTS: .....
K1: 2
Kf: 5.E-4
Kf1: 20
Kf2: 0.2
T1: 2.E-2
T2: 1.E-3

```

**FIGURE 6.48**

Results of Simmon simulation of the block diagram system of Figure 6.47. The ac component of voltage V_2 (trace 2) is proportional to $I(t)$ in the busbar, and would be isolated by a bandpass filter. Other traces are: 1, V_1 ; 3, V_3 ; 4, I_0 ; 5, alpha; 6, zero.

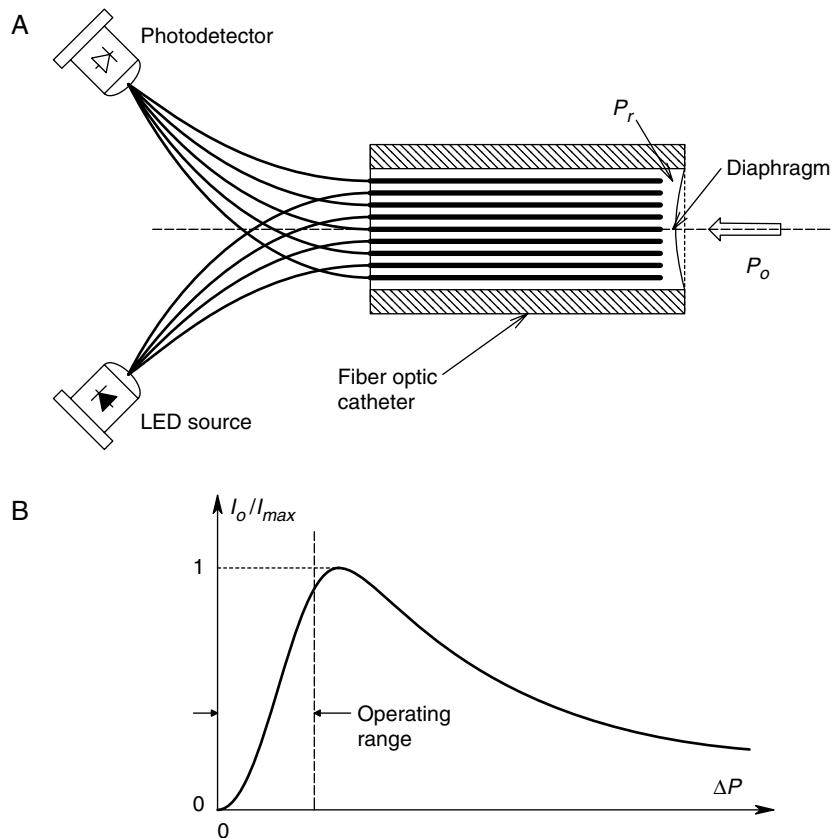
The original Cease and Johnston paper gave no constants for their system's components, so we arrived at the values above by reasonable estimates and trial and error. Figure 6.48 illustrates the results when the program above is run with the constants given above. Note that the photodetector output, V_1 , is a sinusoid of line frequency whose peak-to-peak amplitude in the steady state is proportional to the current in the busbar. The dc level of V_2 , however, must be removed by passing V_2 through a bandpass filter (not shown in the original Cease and Johnston paper). It is the ac voltage output of the bandpass filter which is actually measured to give the value of $I(t)$ in the busbar. In our simulation of this system, we noted that the linear proportionality between the peak-to-peak V_2 and $I(t)$ is lost for $I_{pk} > 1200$ A, for the constants we used. An I_{pk} of 2000 A gives severely distorted peaks on the V_1 waveform.

In conclusion, we have seen that there are a number of ways that can be used to resolve optical rotation angles as small as 0.001° . Such small polarization angles may occur in instrumentation systems that make use of the Faraday magneto-optic effect, the electro-optic effect and the electrogyration effect [Rogers, 1976].

6.7.3 Fiber Optic (FO) Mechanosensors

In this section, we will review the operation of various mechanosensors using optical fibers and light (excluding fiber optic, Sagnac effect gyroscopes). Fiber optic systems have been devised to measure pressure, force, displacement, acceleration, etc., using a variety of mechanisms. Hochberg (1986) places FO sensors in two classes—extrinsic FO sensors, in which the FO cable(s) is/are passive communications link(s) to an optically active terminating element, and intrinsic FO sensors in which some optical property of the FO cable itself is changed by the QUM.

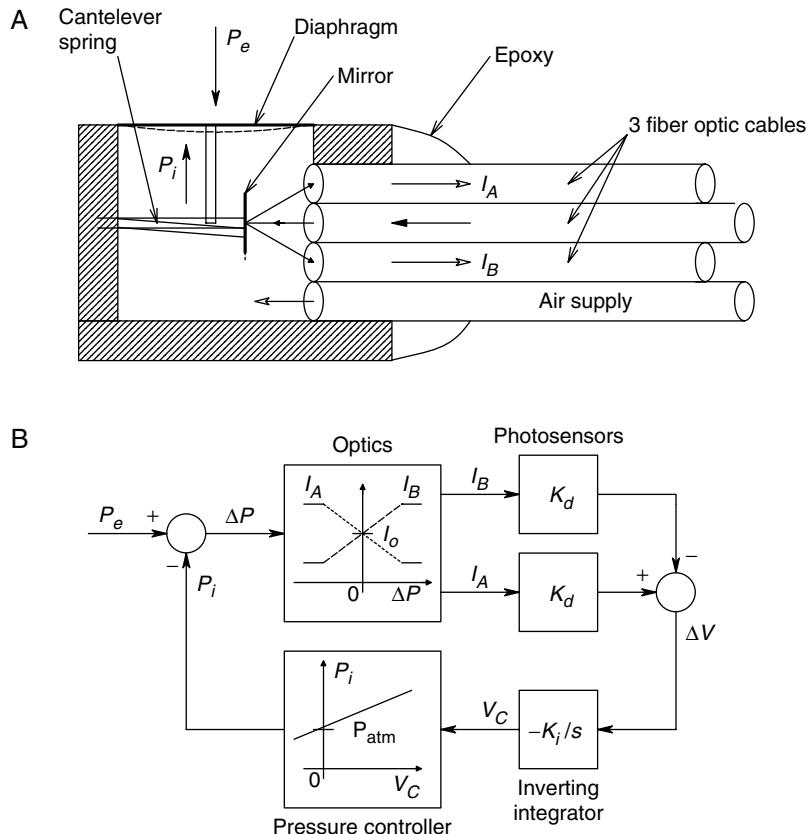
Examples of extrinsic sensors are the intravascular pressure transducer developed by Hansen (1983) (Figure 6.49), and the feedback pressure sensor developed by Neuman (1988) for measuring intracranial pressures. The Neuman sensor, shown in Figure 6.50A,

**FIGURE 6.49**

A. Mixed fiber, FO pressure sensor. The pressure difference across the diaphragm causes a change in the measured light output. B. Relative sensor output intensity *vs* pressure difference for the Hansen (1983) FO pressure sensor.

is a feedback or null system. The microdiaphragm deflects proportional to the pressure difference across it. An increased external pressure, P_e , causes an inward deflection of the diaphragm, in turn causing less light to be reflected from the input optical fiber to fiber A, and more light to enter fiber B. Fibers A and B go to two photodiodes. The conditioned outputs of the photodiodes are subtracted to form an error signal that, if negative, causes the pressure inside the probe, P_i , to increase, forcing the diaphragm outward until the error signal is zero. At this point, the diaphragm is in its equilibrium position, and $P_i = P_e - P_i$ is easily measured by a standard pneumatic pressure sensor, and in the steady state, is equal to P_e . The block diagram for a simple controller for the Neuman probe is shown in Figure 6.50B. Other extrinsic FO sensors are shown in Figure 6.51.

One intrinsic FO mechanosensor makes use of the property of optical fibers in which micro bending causes a net decrease in the transmitted light intensity because of the loss of light power into the cladding—the bending angle prevents internal reflection of light back to the core (Figure 6.52). The microbending obviously requires small deflections of the deformers relative to the fiber and can be the result of the deflection of a diaphragm (pressure sensor), or elongation of a load cell (isometric force sensor). Another intrinsic FO mechanosensor design makes use of interferometry. Figure 6.53, after

**FIGURE 6.50**

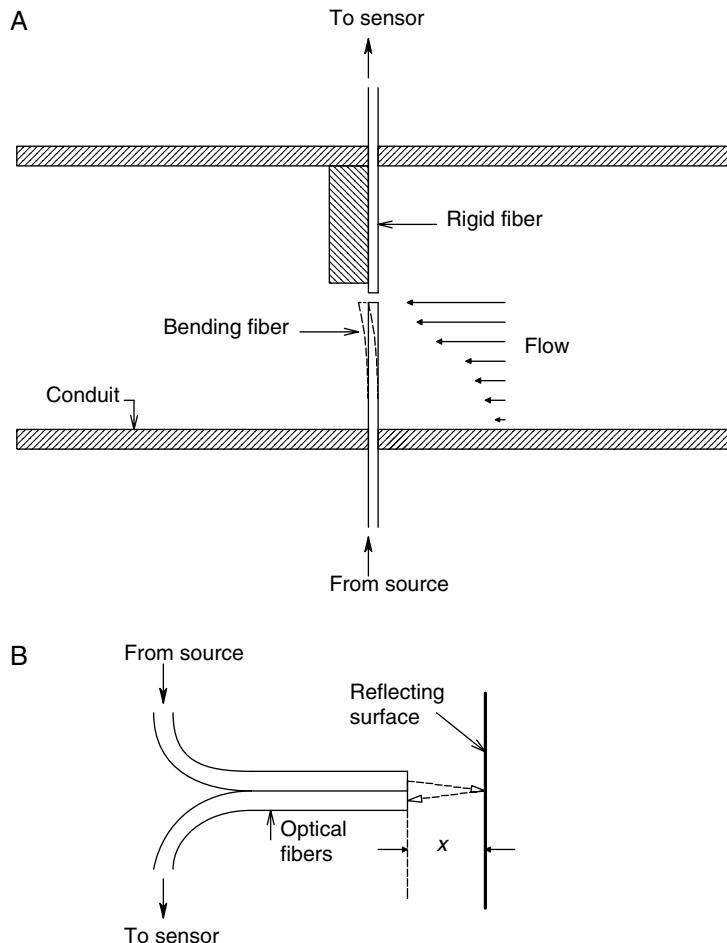
A. The Neuman (1988) pneumatic feedback pressure sensor. When the internal pressure equals the external pressure being measured, there is no deflection of the diaphragm, and the light outputs are equal. B. Block diagram of a simple controller for the Neuman pressure sensor. The system requires a voltage-to-pressure transducer to adjust the internal pressure of the sensor probe.

Hochberg (1986), illustrates an interferometric sensor system. The detector utilizes the light and dark interference patterns caused by combining the reference and measurement fiber light outputs. In this case, some net deformation of the measurement fiber causes a change in the fiber core's refractive index, hence a phase change in the light output and an interference fringe shift which can be sensed by the detector. The spatial distribution of light in an interference pattern at its center generally follows a function of the form

$$I(X) = (I_0/2)[1 + \cos(4\pi X/X)] \quad (6.128)$$

where X is the spatial period of the fringes. The fringe shift occurs because of a small change (δx) in the spatial parameter, x . For maximum sensitivity, the fractional fringe shift should be sensed at the points of steepest slope of the cosinusoidal term in equation 6.128. Thus, $I(\delta x) \cong (I_0/2)(1 + 4\pi\delta x/X)$.

An intrinsic, FO pressure sensor system using polarized light and interferometry was described by Bock *et al* (1990). The Bock system is basically a one fiber, Mach-Zehnder interferometer in which the sensing fiber carries two optical modes differing in polarization and whose phase delays change differentially in response to the applied external pressure. The pressure sensor of Bock *et al* (1990) was shown to have almost perfect compensation for temperature changes. Figure 6.54 shows their experimental

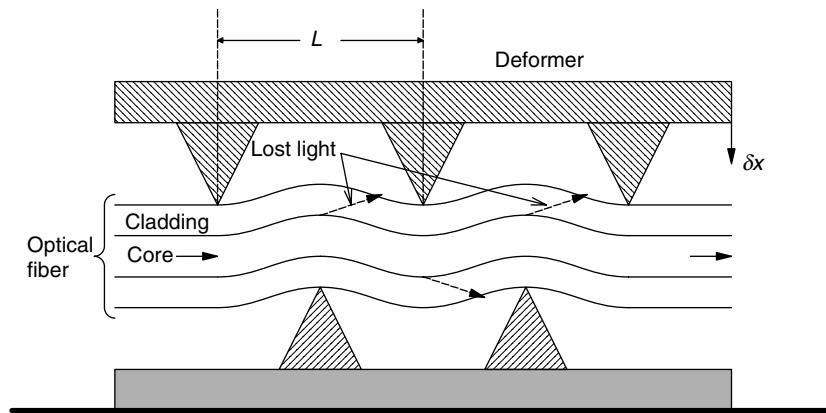
**FIGURE 6.51**

A. A basic, single fiber, extrinsic sensor in which some force (as from flow) causes lateral displacement of one fiber, decreasing light coupling. This type of sensor is inherently nonlinear. B. The optical coupling between two OFs varies with the distance from a plane reflecting surface in a nonlinear manner.

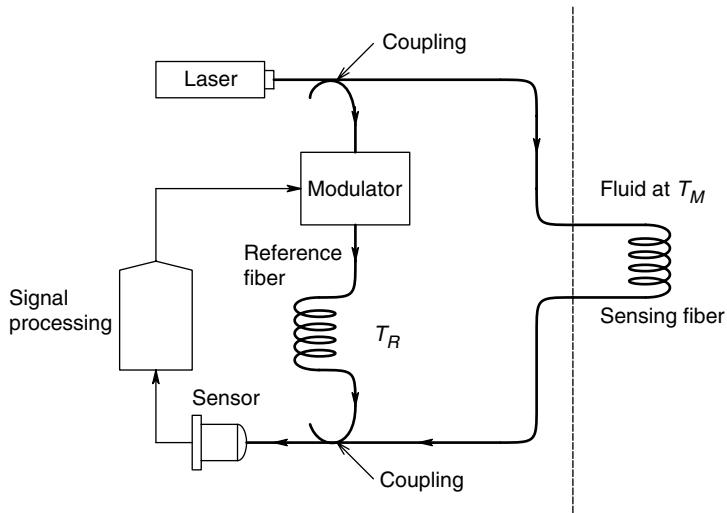
set-up, which used a HeNe laser as a source. The laser output was linearly polarized and launched parallel to one of the two principal axes of a highly birefringent (HB) fiber leading to the sensor. A bidirectional coupler couples light to the measurement fiber. The HB measurement fiber has a 45° axial splice (the reference section is rotated 45° with respect to the input segment). The reference section is of length L_1 . The sensing (terminal) section of HB fiber is axially spliced with a 90° rotation to the reference section. The terminal end of the sensing section of fiber has a gold mirror to reflect the light back down the fiber to the analyzer and detector. In another paper, Barwicz and Bock (1990) reported further on their FO pressure sensing system. The photodetector output voltage was given by the rather complex formula:

$$v(p, T) = (V_o/2)\{1 + \cos[(2\pi L/\lambda)(\lambda/L_B - C_p \beta \Delta T + |C|ap)]\} \quad (6.129)$$

where ΔT is the temperature difference from the reference temperature, p is the gauge pressure the FO sensor is exposed to, λ is the wavelength of light in the optical fibers,

**FIGURE 6.52**

Intrinsic FO sensors may be based on microbending of a transmitting fiber. Microbending causes a decrease in the intensity of the exiting light because light is lost into the cladding as the result of bending. A pressure sensor can be made, in which a membrane presses the microbending fixture into the OF.

**FIGURE 6.53**

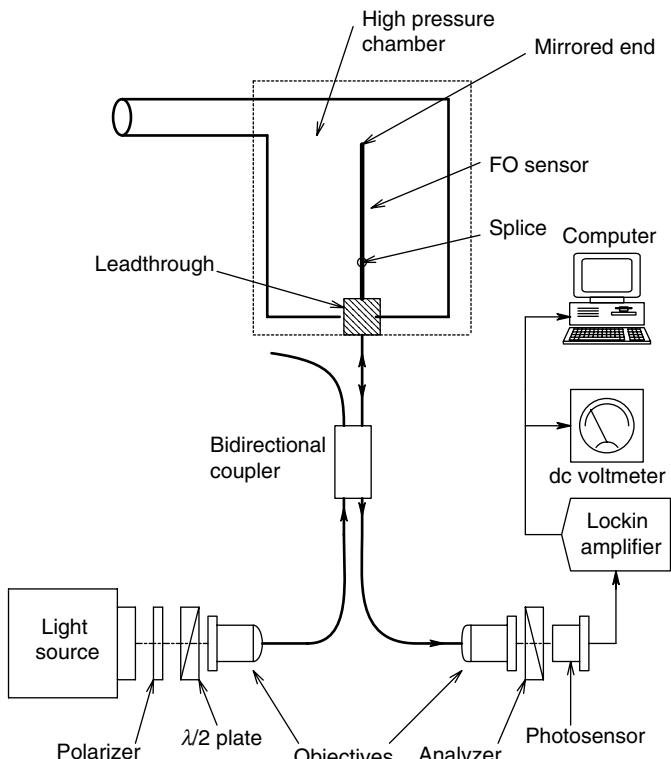
Schematic diagram of a two fiber interferometer in which some physical quantity acting on the sensor fiber causes a phase change between the two light paths, hence a change in the detected intensity of the interference pattern.

L_B is the beat length parameter of the fiber, L is the length of the fiber, a and β are material constants, C_p is the relative photoelastic constant and C is a pressure dependent photoelastic constant given by:

$$C = n^3(1 + \sigma/2E)(p_{12} - p_{11}) \quad (6.130)$$

in which n is the refractive index of the fiber core, σ is the Poisson ratio, E is Young's modulus, and p_{11} and p_{12} are optical strain coefficients.

Barwicz and Bock (1990) showed that $v(p)$ obtained experimentally follows equation 6.129 quite closely—the period of $v(p)$ varying inversely with sensing fiber length, L . For example, with $L=26$ mm, the period in $v(p)$ is 26 MPa (i.e. the sinusoidal curve repeats itself every 26 MPa over a 100 MPa range). Note that 1 psi = 6895 Pa. By using

**FIGURE 6.54**

Schematic diagram of the interferometric, high pressure sensor developed by Bock *et al* (1990). It is a one fiber, Mach-Zehnder interferometer which makes use of the pressure dependence of the sensing fiber's refractive index.

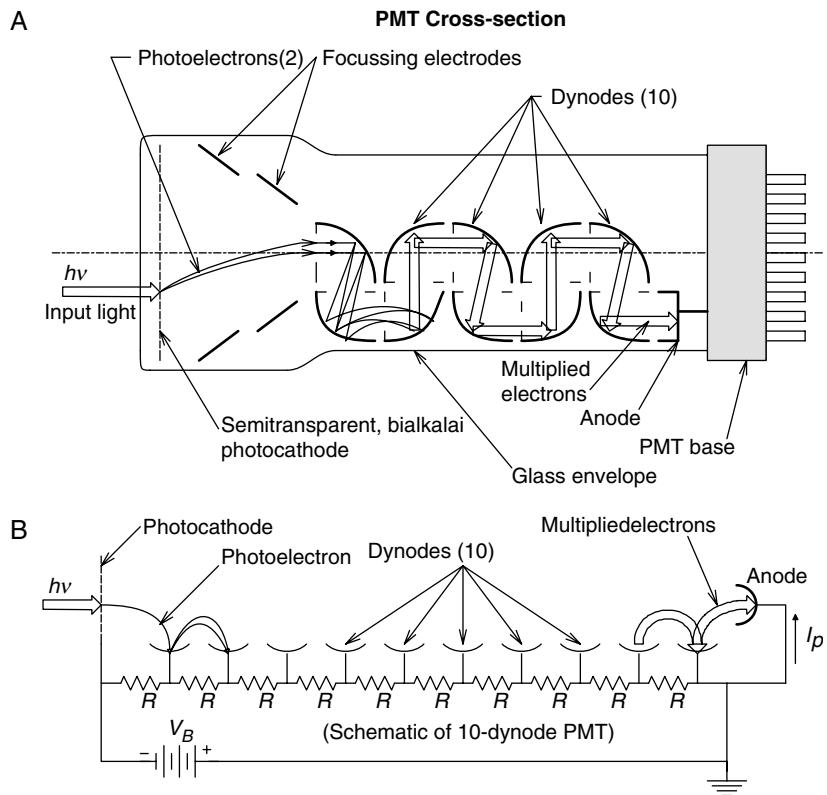
three different lengths of FO sensor and computer processing of the sensor outputs, Barwicz and Bock were able to cover a 100 MPa range with 0.01 MPa resolution. This implies an 80 dB SNR for their system for high pressure measurements.

Additional examples of FO mechanosensors may be found in the texts by Culshaw (1984), Marcuse (1981), Davis (1986) and Cheo (1990). The design of FO mechanosensors is a rapidly growing field, which has an active literature in instrumentation and in optoelectronics journals.

6.8 Photomultiplier Tubes and Related Electron Multiplication Devices

6.8.1 Introduction

Photomultiplier tubes (PMTs) are the most sensitive of visible range photon sensors. Under special operating conditions, they can respond in an almost 1:1 manner to single photons that strike their photocathodes. PMTs have been replaced in many applications by solid state sensors such as avalanche photodiodes. However, they remain the photon sensor of choice for ultra low light intensity sensing such as in fluorescent molecular tagging applications, Raman spectrophotometry, single bubble sonoluminescence (SBL) research and spectrophotometry. PMTs are vacuum tubes. Early PMTs had glass vacuum envelopes about the size of a beer can. Newer PMT designs are smaller. Figure 6.55 schematically illustrates a cross-section through a cylindrical end-on, PMT. Inside, is a photocathode, electron focusing electrodes and a series of 8-14 dynode

**FIGURE 6.55**

A. Cross-section of a 10-dynode, end window PMT. B. Schematic of the 10-dynode PMT. The dynodes are biased with a simple resistive voltage divider.

electrodes where electron multiplication takes place. Ten dynodes is a more common number. The multiplied electrons are collected by the anode and form the PMT's photocurrent. In side-on PMTs, the geometry is changed so that the photon (light) window is at right angles to the long axis of the envelope.

The photocathode is generally semi-transparent to permit greater efficiency in photon capture. It is often made from bialkali, low work function materials which determine its spectral sensitivity in mA of photoelectrons emitted per watt of light at a given wavelength. If a quantum of light with energy, $h\nu \geq \phi$, strikes the photocathode, an electron may be emitted. ϕ is the cathode work function in joules. Statistically, the ratio of emitted electrons to incident photons is the quantum efficiency, η . In general, $0 < \eta < 1$.

One photoelectron emitted from the cathode is accelerated in the PMT by the electric field between the cathode and the first dynode to the more positive first dynode electrode. It hits the first dynode with sufficient kinetic energy to release perhaps four electrons (on the average), which in turn are accelerated to the second dynode, which is at a fixed potential that is more positive than the first dynode. These electrons, too, gain enough kinetic energy to each dislodge say another four electrons. Now a total of 16 electrons strike the third dynode, and the process repeats itself, and so on. Thus, it is easy to see that $4^{10} \cong 1.05 \times 10^6$ electrons strike the anode as the result of one photoelectron being emitted in a 10-dynode PMT. The dynode-to-dynode potential is maintained by a simple resistive ladder voltage divider biasing circuit, as shown in the dynode schematic in Figure 6.55. The delay time between the emission of the

photoelectron from the photocathode until the current pulse is seen is typically ca. 20 ns in a 10-dynode PMT. A typical dynode-to-dynode potential is ca. 100 V.

Desirable characteristics in a PMT are:

1. High responsivity (A/W)
2. Low dark current
3. High output SNR
4. Wide spectral response
5. Low transport delay
6. High stability

Disadvantages of PMTs are:

1. Mechanically fragile (glass envelope)
2. Shapes and sizes are limited, physically large
3. Need stable high voltage power supplies
4. Expensive (hundreds of dollars)
5. Responsivity affected by magnetic fields, needs magnetic shielding in critical applications
6. Needs cooling to LN₂ temperatures for noise reduction in critical applications

Competing with PMTs are channel plate photomultiplier (CPM) devices and single channel PMTs, both of which are described below.

6.8.2 Operation of PMTs

Figure 6.56 illustrates the anode (output) spectral responsivity of five different Oriel PMTs. Each uses a different photocathode material. Note that some cathode materials

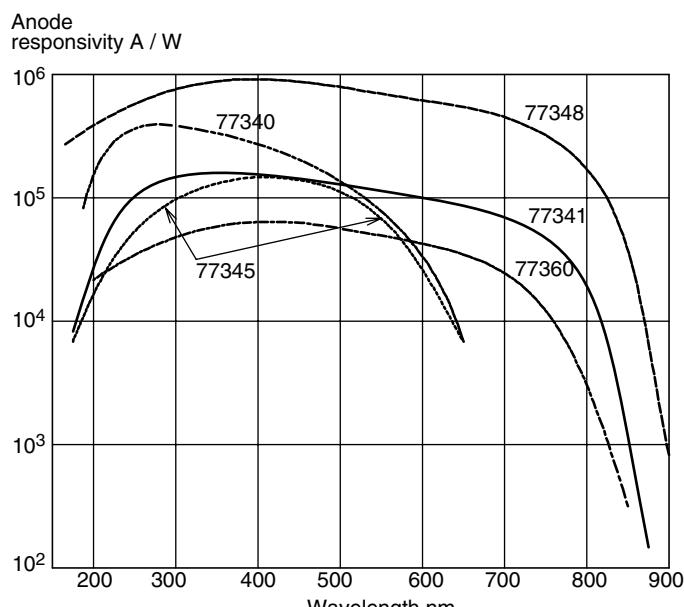


FIGURE 6.56
Spectral responsivity of five different Oriel PMTs.

permit PMT operation at UV wavelengths down to 185 nm (e.g. Hamamatsu R6925, a 9-dynode, side-on PMT which uses a bialkali cathode and a UV glass window), and others in the NIR to wavelengths as long as 1,700 nm (e.g. Hamamatsu R5509-72, a 10-dynode PMT which uses an InP/InGaAs cathode and a borosilicate glass window).

The gain or amplification of PMTs can be expressed in several ways:

1. An anode spectral responsivity, $R_A(\lambda)$, in anode A/W of incident light power on the photocathode
2. A pure current gain between cathode and anode, G_I
3. Photocathode responsivity $R_C(\lambda)$ in mA/W

$R_A(\lambda)$ and G_I depend on the inter-dynode potentials and the number of dynodes. For example, the Oriel model 77341 PMT operated at 1,000 V has an inter-dynode potential of 100 V, a maximum $R_A(\lambda) = 2 \times 10^5$ A/W, $G_I = 5 \times 10^6$ and $R_C(\lambda) = 0.04$ A/W. In addition, this PMT has a maximum anode current, $I_A = 100 \mu\text{A}$, from which we can conclude that the *maximum* input light power is 0.5 nW! The model 77341 PMT also responds to light wavelengths from 185 nm (UV) to 870 nm (NIR). It has a broad peak $R_C(\lambda)$ from ca. 300–550 nm. Its photocathode is 8 × 24 mm, its 10 to 90% rise time (anode current in response to a step of light) is 2.2 ns, and the transit time between a step of light and the anode current step is 22 ns (note that the electrons moving between dynodes travel at finite velocity, being accelerated by the 100 V inter-dynode potentials). In addition to the anode photocurrent, all PMTs have an anode dc dark current, I_{DA} . The I_{DA} has two principal components—amplified thermionic emission from the dark photocathode and an ohmic leakage component. Ohmic leakage can be the result of fingerprints and solder flux on the PMT base and socket. The model 77341 PMT has $I_{DA} = 2 \text{nA}$. The thermionic emission from the cathode I_{TC} can be modeled by:

$$I_{TC} = KT^2 \exp(-\phi/kT) \text{ Amps} \quad (6.131)$$

where K is a constant, T is the absolute (kelvin) temperature, ϕ is the cathode work function in joules, and k is Boltzmann's constant ($1.380 \times 10^{-23} \text{ J/K}$).

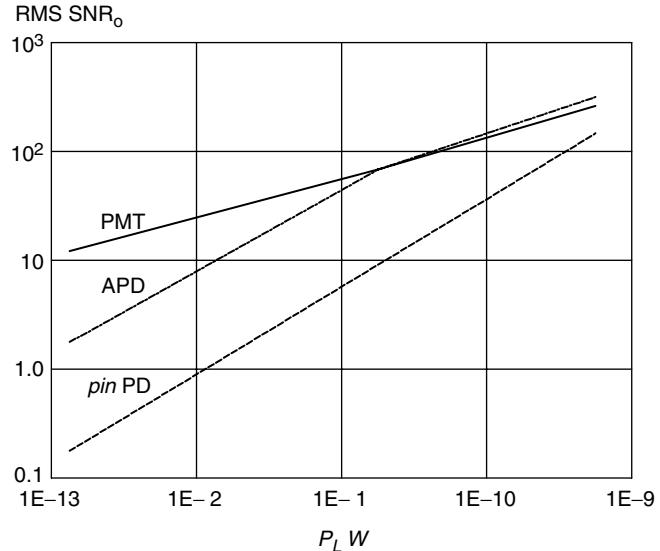
All PMTs are noisy. Noise in the anode current comes from four sources:

1. White Johnson or thermal noise from the series resistor used in the anode circuit
2. White noise from thermionic emission from the dynode surfaces (the resistors' Johnson and dynodes' shot noises are generally neglected in calculating PMT output signal-to-noise ratio, SNR_o)
3. White shot noise from the thermal emission of the (dark) photocathode
4. White shot noise from the photocathode's actual photocurrent

Thus, the MS anode (output) noise current can be written:

$$\overline{i_{an}^2} = G_I^2 [2q(I_{CP} + I_{TC})]B \text{ MSA} \quad (6.132)$$

where B is the Hz noise bandwidth (very high if we are counting pulses, low if dc light levels are concerned), the cathode photocurrent is $I_{CP} = R_C(\lambda)P_L$. P_L is the light power in watts at a wavelength, λ , incident on the photocathode. I_{TC} is the cathode's average thermionic emission current. It increases with the cathode's kelvin temperature.

**FIGURE 6.57**

Representative comparison of output SNRs for the three major classes of photosensors under varying input light power conditions. Note: PMT = photomultiplier tube, APD = avalanche photodiode, *pin*-PD = *pin*-photodiode.

Assume the output signal at the anode is the RMS photocurrent from the cathode multiplied by G_I . That is, $I_{AP} = I_{CP} G_I = G_I R_C(\lambda) P_L$. Thus, we can write the MS output SNR as:

$$\text{SNR}_o = \frac{G_I^2 R_C^2(\lambda) P_L^2}{2qB G_I^2 [I_{TC} + R_C(\lambda) P_L]} = \frac{R_C^2(\lambda) P_L^2}{2qB [I_{TC} + R_C(\lambda) P_L]} \quad (6.133)$$

Note that SNR_o is independent of the PMT's current gain, and increases with increasing input light power to the photocathode, P_L . A commonly used technique to increase SNR_o is to decrease I_{TC} by operating the PMT at the temperature of boiling liquid nitrogen (LN_2) (-196°C), or by cooling it thermo-electrically. Lion (1959) reported that there can be a 40 dB increase in a PMT's RMS SNR_o between room temperature operation and operation at LN_2 . Figure 6.57 shows that PMT SNR exceeds that of *pin* PDs and exceeds APDs for intensities $< 1 \times 10^{-11} \text{ W}$.

Let us now calculate the light power at 555 nm wavelength to give a PMT $\text{SNR}_o = 1$. We assume the PMT is cooled with LN_2 , so the cathode thermionic emission is ca. 1 electron/s, or $I_{TC} = 1.6 \times 10^{-19} \text{ A}$. We set the $\text{SNR}_o = 1$:

$$1 = \frac{R_C^2(\lambda) P_{L1}^2}{2qB [I_{TC} + R_C(\lambda) P_{L1}]} \quad (6.134)$$

Equation 6.134 yields a quadratic equation in P_{L1} :

$$-P_{L1}^2 = [2qB/R_C(\lambda)]P_{L1} + 2qBI_{TC}/R_C^2(\lambda) = 0 \quad (6.135)$$

Solving this quadratic equation for its real root, we find to a good numerical approximation:

$$P_{L1} \cong 2qB/R_C(\lambda) \quad (6.136)$$

Let us use the numerical values from the Oriel 77431 PMT cited above: $R_C(\lambda) = 0.04 \text{ A/W}$ and $B = 10^3 \text{ Hz}$. Thus $P_{L1} \cong 8 \times 10^{-15} \text{ W}$, which is mighty small!

A LN₂-chilled, 12-dynode PMT, properly operated, can be used to statistically detect single photons. Under extremely low light power conditions, a single photon may or may not generate a single photoelectron. If it does, the electron is multiplied and gives an output current pulse. Under these conditions, the dark current is due to single thermionic electrons from the cathode, which occur at about one per second. One means of resolving photoelectrons over thermionic events is differential counting. A count of output current pulses over a fixed time is made with the PMT covered to get the mean emission rate for cathode thermionic electrons. Then the constant photon source is presented to the PMT and the count redone. The extra counts are presumed to be due to photons. A pulse height window is used to exclude smaller anode output pulses coming from thermionic emission from the dynodes. Cathode thermionic electrons and photoelectrons are assumed to follow Poisson distributions. Maximum sensitivity of a LN₂-chilled, magnetically shielded 14-dynode PMT is of the order of 10^{-16} W at $\lambda = 555 \text{ nm}$, using a lock-in amplifier signal recovery system [Oriel Merlin™ system].

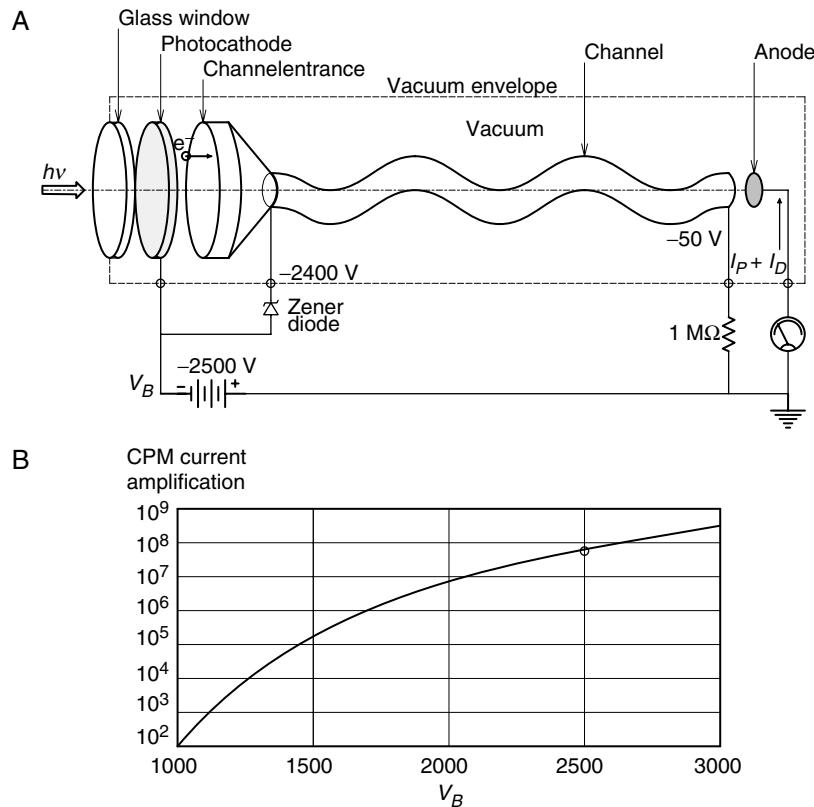
For fast pulse operation of a PMT, the dynodes #10, 11 and 12 are bypassed to ground with capacitors of the order of $0.02 \mu\text{F}$. The size of the capacitor bypassing the last (12th) dynode should be at least 100 times the output charge per pulse, sic:

$$C_{12} \geq 100 (I_{Appk} T_p) / V_{12} \quad (6.137)$$

Here I_{Appk} is the peak anode output current, T_p is the output pulse width and V_{12} is the potential to ground at the 12th dynode [Hamamatsu online catalog, Section 11, *Voltage Divider Consideration*].

6.8.3 The Single-Channel Photomultiplier

The Perkin-Elmer channel photomultiplier (CPM) tube has an incredibly simple and ingenious design. The envelope package is cylindrical, 75 mm in length and 10.5 mm in diameter. Four electrical leads come out from the bottom. A semitransparent cathode is deposited on the inner surface of the entrance window. Figure 6.58A illustrates a cross-section of the P-E CPM tube. (The photocathode is shown separated from the glass window for artistic simplicity.) In a major departure from PMT architecture using dynodes for electron multiplication, P-E found that by accelerating photoelectrons from the anode into the corkscrew shaped, hollow, resistive channel, they would collide with the inner walls of the channel, releasing secondary electrons which would also be accelerated and release still more electrons, etc. Current amplification is given as 5×10^7 with a supply voltage of $V_B = 2400 \text{ V}$. Supply (bias) current flowing up the resistive channel is about $50 \mu\text{A}$ with $V_B = 2400 \text{ V}$. Thus the total channel resistance is about $47 \text{ M}\Omega$. Anode sensitivity is $R_A(\lambda) = 3 \times 10^6 \text{ A/W}$ at $\lambda = 410 \text{ nm}$ for the P-E model C492 CPM. This CPM has a quartz window and a bialkali photocathode. It has a spectral response that lies between 165 and 650 nm. Also, $I_D = 80 \text{ nA}$, the anode current rise time is 3 ns and the pulse width (FWHM) is 6 ns for this CPM, with a 2400 V supply. One P-E CPM (model C911) works into the vacuum UV, down to 115 nm—it has a magnesium fluoride window. P-E claims that because the channel electron multiplication process is essentially silent, the CPM has a noise level one or two orders of magnitude lower than dynode PMTs, resulting in a significantly higher dynamic range.

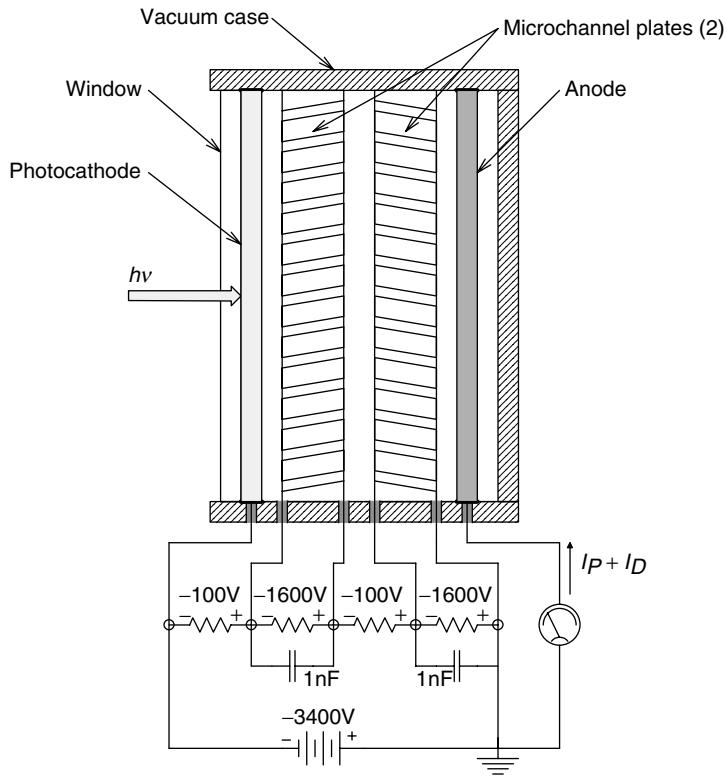
**FIGURE 6.58**

A. Cross-section of the Perkin-Elmer channel photomultiplier tube (PE-CPM tube). B. Plot of current gain (anode current/primary photocurrent) *vs* operating voltage, V_B .

6.8.4 Microchannel-Plate Photomultipliers

The microchannel plate PMT (MCP) is another dynodeless PMT design. The heart of an MCP is a lead glass disc through which have been inserted 10^4 – 10^7 small holes. The holes are typically 8 – $25\text{ }\mu\text{m}$ in diameter on 10 – $40\text{ }\mu\text{m}$ centers and 40 – 100 times longer than their diameters. Each hole acts as a miniature CPM device. A potential of ca. 4000 V is placed across the ends of the holes to accelerate the multiplied photoelectrons traversing the holes [Pankratz, 1998]. It has been found that the electron multiplier effect of the holes does not depend on the length l of the holes or on their diameters, d , but rather on the ratio, $\alpha = l/d$. Hence size reduction is possible, permitting many small holes to be placed in parallel in a MCP disk. The glass disk thickness itself is of the order of 2.5 mm . The axis of each hole is perpendicular to the disk's surfaces, or can be inclined ca. 8° from perpendicular. To establish the E -field across the tubes, a thin layer of a metal such as Nichrome or Inconel is vapor deposited on the flat surfaces of the disk. The resistance between electrodes is of the order of $10^9\text{ }\Omega$. MCPs not only respond to photoelectrons emitted from a photocathode, but can also directly detect X-ray and γ -ray photons. The electron gain of each MCP hole is 10^4 – 10^7 [Wiza, 1979].

As an example of a MCP, examine the Burle model 85104 MCP device. This MCP uses a semi-transparent GaAs photocathode, which responds over $350 \leq \lambda \leq 860\text{ nm}$. It uses a two stage microchannel plate structure, similar to that shown schematically in cross-section in Figure 6.59. The 85104 is physically small, the case is a cylinder 52.5 mm in

**FIGURE 6.59**

Cross-section (not to scale) of a micro-channel plate PM. Note biasing resistors.

length and 44.5 mm in diameter. The window is 18 mm in diameter. The maximum $V_B = -3,600$ V and the maximum average anode current is 300 nA. The peak cathode spectral sensitivity is $R_C(\lambda) = 120$ mA/W at 860 nm, the net electron gain, $G_I = 2 \times 10^6$. Thus $R_A(\lambda) = 2.5 \times 10^5$ A/W at 860 nm. The dark current, $I_D = 10$ nA, max. Due to the small length of the holes, typical electron transit time is ca. 150 ps, and the pulse spread is ca. 550 ps.

MCP applications include all the applications of dynode PMTs, including single photon detection when cooled with LN_2 , image intensifiers. MCPs are also used in the design of ultra high speed CRT oscilloscopes. The MCP device is placed in proximity focus with the phosphor screen of a CRT. This allows use of low beam currents and low acceleration potentials, necessary for high writing speeds. The latter are necessary in a high bandwidth (\sim several GHz) real time oscilloscope. MCP devices are also used for the direct detection of charged particles and energetic photons (X- and γ -rays). In this application, they compete with the Hamamatsu position sensitive, dynode PMTs used with scintillation crystal plates. MCP devices can also be gated by placing a gate grid between the photocathode and the outer channel plate. Holding the grid negative repels photoelectrons and effectively turns the MCP device off.

6.8.5 Summary

In this section, we have described the salient characteristics of the three classes of photomultiplier sensors—PMTs with dynodes, the single channel PMT (CPM) and microchannel plate PMTs (MCPs).

All three are capable of counting single photons, equivalent to incident light levels on the order of 10^{-17} W. Resolution was seen to be enhanced by reducing device shot noise. Thermionic emission white noise from the cathode and Johnson noise from resistors are reduced by cooling the sensor with LN₂ or thermo-electrically.

Due to their extreme sensitivity, all types of PMTs can be damaged by excessive input light power. Manufacturers generally specify a maximum anode current for an active device. Exposure to high light intensity can also compromise a photomultiplier's sensitivity, even when it is not energized, however, the tube will recover if given 'rest' in the dark [Lion, 1959].

Hamamatsu [PMT Handbook, available online] lists and illustrates sixteen applications of PMTs. These include:

1. Spectrophotometry
2. Gamma cameras (used in nuclear medicine)
3. Positron Emission Tomography (PET)
4. *In vitro* radioassay
5. Liquid scintillation counting systems
6. Biotechnology (cell sorters/counters, fluorometers, DNA sequencers)
7. High energy physics experiments
8. Oil well logging
9. Mass spectrometry/Solid surface analysis
10. Environmental measurements
11. Applications of laser measurements (ladar, side-looking ladar, Doppler laser velocimetry)
12. Plasma applications
13. Color scanners
14. Industrial measurements applications
15. Space research applications
16. Low light measurements (Raman spectrophotometry)

To these, we can add single and multiple bubble sonoluminescence and surface plasmon resonance.

6.9 Ionizing Radiation Sensors

Ionizing radiation can be either high energy, electromagnetic radiation, which includes photons, gamma radiation and X-rays, or can be a directed beam of high energy, subatomic particles such as beta particles (electrons), alpha particles (helium nuclei, less the two orbital electrons) or certain ions, moving at high velocity. Ionizing radiations are given that name because, as they pass through various media which absorb their energy, additional ions, photons, or free radicals are created. In a biological context, such ions or free radicals can, under certain circumstances, lead to genetic damage and mutations. On the other hand, the production of ions, electrons or photons under controlled circumstances can lead to a means of counting the incident ionizing particles and determining their energy. We describe below several designs for radiation sensors,

including the familiar gas filled Geiger Muller tube, solid state crystal detectors, and scintillation counters. The basic unit of radioactivity is the curie, defined as 3.7×10^{10} disintegrating nuclei per second. The unit of exposure is the roentgen, defined as producing 2.58×10^{-4} coulombs of free charge per kilogram of dry air at standard temperature pressure. The biologically absorbed dose is the rad, defined as 100 ergs of absorbed radiation per gram of absorber. The rem is an equivalent dose of radiation that has the same 'effect on a man' as does one rad of 200 kV X-rays.

6.9.1 Geiger-Muller Tube

A basic Geiger-Muller tube (GMT) with coaxial symmetry is shown in Figure 6.60A. The outer wall of the tube is generally a thin metal such as aluminum. The outer wall is the negative electrode, or cathode, of the GMT. The anode is a thin wire of tungsten or platinum, run axially at the center of the tube and is, of course, insulated electrically from the cathode. The other major type of GMT is the end-window design, shown in Figure 6.60B. The end window is a thin sheet of mica or mylar and serves to admit particles which have low penetration, such as α and β particles (although α particles have very high energies, in the range of 4–10 MeV, their penetration in air is relatively short—a 3 MeV α particle travels about 1.7 cm in air, and α particles are stopped by a sheet of paper or aluminum foil! β particles (electrons) also arise from the decay of certain natural and manmade radioisotopes. An electron with 3 MeV energy will travel about 13 m in air before being stopped, and may penetrate as much as 1 mm of aluminium).

Geiger-Muller tubes are generally filled with an inert gas such as helium, argon, or nitrogen, at pressures ranging between a few mm Hg and atmospheric. Most modern GMTs are self-quenching (i.e. they also contain a polyatomic gas such as methane or alcohol vapor which prevents energetic photons emitted from the activated inert gas molecules from reaching the cathode and stimulating the release of secondary electrons by photoelectric emission). The polyatomic gas is either dissociated or decomposed by the high energy photons in the process of absorbing their energy. Since the polyatomic gas is decomposed, the self-quenching GMT gradually loses its quenching

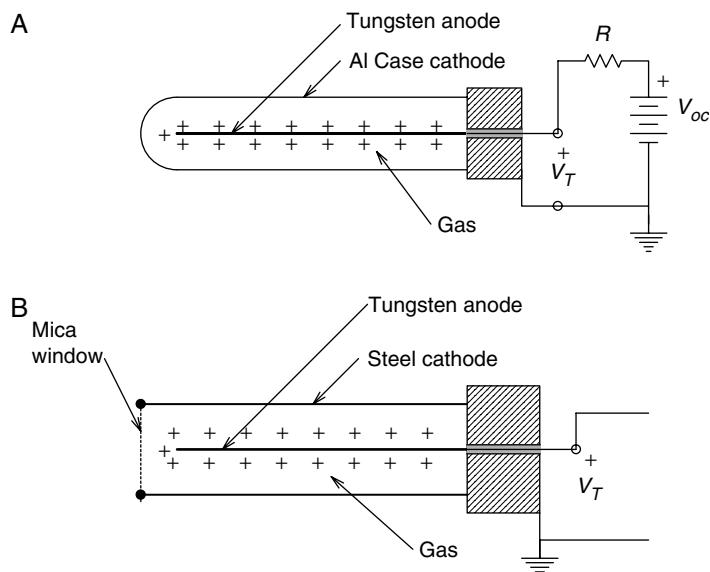


FIGURE 6.60

A. Conventional, coaxial, Geiger-Muller radiation counting tube (GMT). The thin aluminum shell passes most ionizing radiation particles except alpha and low energy electrons. B. A GMT with thin mica end window to enable counting of alpha particles.

ability with its total number of counts. The effective life of a GMT may be as high as 10^{10} counts. Methane, as a quenching gas, gives a shorter count life (10^8) than does alcohol vapor.

The dc potential difference across a GMT determines its properties as a radiation counter. Lion (1959) describes five distinct ranges of applied voltage to the GMT. The voltage ranges given are examples; every type of GMT will have numerical values peculiar to its own design. The *first region* (100–400 V) is called the ionization chamber region. In this region, an ionizing particle entering the GMT causes the generation of one or more electron-ion pairs. These charged particles drift toward the electrodes in the E -field of the tube. If the field is high enough so that they reach their destinations before an appreciable number of them recombine, and not so strong that additional ionization by collision occurs, the number of charges moving to and arriving at the electrodes is equal to the number initially produced by the collision(s) of the ionizing particle with the inert gas molecule(s). Each surge of ions from an ionizing particle collision thus causes a very small current pulse, the time integral of which is proportional to the number of ions, n_j , produced by the j th collision.

The *second region* (400–700 V) is called the proportionality region. Due to the higher electric field strength, ions and electrons produced in the GMT by an energetic particle are accelerated to higher kinetic energies, and thus cause the generation of new ions and electrons by inelastic collisions. This process lacks the energy to become self-sustaining, but a current pulse produced has an area proportional to kn_j . The multiplication factor, k , depends on many factors, including the type and pressure of inert gas used. The value of k can range 10 – 10^4 . Since the current pulses caused by the collisions of ionizing particles with the inert gas in the GMT are larger, proportional operation of the GMT allows the measurement of n_j and thus, the energy of the ionizing particles.

The *third region* (700–1000 V) is the region of limited proportionality. In this region, the sizes of the pulses are no longer proportional to the initial number of ions formed by the collision and hence, not proportional to ionizing particle energy. This region has little practical use.

The *fourth region* (1000–1300 V) is the well-known Geiger counter region. Here, a collision of an ionizing particle with an inert gas molecule in the GMT causes output pulses that are all the same size, independent of n_j and the particle's energy. The mechanisms by which this mode of operation occurs have been described by Lion (1959). The initial positive ion and electron formed in the collision are accelerated in the strong E field and gain sufficient kinetic energy to cause further ionizations through collisions with the inert gas molecules. The additional ions and electrons so produced form an 'avalanche' of charge. Photons are emitted from the excited gas atoms in the avalanche, and the discharge forms along the center wire anode, where the field strength is highest.

According to Lion, "The light [photon] emission proceeds in every direction, but only in the vicinity of the central wire is the field strength high enough to facilitate the formation of new avalanches. The discharge spreads along the wire, therefore, until the entire anode is surrounded by a narrow cylinder of ions, and the discharge becomes self-sustaining."

Lion continues, "The termination of the discharge comes about in the following manner. The positive ions surrounding the central anode move toward the cathode with a velocity which is considerably smaller (because of their mass) than that of the electrons. These ions cause a positive space charge surrounding the positive wire; as they move toward the cathode, the positive space charge cylinder ('the virtual anode') increases in diameter and, therefore, the field strength in the counter decreases."

The recovery time in GMTs following a 'count' ranges 30–300 μs [Lion, 1959], thus the maximum counting rate can be in the range 3,000–30,000 counts per second. The action of the quenching gas is described above.

The *fifth region* ($> 1300 \text{ V}$) is called the unstable region. Here, the field strengths are so large that the quenching gas is ineffective and one ionizing particle collision with an inert gas molecule can lead to multiple pulses, or in extreme cases, a sustained plasma discharge, which may destroy the GMT.

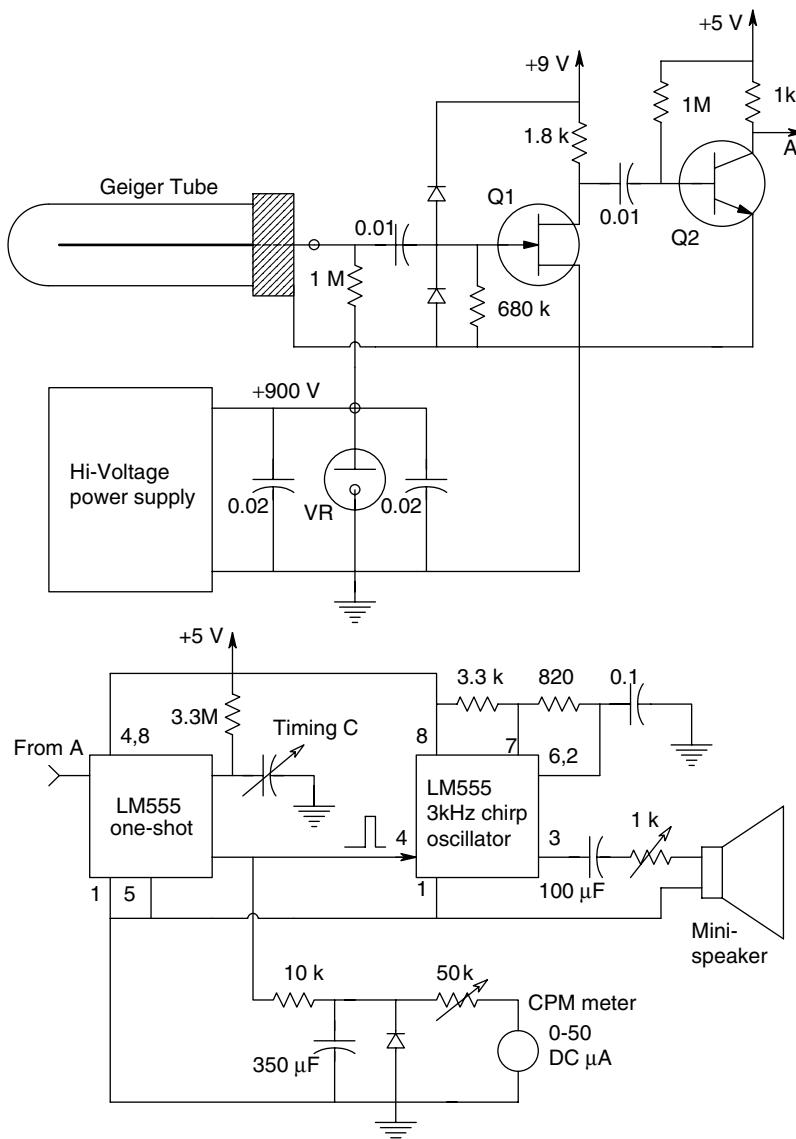


FIGURE 6.61

Circuit of a simple, analog, radiation (Geiger) counter. Voltage pulses from the GMT are conditioned by transistors Q1 and Q2, and used to trigger a LM555L-1 which acts as a one shot multivibrator to produce pulses of standard width and height. The output pulses from LM555-1 are averaged by a low-pass filter preceding a dc microammeter. The dc current in the microammeter is proportional to the average number of GMT pulses per minute. The output pulses from the LM555L-1 also gate a second LM555L, acting as a chirp oscillator. The chirp is sent to headphones or a small loudspeaker to give the operator a subjective reading.

It should be stressed that the current pulses from a GMT operated in the Geiger or plateau region are about 1000 times the size of the pulses observed when the GMT is operated in the proportionality region. The circuit of a simple, portable Geiger counter is shown in Figure 6.61. Several output modes are commonly used—an audio output through headphones or a loudspeaker, an analog meter output and a digital counter output reading in counts per minute, or other time unit. A GMT counter system does not provide information as to what the ionizing particles or radiation is. This information must be inferred from an *a priori* knowledge of the source of radiation.

6.9.2 Solid State Crystal Radiation Sensors

Crystal radiation sensors have used crystals of materials such as cadmium telluride lightly doped with chlorine (a *p*-semiconductor), thallium chloride-thallium bromide, silver chloride, silver bromide and diamond on which metallic electrodes have been deposited on opposite surfaces. A dc potential is applied to the electrodes, generating an internal electric field in the crystal of several kV/cm. If an ionizing particle enters the crystal and collides with its atoms, a burst of free electrons is released, which then drifts through the crystal to the positive electrode where its arrival causes a brief current pulse. Only 1.0 eV is required in a Cl doped CdTe crystal to put an electron into the conduction band.

Let us assume the total electron charge released by the ionizing event is n_0q coulombs (n_0 is the number of electrons released, q is the charge on one electron) and these electrons are created near the cathode and drift *en masse* over the thickness, d , of the crystal to the positive electrode. In their travel to the anode, some of the electrons recombine with positive charges (holes), or are trapped by impurities in the crystal such that the charge reaching the anode is given by:

$$Q_A = n_0q \exp(-d/\lambda) \quad (6.138)$$

Thus, n_0q coulombs start to move at some velocity, v , toward the anode, decreasing in number as they travel. Since the applied electric field in the crystal is high (about 5 kV/cm) and d is small, the travel time of the electrons, $T_d = d/v$, is small compared with the RC time constant of the external circuit in Figure 6.62. Note that the collecting (right

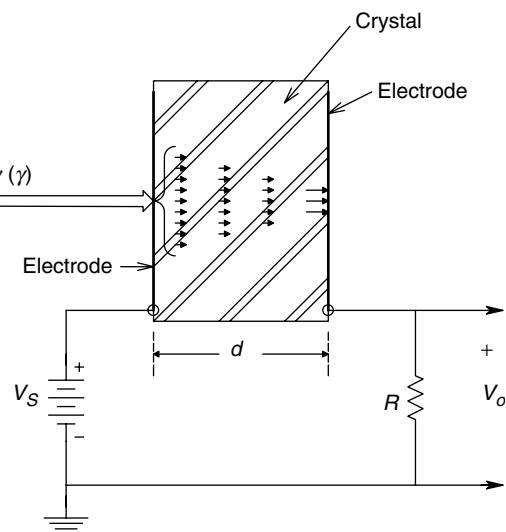


FIGURE 6.62

Schematic of a solid state, crystal radiation detector.

hand) electrode is a solid disk, but the thin metal film anode (left hand) electrode is annular, providing a center hole through which the photons strike the crystal. The magnitude of the charge bundle in the crystal may be found as a function of time:

$$\begin{aligned} Q(t) &= n_o q \exp(-tv/\lambda) \quad \text{for } 0 \leq t \leq T_d \\ Q(t) &= 0 \quad \text{for } t > T_d \end{aligned} \tag{6.139}$$

We note that the peak voltage on the capacitor formed by the crystal and its electrodes is given by $\bar{V}_c = \bar{Q}/C$. Q is the average charge flowing over the interval, $[0, T_d]$. Integrating equation 6.139 from 0 to T_d and dividing by CT_d yields the peak value of the voltage induced by the moving electrons:

$$V_c = (n_o q \lambda / Cd) [1 - \exp(-d/\lambda)] \tag{6.140}$$

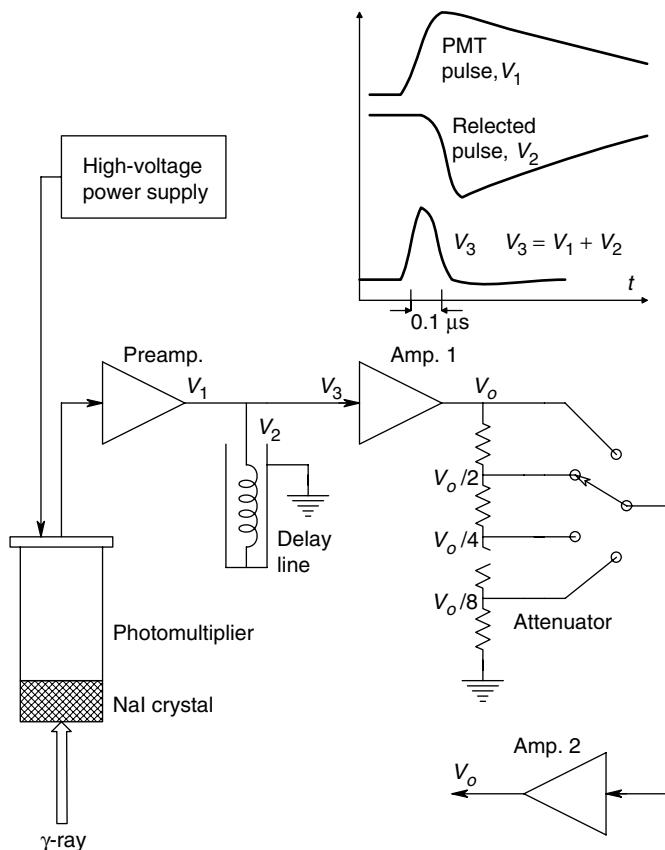
V_c decays exponentially as $v_c(t) = V_c e^{-t/\lambda C}$. In the equations above, the constant, λ , is known as the *Schubweg* of the crystal. This is the space constant for electron decay in the crystal, as seen from equation 6.138 above. The value of the Schubweg depends on the crystal material, and is proportional to the applied field strength. Lion (1959) states that the crystal will behave as a good proportional counter when the Schubweg lies between a few tenths of a millimeter and one centimeter. He states, "If the incident particle penetrates deeply into the crystal and causes the release of electrons uniformly throughout the crystal volume,..." The peak voltage is then given by:

$$V_c = (n_o q \lambda / Cd) [1 - (d/\lambda)(1 - e^{-d/\lambda})] \tag{6.141}$$

As in the case of GMTs, there is a critical range of applied dc voltage, where the pulse size is proportional to the energy of the ionizing particle. The particle energy for single ion pair production in a crystal is of the order of 5–10 eV. Lion states that a 1 MeV beta particle causes the release of about 1.2×10^5 electrons in a AgCl crystal. The lowest detectable energy for an incident particle is in the order of 1 keV, produced from 100–200 electrons. With a 5 kV/cm internal dc field, an AgCl crystal can resolve about 10^6 cps and a diamond can count about 10^7 events/second. According to Lion, space charges may build up in the crystal due to the accumulation of immobile holes in the crystal lattice, or from electrons trapped near the anode. Such space charges may seriously limit the counting life of a crystal, but their effect may be overcome by irradiating the crystal with infrared light, or periodically reversing the polarity of the dc applied to the crystal.

6.9.3 Scintillation Counters

A scintillation counter has three main components—a scintillation crystal, a photomultiplier tube which is used to count flashes (scintillations) from the crystal and an electronic pulse forming and pulse height discriminating circuit. The scintillation crystal is generally sodium iodide, about 0.5 in thick. The energy from ionizing radiation which collides with the crystal lattice is converted, in part, to a flash of light (secondary photons), the intensity of which is proportional to the energy of the absorbed radiation. The intensity of the flashes, or scintillations, are very weak and must be converted to electrical pulses by a PMT. The PMTs used in scintillation counters have flat, transparent photocathodes. Photons from the scintillation crystal strike the PMT's photocathode

**FIGURE 6.63**

Schematic of a sodium iodide based, scintillation radiation detector used in nuclear medicine. The energy of a light flash from the NaI crystal which results from the capture of a gamma particle is proportional to that particle's energy. Light flash energy is converted to voltage pulses by a fast photomultiplier tube and associated pulse sharpening circuitry. Electrical pulses are scaled and counted to isolate the activity of a certain radioisotope.

and cause the release of a number of electrons, proportional to the intensity if the flash, the intensity of which is proportional to the energy of the ionizing particle. The few photo-electrons released at the photocathode are amplified by the dynode electrodes of the PMT to form a large output current pulse. The current amplification of a PMT (anode current/photocathode current) depends on the supply voltage used, the tube's geometry, etc., and can range $10^2\text{--}5 \times 10^6$ [Lion, 1959]. Overall photosensitivity of PMTs can range $0.01\text{--}2 \times 10^4$ A/lumen. The PMT current amplification increases proportional to the square root of the supply voltage. With liquid nitrogen cooling to reduce noise, PMTs can resolve scintillations as small as 2×10^{-16} lumen. The response time of a PMT is, in general, <10 ns, hence such tubes should be capable of counting as fast as 10^7 cps. The output pulses from PMT circuits are often reshaped to make them sharper. One such sharpening circuit, taken from Webster (1978), uses a shorted delay line to cancel the long tail on the primary output pulse, making the pulse narrow (Figure 6.63).

Scintillation systems are widely used in nuclear medicine in conjunction with γ -emitting isotopes to image tissues with cancerous growth. The patient swallows, or is injected with, a small amount of an isotope which is selectively taken up metabolically by the organ in question. The presence of a tumor generally means that more of the radioisotope will be taken up than by normal tissue. A scintillation system with a lead collimator is used to scan over the tissue in question and to detect areas of abnormal (high) radioactivity [Northrop, 2002]. An example of such a system is the use of the isotope ^{131}I , which has a half-life of 8.1 days. ^{131}I is selectively taken up by cells in the thyroid gland in the process of the formation of the hormone thyroxin. Planar arrays of scintillation crystals and their associated PMTs are called gamma cameras and are

used in nuclear medicine to visualize parts of the body that have selectively absorbed a radioisotope. This type of gamma camera provides relatively coarse resolution because of the physical size of the scintillation crystals and PMTs. In the Anger camera, a single, large scintillation crystal is used with an array of 19 PMTs. The location of a single radioactive event in the crystal is estimated by a formula involving the intensity of the flash produced at each PMT and the location of each PMT. Another configuration of scintillation camera used in nuclear medicine is the pinhole collimator. In this mode of operation, pixel resolution is traded off for sensitivity [Macovski, 1983].

6.10 Electro-Chemical Sensors

This class of sensor is generally used in a ‘wet’ environment (i.e. one in which the substance to be measured is dissolved in water or another solvent). The outputs of such sensors are generally currents or voltages which are functions of the concentration of the substance to be measured. Electrochemical sensors may be subdivided into three categories—those which generate EMFs as the result of the sum of two half-cell potentials, (i.e. they are electrochemical cells), those which operate polarographically and those which are basically fuel (redox) cells. All three categories involve electrochemical reactions.

6.10.1 pH and Specific Ion Electrodes

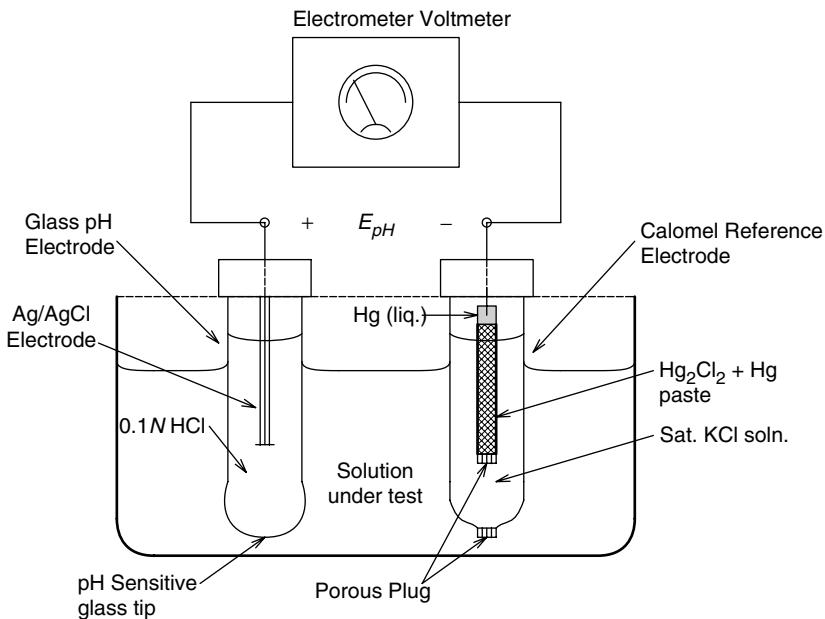
The traditional pH measurement system is composed of two half-cells forming a ‘battery’ whose EMF is a function of pH. pH is defined as the logarithm to the base 10 of the reciprocal of the hydrogen ion concentration. In terms of an equation:

$$\text{pH} \equiv \log_{10}(1/\text{[H}^+]) \quad (6.142)$$

A traditional pH cell is shown in Figure 6.64. The pH half-cell electrode itself is made from a special glass membrane, across which the EMF is a function of hydrogen ion concentration. The pH electrode is filled with 0.1 N HCl and electrical contact is made internally through a silver chloride coated silver wire (Ag/AgCl). The resistance of the glass pH electrode is quite high, around $50\text{ M}\Omega$ or more. This limits the current which may be drawn from the pH cell, so that an electrometer amplifier interface must be used to measure the cell’s EMF without IR drop errors. The reference half-cell electrode is traditionally a saturated calomel design, shown on the right in Figure 6.64. Undissolved potassium chloride crystals saturate the internal solution of the reference half-cell. The total pH cell’s EMF is given by summing the half-cell EMFs:

$$\begin{aligned} E_{GL} &= E_0 GL - (RT/\mathfrak{F}) \ln(a_{H+}) \\ &= E_{0GL} + (2.3026RT/\mathfrak{F})(\text{pH}) \end{aligned} \quad (6.143a)$$

$$E_{CAL} = E_{0CAL} - 7.6 \times 10^{-4}(t - 25^\circ\text{C}) \quad (6.143b)$$

**FIGURE 6.64**

A basic pH measuring system. The pH EMF is generated between a pH sensitive glass electrode and a calomel reference electrode whose half-cell potential is independent of pH.

$$\begin{aligned}
 E_{NET} &= E_{GL} - E_{CAL} = (E_{0GL} - 0.2415) \\
 &\quad + 7.6 \times 10^{-4}(t - 25) + (2.3026 RT/\mathfrak{F})(\text{pH})
 \end{aligned} \tag{6.143c}$$

where a_{H+} is the activity of the hydrogen ions in solution (activity may be considered to be equal to concentration at low concentrations). E_{GL} is the glass half cell electrode's EMF, E_{CAL} is the calomel electrode's half-cell EMF, R is the MKS gas constant (8.3143 J/K), T is the temperature in kelvin, t is the Celcius temperature, \mathfrak{F} is the Faraday number (96,500), E_{0CAL} is the standard potential for the calomel half-cell (0.2415 V) and E_{0GL} is the standard potential for the glass electrode. E_{0GL} varies between glass electrodes and is compensated for when the individual pH measuring system is calibrated. Note that acidic solutions have pHs ranging 0–7.0, while basic solutions have pHs ranging 7.0–14. Also note that the pH potential, E_{NET} , is highly temperature dependent and therefore, all pH meters use a temperature probe of some sort, also immersed in the solution under measurement. The signal from the temperature probe is used to null the $(t - 25)$ term in the pH cell EMF, E_{NET} , and to correct for changes in the $(2.3026 RT/\mathfrak{F})$ term if the t of the measurement solution differs from the temperature of the standard pH buffer solutions used to calibrate the pH meter. It should be noted that there are many physical variations on the standard glass pH electrode and the calomel reference electrode. Designs exist which have been optimized for the measurement of blood pH, the pH of high temperature solutions, pH measurements in the stomach, etc. In addition to the common glass pH electrodes, under limited conditions, other half-cells can be used whose half-cell potentials are proportional to pH. Tungsten metal, quinhydrone and palladium oxide electrodes are responsive to pH [Maron and Prutton, 1958; Grubb and King, 1980].

Specific ion electrodes are used in conjunction with a reference electrode to measure the logarithm of the activity (concentration) of a specific anion or cation in solution.

Specific ion electrodes are tempermental, in that most have ranges of pH for optimum operation and they may also require the complete absence of certain ion(s) which compete in the electrode's half-cell chemical reaction and therefore, cause errors. We mention a few representative, specific ion electrodes and their properties below.

The Orion model 94-17 Chloride Electrode has a log linear half-cell potential which increases by 59.2 mV/decade change in concentration over a 10^{-4} to 1 M range. This electrode tolerates a pH range of 0–14, as well as nitrate, sulfate, phosphate, fluoride and bicarbonate ions. It will not function in strongly reducing solutions, nor in the presence of sulphide, iodide, or cyanide ions. The chloride level must be at least 300 times the bromide level, 100 times the thiosulfate level and 8 times the ammonia level for proper operation.

The Orion model 94-06 cyanide ion electrode has a log linear half-cell potential over a 10^{-6} – 10^{-2} M range. The cyanide electrode does not respond to cations, except silver, and most common anions, including nitrate, fluoride and carbonate. The electrode will not work with sulfide ions in the solution. The chloride concentration may not be more than 10^6 times the cyanide concentration, the bromide level may not be more than 5000 times the cyanide level and the iodide level may not be more than 10 times the cyanide to prevent error from occurring in measuring the cyanide level.

The Orion model 92-19 potassium ion electrode produces a log linear, 59.6 mV/decade concentration change of K^+ ions over a 10^{-6} –1 M concentration range. It operates in pH 1–12. Selectivity constants, given as the ratio of the electrode's response to the interfering ion to the response to K^+ , are given by Orion as: Cs^+ (1.0), NH_4^+ (0.03), H^+ (0.01), Ag^+ (0.001), Na^+ (2×10^{-4}), Li^+ (10^{-4}). Thus we see that the potassium electrode is relatively insensitive to other metal ions, but has a 1:1 ambiguity with cesium ions and is affected at the 1% level by pH.

The concentration of cadmium ions may be measured with the Orion model 94-48 electrode. The EMF of this half-cell is log linear in the 10^{-7} – 10^{-1} M [Cd^{++}] range. The pH range is 1–14 and silver, mercury and copper must not be present in the solution. The level of free lead or ferric ions must not exceed the cadmium ion concentration to prevent error.

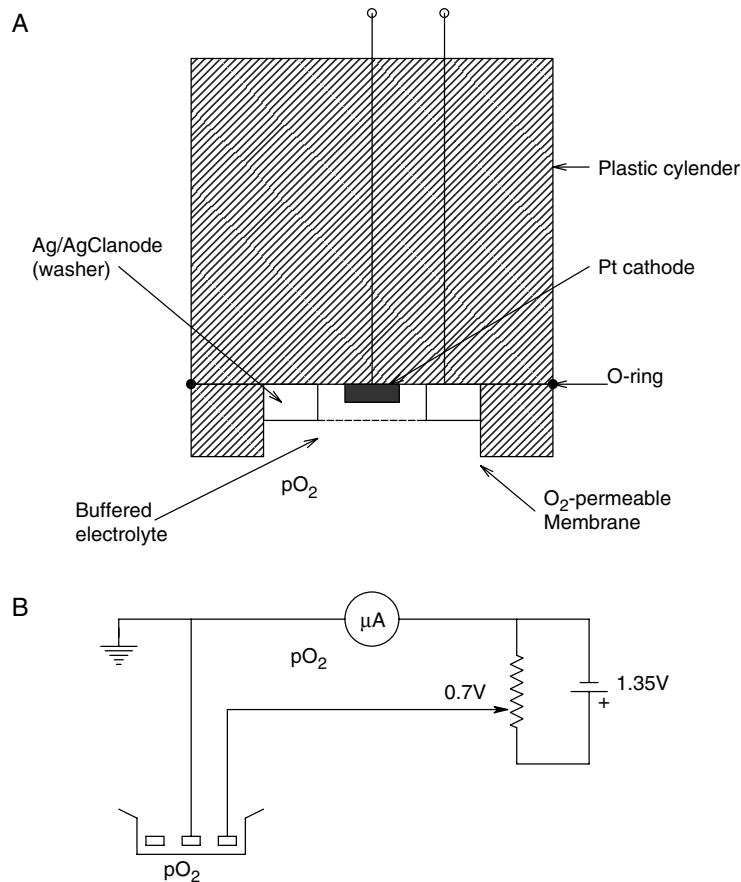
The Severinghaus pCO₂ electrode, described by Wheeler [Webster, 1978] makes use of the property that the pH of a buffer solution is proportional to the log[pCO₂] over the range of 10–90 mm Hg partial pressure. A teflon membrane separates the solution in which CO₂ gas is dissolved (blood) and the buffer solution. The membrane is permeable to CO₂ gas, but blocks the passage of ions such as H⁺ and HCO₃⁻. The pH of the Severinghaus cell may be shown to be equal to:

$$pH = \log[HCO_3^-] - \log(k) - \log(a) - \log[pCO_2] \quad (6.144)$$

where k is the equilibrium constant for CO₂ gas in water changing to bicarbonate ion and hydrogen ion, a is the constant relating the CO₂ gas pressure above blood to the concentration of CO₂ dissolved in blood, and [HCO₃⁻] is the concentration of bicarbonate ions in the buffer solution around the pH electrodes. The Severinghaus pCO₂ electrode requires calibration with two known partial pressures of CO₂ and, like any chemical cell, is affected by temperature.

6.10.2 Polarographic Electrodes

Polarographic electrodes differ from pH and ion specific cells in that a polarographic cell is run at a constant potential in order to force two electrochemical reactions to take place. The current which flows in the cell is determined by, and is proportional to, the

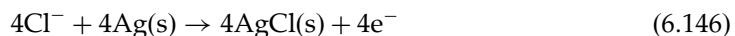
**FIGURE 6.65**

A. A cross-section through a Clark, oxygen sensing, polarographic electrode. B. Schematic of basic circuit used with a Clark O₂ probe. No temperature compensation is shown.

limiting concentration of one reactant. The *Clark Cell* for the measurement of the partial pressure of oxygen dissolved in water, blood, or of oxygen in the air, is shown schematically in Figure 6.65A. The Clark cell has four components—a platinum or gold cathode, an Ag/AgCl anode, an electrolyte solution and a plastic membrane (polypropylene, teflon or mylar) which is permeable to O₂ gas and little else. The electrolyte is generally a buffered, saturated KCl solution. At the noble metal cathode, a reduction reaction takes place, which is described as:



At the Ag/AgCl anode, the oxidation reaction given below takes place:



It is found experimentally that, when the anode is held at 0.7 V positive with respect to the cathode, a current flows, which is described by the relation:

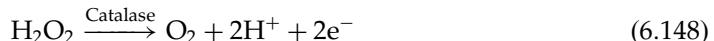
$$I = I_o + \beta[\text{pO}_2] \quad (6.147)$$

β is typically about 10 nA/mm Hg pO₂. Both I_o and β are increasing functions of temperature. The response time of the Clark electrode to step changes in pO₂ is limited by the ease of O₂ diffusion through the membrane to the cathode surface. Typically in the tens of seconds, the response time of a fast Clark O₂ electrode was been cited by Geddes and Baker (1968) as 0.3 ms. According to Wheeler [Webster, 1978], the settling time for a Clark cell is proportional to the pO₂ being measured.

The Clark cell has two major sources of error in measuring pO₂. One is the effect of temperature. The Clark cell must be calibrated and operated at a controlled, fixed, temperature which varies no more than $\pm 0.1^\circ\text{C}$, or, knowing its tempco, we can measure its temperature and calculate compensation terms for equation 6.102 above.

6.10.3 Fuel Cell Electrodes

A fuel cell is a battery (cell) in which the chemical reactions producing the EMF and output current occur at one or both half-cell electrodes and consume two or more reactants, which are broken down to reaction products. If one of the reactants is present in excess concentration and the other is present at a reduced concentration, then the cell's EMF may be proportional to the logarithm of the activity of the limiting reactant. A specific example of a fuel cell system is the catalyzed oxidation of glucose to form gluconic acid, hydrogen peroxide and heat. A glucose fuel cell described by Wingard *et al* (1982), used a mixture of the bioenzymes glucose oxidase and catalase immobilized on the surface of a platinum screen electrode. Glucose oxidase catalyzes the oxidation of glucose to gluconic acid and H₂O₂. Catalase catalyzes the breakdown of H₂O₂ according to the reaction:



The authors state that the source of the cell's EMF is the hydrogen peroxide reaction, although firm evidence for this contention was not given. They speculate that the reaction at the platinum anode is the oxidation of H₂O₂, and the cathodic half-cell reaction is the reduction of AgCl to Ag. Wingard *et al* showed that the open circuit EMF of their fuel cell was proportional to the logarithm of the glucose concentration from 5–500 mg/dl. Another type of glucose fuel cell is described in Chapter 8 of Wise (1989).

It should be pointed out that there are many other electrochemical cells which have been devised to measure glucose. Some use polarographic methods in which the current through the cell is a function of the glucose concentration, some measure the decrement of pO₂ across a membrane caused by the oxidation with a Clark O₂ cell, others measure the temperature rise in a chamber resulting from the oxidation. The interested reader should see the review of glucose sensors by Peura and Mendelson (1984).

6.11 Electronic 'Noses'

6.11.1 The Rationale for Artificial Noses

The need for a reliable, portable instrument that will sense contraband drugs, explosives of all sorts, including buried land mines, organic pollutants in soil, water and air

pollutants, is a driving force for the development of a measurement system that will detect and measure the minute concentration of certain volatile organic compound (VOC) analytes in the air. Such an electronic nose can also be used for the grim purpose of locating buried corpses after natural disasters or bombings, and evaluate the freshness and quality of meat, fish, fruit, dairy products, vegetables and beverages. At present, the best means we have for many of these tasks are trained dogs. Dogs get tired, have to eat and sleep and have some 'down time' between their searches. An electronic nose is available 24/7. It requires calibration, however, just as a dog requires careful training to respond to specific scents. Note that many less portable or fixed technologies exist for VOC analysis, particularly in airports. These include gas chromatographs, mass spectrometers and spectrophotometers. Even these instruments have been reduced in size so that they can be brought into the field. However, they generally are not hand portable.

In the sections below, we will describe the anatomy and physiology of vertebrate olfaction, the state-of the art in manmade VOC sensors and the overall design of the electronic nose using artificial neural networks.

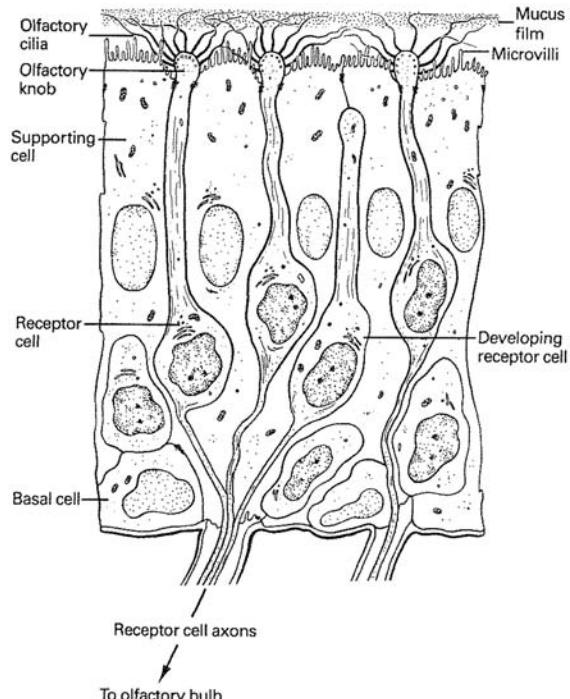
6.11.2 Olfactory Anatomy and Physiology

Smell is one of the oldest and most primitive senses in vertebrates. The olfactory system is designed to detect gas phase odorant molecules. Matter is generally in one of three phases, depending on its temperature and pressure—solid, liquid or gas (vapor). Olfactory systems operate on molecules in the gas phase (i.e. that are airborne). Early in vertebrate evolution, olfactory systems developed as a means of finding a mate (pheromones and sex attractants), food, water, and avoiding predators, danger and intraspecies conflict.

The sensing and identification of complex odorants involves an animal's entire olfactory system which is generally composed of many millions of olfactory sensory cells. Each olfactory cell has cilia, with surface protein receptors which bind specifically to a unique odorant molecule. The binding generally triggers a complex, internal sequence of biochemical events that lead to depolarization of the olfactory cell and spike generation on its axon. The spike frequency increases monotonically with the odorant concentration.

The olfactory systems of vertebrates can generally be considered to be massively parallel arrays of chemosensory neurons organized in a mucous covered, olfactory epithelium open to nasal air. The area of the olfactory epithelium in humans is ca. 5 cm^2 in humans, ca. 25 cm^2 in cats and $75\text{--}150\text{ cm}^2$ in dogs. About 5% of the human cortex is estimated to process olfactory sensory information; in dogs, ca. 33% of the cortex is devoted to this operation.

In humans, the ca. 5×10^7 olfactory sensory cell axons project through the cribiform plate on the roof of the nasal cavity into the olfactory bulb in the brain, which in turn sends interneurons to the amygdala. Figure 6.66 illustrates the anatomy of the olfactory epithelium schematically in a sagittal section. Comparative anatomy teaches us that behavior governed by the sense of smell evolved long before mammals evolved cortices and the ability to reason. Airborne molecules evoke complex behavioral patterns, ranging from 'keep out' (animals mark territory with urine, feces, scent glands), to sexual receptivity (pheromones), and a food source is near (blood, amniotic fluid, etc.) Scent can also alter mood (infant smells mother, leading to nursing behavior and the release of endorphins), or elicits fear (animal smells predator) leading to the release of adrenaline. Danger may also be signaled by the odor of spoiled food. Many animals have specialized glands they use for territorial marking, usually in conjunction with

**FIGURE 6.66**

Schematic section through the vertebrate olfactory epithelium, found in the nasal cavity. The mature sensor cells have olfactory cilia. (Source: With permission, Kandel *et al.*, 3rd ed., 1991.)

urine. Odorants have been extracted from the feces of predators such as African lions and marketed for use in keeping deer out of gardens. The study of odorant molecules that affect animal behavior is called semio-chemistry [Albone, 1997].

Much research today is being directed towards understanding how certain animal olfactory systems can detect (as evidenced by behavior) nanomole concentrations of certain odorants. The ability of bloodhounds to track a specific person through a crowd, or follow a days-old trail when searching for a lost person or fugitive is legendary. The noses of trained dogs are also the most effective system known for locating buried land mines, explosives, or drugs in luggage and vehicles. Presumably, the dogs are responding to the vapor phase of the analyte (e.g. C4, TNT, cocaine, etc), or its component impurities. Dog threshold olfactory sensitivity to substances they have been trained to detect (e.g. cocaine, semtex, C4, etc.) has been estimated to be on the order of tens of ppb ($1/10^8$), down to ca. 500 ppt ($5/10^{10}$) [Waggoner *et al.*, 1997].

Due to the commercial importance of odors in the food and cosmetic industries, and the need to detect drugs and explosives, much research is being invested in 'artificial (or electronic) noses', using a variety of physical, immunochemical, electronic and photonic means. It appears that man has a long way to go to make a sensor with the threshold sensitivity of the bloodhound's nose or certain insect pheromonal chemoreceptors.

6.11.3 The Vertebrate Olfactory Chemoreceptor System

Vertebrate chemoreceptor neurons are unique in that they undergo apoptosis (programmed self-destruction) about every 40–60 days and are replaced by new ones. The new cells send their axons into the olfactory bulb (first synaptic relay point for olfactory information) where they synapse with appropriate target mitral cells. The mitral cells, like all other neurons in the CNS, do not divide (Figure 6.67). Each olfactory cell has

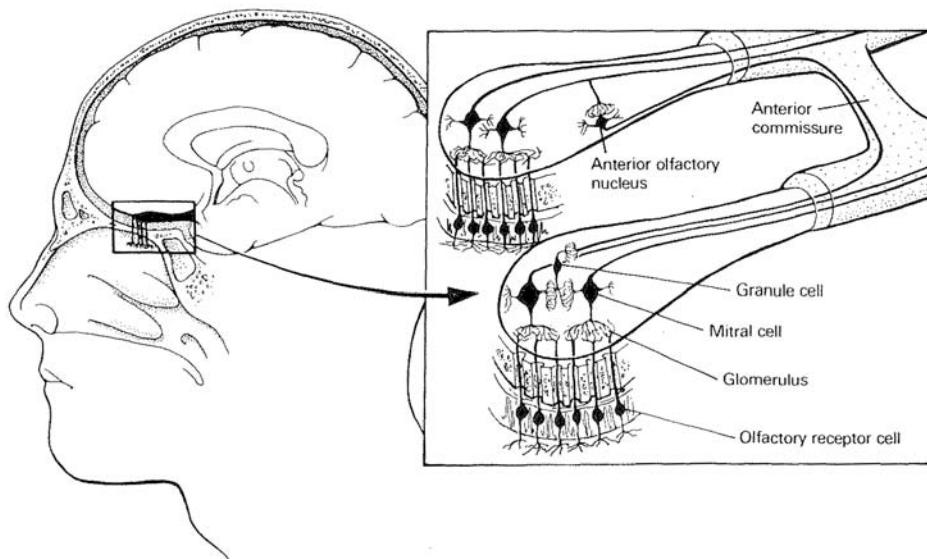
**FIGURE 6.67**

Diagram showing how human olfactory neuron axons project through the bone cribiform plate into the interneurons of the olfactory bulbs. Mitral cells in the olfactory bulbs project axons into the brain. (Source: With permission, Kandel *et al*, 3rd ed., 1991.)

8–20 cilia at its end immersed in a mucous layer, which is about $60\text{ }\mu\text{m}$ thick. The cilia are $30\text{--}200\text{ }\mu\text{m}$ in length and their surfaces contain the odorant receptor molecules that initiate receptor spike generation. A mouse may have 500–1,000 different odorant, G-protein coupled receptor proteins [Leffingwell, 2001; Eisthen, 2002]. It is estimated that about 1% of the total genes specifying an entire mammal (mouse) are used to code its olfactory receptor proteins.

Airborne odorant molecules must dissolve in the mucus before they can reach receptor molecules on the olfactory cilia. Also floating in the mucus are odorant binding proteins (OBPs) known as lipocalins which have been postulated to act as carriers that transfer odorants to their specific receptors on the cilia. The OBPs may act as cofactors that facilitate odorant binding and they may also participate in the destruction of a bound odorant, freeing the receptor to bind again. OBPs are made by the lateral nasal gland at the tip of the nasal cavity [Kandel *et al*, Chapter 34, 1991]. The exact role of the OBPs is unknown.

A putative mechanism of olfactory transduction is as follows. An odorant molecule dissolves in the mucous layer overlying the receptor cells and their cilia. It combines with a specific OBP and then collides with and binds to a specific surface membrane receptor protein on a cilium that has a high affinity for that odorant. The presence of the OBP may be necessary to trigger the next step, which is the activation of a second messenger biochemical pathway that leads to a specific ionic current and receptor depolarization. Odorant binding to the receptor protein activates a G-protein on the intracellular surface of the cilium. The α -subunit of the G-protein activates a molecule of the enzyme, type III adenylate cyclase, which in turn catalyzes the formation of a cyclic nucleic acid, cyclic-3',5'-adenosyl monophosphate (cAMP) from ATP. cAMP then binds to and opens a cAMP-gated Ca^{++} channel, causing an inward $J_{\text{Ca}^{++}}$. The increase in intracellular Ca^{++} appears to activate calcium dependent chloride channels, producing an outward chloride current, $J_{\text{Cl}^{-}}$, that depolarizes the receptor [Leffingwell, 1999]. The more of a given odorant that binds to a specific chemoreceptor cell, the larger

the total $J_{\text{Ca}^{++}}$ and J_{Cl^-} , and the greater the depolarization, and the higher the spike frequency on the receptor's axon. cAMP then degrades by hydrolysis to AMP, and the G-protein also returns to its resting state.

The receptors on a given chemoreceptor's cilia can have affinity to more than one odorant molecule. Also, more than one receptor can be activated (in varying degrees) by a given odorant. Hence the very large chemoreceptor array must have a combinational logic mechanism for eliminating odor ambiguity, since several receptors may respond to a given odorant in different degrees. Note that perception of certain scents can be the result of different parts of the same odor molecule fitting into different receptors [Leffingwell, 2001a]. (Again, Nature dazzles us with her complexity.)

How are specific odorant receptors recycled? A problem in any sensory communication system is to preserve sensitivity to a new stimulus. We know that when an odorant binds to its receptor, the second messenger cascade is initiated. It has been recently discovered that the second messenger cascade is inhibited by the second messengers activating a protein kinase, which phosphorylates the receptor protein, interrupting the transduction process [Breer, 1997]. It is also reasonable to assume that the odorants are slowly broken down enzymatically, freeing up the receptors for new inputs. Olfactory systems are extremely sensitive, but (perhaps mercifully) adapt profoundly to a sustained, external odorant concentration, to the point where one is not conscious of the odor.

Bundles of 10–100 axons from the ciliated olfactory chemoreceptors project through the pores in the ethmoidal cribriform plate on the roof of the nasal cavity into the olfactory bulbs of the brain, where the first level of signal processing takes place. In structure, the olfactory bulb has several layers of interneurons, not unlike a retina. Figure 6.68 illustrates schematically the basic, 5-layered structure of a mammalian olfactory lobe. Not shown is the tremendous convergence of information in the olfactory bulb. For

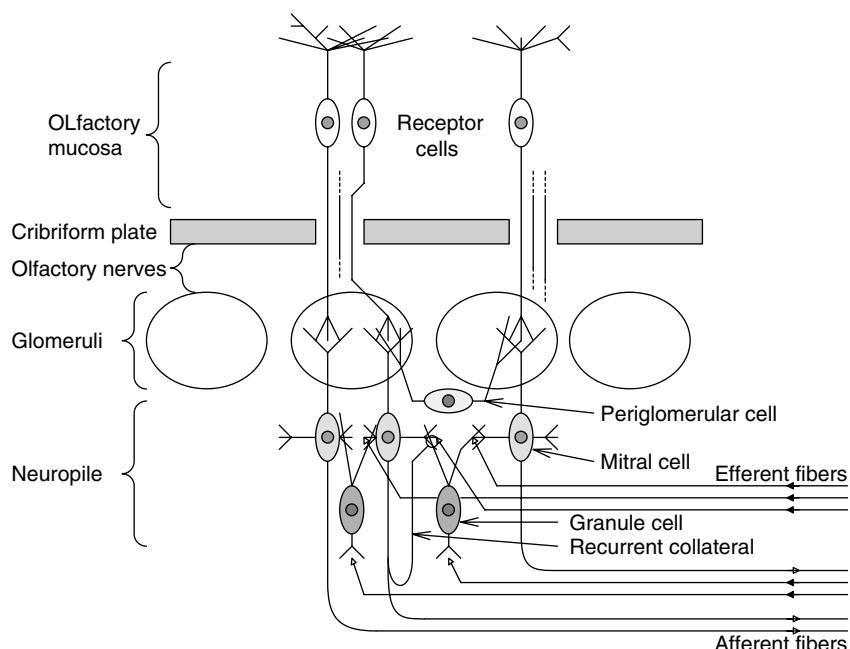


FIGURE 6.68

Schematic of some of the relations found between interneurons in the olfactory lobes.

example, in the rabbit, about 26,000 receptors converge into ca. 200 glomeruli, which then converge at 25:1 on each mitral cell [Leffingwell, 1999]. Strangely, humans have only about 347 olfactory receptor proteins [Zozulya *et al.*, 2001 in Leffingwell, 2001] (more may be found). The olfactory cells having one specific receptor appear to be distributed randomly in the nasal epithelium. However, their axons sort themselves out and converge on the same area of the optic bulb. Thus, there appears to be order out of chaos and a basis for some sort of fuzzy combinatorial logic to identify complex scents containing several odorants [Berrie, 1997]. Indeed, this layered neural architecture has inspired the application of artificial neural networks (ANNs) in the design of electronic 'noses'.

The periglomerular cells and granule cells in the olfactory lobe are known to be inhibitory on the mitral cells. Centrifugal (efferent) control fibers from the anterior olfactory nuclei stimulate these inhibitory cells. The receptor cell axons synapse excitatorily on the mitral, tufted and periglomerular cells. The mitral and tufted interneurons send their (afferent) axons to the lateral olfactory tract (LOT). The LOT axons synapse further in six different areas of the olfactory cortex: the amygdala, septal nuclei, pre-pyramidal cortex, entorhinal cortex, hippocampus, and subiculum.

It is well-known that human olfactory sensitivity can vary more than a factor of 10^3 between normal individuals. Some individuals are totally unable to smell certain odors, while others can. Certain odors such as the 'rotten egg' scent of H_2S can be sensed in concentrations of several parts per trillion. Mercaptan can be detected in concentrations as low as $7 \times 10^{-13} M$ (700 nM). Olfaction in humans (and presumably in other mammals) is ca. 10^4 times more sensitive than taste (gustation). As persons age, there is a loss of threshold sensitivity for certain odors, and the ability to discriminate between odors [Jacob, 2003].

The noses of trained dogs still remain the most reliable and most sensitive for detecting drugs and explosives. Dogs have also been trained to detect the scent of lung cancer on a patient's breath. Certainly, we have a long way to go to invent an artificial nose with the sensitivity of a dog.

6.11.4 Sensors for Gas-Phase Analytes

In man's attempt to design artificial olfactory systems, it soon became evident that certain classes of sensors (e.g. conductive polymers, metal oxides) had broad sensitivities to various VOCs (analytes). Thus, a sensor in an array would respond maximally to a specific analyte, but also produce smaller outputs in response to other VOCs. In some respect, this behavior mimics a live olfactory receptor array, which, when viewed at the receptor level, gives ambiguous responses. Elimination of this 'cross-talk' is done neurally by the olfactory bulb and other parts of the brain. As you will see below, man has relied on computer algorithms and artificial neural networks (ANNs) to sharpen the selectivity of his electronic noses to desired vapor phase analytes.

In order to perfect an electronic nose that will compete with dogs, much effort has been directed to improving the sensitivity and molecular specificity of 'nose' sensors to a desired vapor phase, odorant molecule. In this section, we will examine some of these sensors. There are currently a plethora of sensor technologies competing for the high (specificity \times sensitivity)/cost index for electronic nose front ends. These sensors are summarized in Table 6.7.

The first class of VOC sensor we shall examine is the conductive polymer film (CPF). Poly(3,4-ethoxy-dioxy)thiophene-poly(styrene sulfonate) polymer composites doped with carbon particles have been used at CalTech to develop VOC sensor arrays [Sotzing *et al.*, 2000a]; polyaniline/carbon composites have also been examined for the same

TABLE 6.7
Some sensors useful in electronic noses

Sensor Type	Operating Mechanism	Sensitivity	Advantages	Disadvantages
Conducting polymer	Adsorbed analyte causes ΔG	$> 10 \text{ ppt} > 5 \text{ ppm}$ more common 5–500 ppm	Inexpensive, microfabricated, operates at room temp.	Very sensitive to humidity
Metal oxides	Adsorbed analyte causes ΔG		Inexpensive, microfabricated	Operates at high temperatures. Big cross-sensitivities
MOSFET	Gate charge altered by analyte	ppm	Can be integrated with interface ICs.	Analyte bound to gate must be removed.
Surface acoustic wave (SAW) sensor	Analyte absorption causes phase or Δ frequency	1 pg mass change	High sensitivity	Noise, poor sensitivity
Surface Plasmon Resonance	Photon absorption due to $\Delta\epsilon$ from analyte	$> 0.01 \text{ ppb}$	High sensitivity	Complex interface signal conditioning
Fluorescence	Analyte molecules tagged with fluorophore molecule	ppb range	High sensitivity, high specificity	Expensive. Bound analyte must be removed to recycle
Mass spectrometry	Atomic mass spectrum	low ppb	Accuracy, high sensitivity	Expensive, hard to make portable
Optical spectroscopy	Beer's Law, λ -dependent transmittance	low ppb	Accuracy, high sensitivity	Expensive, requires pattern recognition algorithm, not portable
Gas chromatography	Absorbance of analyte on column. Eluent peaks at different times	low ppb	Accuracy, high sensitivity	Expensive, requires pattern recognition algorithm, not portable

purpose [Sotzing *et al*, 2000b]. Other CPF sensors have used polymers based on pyrroles, indoles, thiophenes and their derivatives [Tan *et al*, 2001].

A typical CPF sensor array contains up to 32 different polymer membranes having different but overlapping analyte sensitivities. The polymers are doped with conductive colloidal particles of carbon, gold or silver. When a conductive particle doped polymer film sensor is exposed to a VOC analyte, the analyte is adsorbed on the sensor's surface and causes a physical swelling of the polymer membrane. The swelling causes the electrically conductive matrix to become less dense, in turn lowering the longitudinal conductance of the membrane (the polymer is basically an insulator). Unfortunately, other physical factors such as temperature and humidity also affect the conductance, requiring that the sensor array be held at a constant temperature and humidity. The doped polymer films used at CalTech are of the order of 1–0.1 µm thick.

Response times for the CalTech CPF sensors to reach steady state output, given a suddenly applied analyte, are reported to range from ca. 0.1 to >100 s, and depend on the specific VOC, the polymer, as well as the sensor thickness [Lewis, 2000]. Lewis illustrates a sensor responding to 880 ppb of nitrotoluene, the signal well above the noise, and a sensor sensitivity of 10 ppt for biogenic amines. Note that the ΔR of the CPF sensor, caused by the analyte, can easily be sensed with a Wheatstone bridge or an op-amp transresistor (cf. Sections 4.1, 6.2). A commercially offered, handheld electronic nose using the CalTech developed conductive polymer sensors is now marketed as the Cryranose® 320; this e-nose uses a 32 CPF sensor array.

The second class of VOC analyte sensor is the metal oxide semiconductor (MOS) film [Moseley *et al*, 1991]. This class of sensor is operated at elevated temperatures (175–500°C). MOS materials investigated by Moseley *et al* include the oxides of Ga, In, Mo, Nb, Sn, Ta, Ti, W and Zn. Also examined for their analyte specificity were binary oxides such as CaSnO₃, trinary oxides such as SrFe_{0.8}Mn_{0.2}O₃, various niobates such as BiNbO₄, certain tantalates, perovskites, bronzes and pyrochlores. Not all materials showed increased resistance in response to a specific analyte; in fact, some showed a resistance decrease for certain VOC analytes. At 500°C, a BaSn_{0.9}Zr_{0.1}O₃ sensor prepared as a porous film 50 µm thick, deposited on an alumina substrate, showed responses (as + ΔR) to Cl₂, NO₂ and SO₂ at 1% concentrations. It also detected H₂S at <2 ppm. Curiously, the + ΔR of this sensor increased monotonically to [H₂S] up to 30 ppm, then at \geq 100 ppm, ΔR changed sign.

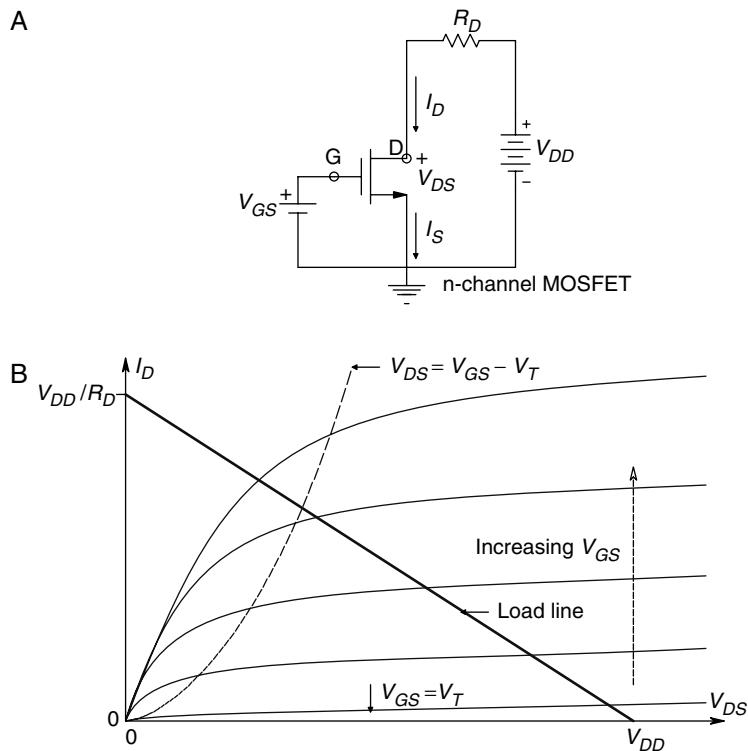
Moseley *et al* concluded:

"The expectation is, and observation confirms, that most [MOS] materials show a response to a range of gases, and that spectacular selectivity is not commonly to be found. Even so, there is plenty of useful variability among the different materials [tested], and in some cases very great selectivity."

Moseley *et al* go on:

"One can speculate that such selectivity tends to occur under conditions where the material is in a transition zone between n-type and p-type behaviour, where the effect of variations in the surface oxygen species has little effect on the conductivity and where the effect of specific surface coordination reactions might therefore be observed."

An experimental electronic nose, using nine various Figaro tin oxide sensors with a temperature and a humidity sensor, was reported by Keller *et al* (1995). Sensor outputs were the 11 inputs to an artificial neural network (ANN) which eliminated ambiguities and allowed the system to discriminate between the test substances—none (control),

**FIGURE 6.69**

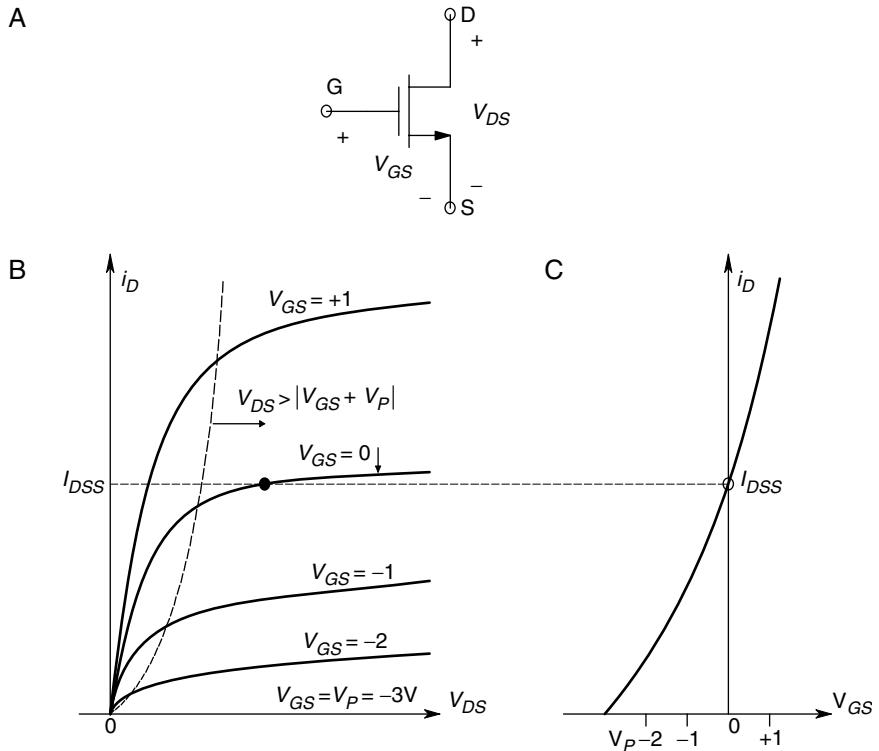
A. Basic grounded source, *n*-channel MOSFET amplifier. B. Characteristic $I_D(V_{GS}, V_{DS})$ curves for the *n*-channel MOSFET. The device is in ohmic operation for points to the left of the $V_{DS} = V_{GS} - V_T$ curve, and in saturated channel operation for $V_{DS} > (V_{GS} - V_T)$.

acetone, ammonia, isopropanol, lighter fluid and vinegar. Binary mixtures of the test substances were also presented—ammonia + isopropanol, ammonia + lighter fluid, ammonia + vinegar and isopropanol + vinegar. The fuzzy ARTmap ANN algorithm was found to give the best detection performance, averaging 93.4% correct over all its trials.

The third class of VOC analyte sensor is the insulated gate MOSFET. Figure 6.69 illustrates the basics of a simple *n*-channel MOSFET dc amplifier. When operating in the saturated channel mode, the dc drain current, I_D , is seen to increase as the gate source (actually, the gate channel) potential increases positively. When the MOSFET has depletion doping, $V_{GS}=0$ produces a constant drain current, I_{DSS} . Thus, when the gate is shorted to the source, the MOSFET behaves like a dc current source. This behavior is illustrated in Figure 6.70. (MOSFETs can be made from doped silicon, or gallium arsenide (GaAs)).

A chemical reaction at the gate that alters the gate charge can change V_{GS} and will change I_D . This ΔI_D is sensed; it is generally a monotonic function of the analyte concentration. For an effective MOSFET sensor, the induced ΔI_D must be resolved above noise and drift. Chemically sensitive MOSFETs are generally called ChemFETs. They can be subdivided into ion sensitive FETs (ISFETs), SURFETs, MEMFETs and POLFETs, depending on the details of their construction and application [Wroblewski, 2002]. Most ChemFET designs have been used to sense analytes in aqueous solution.

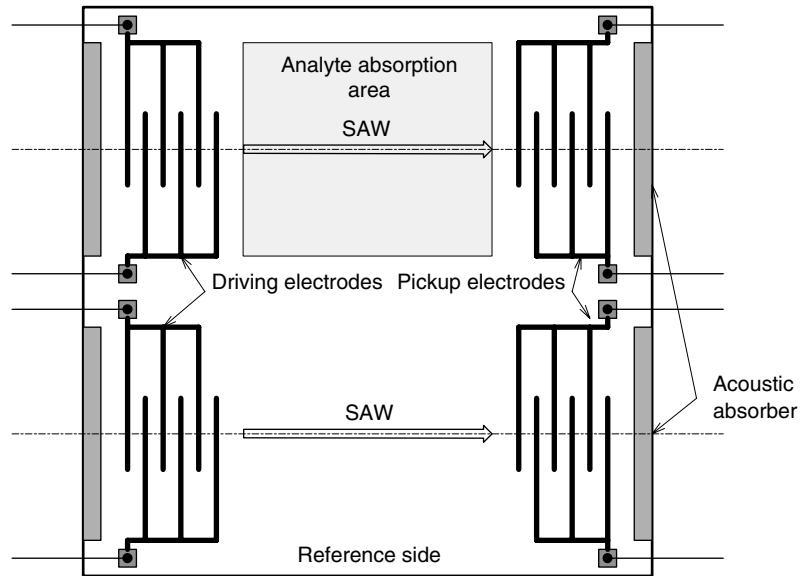
In one embodiment of a ChemFET gas sensor, the gate is covered with a thin polymer membrane, such a gas sensor, and is called a POLFET. The polymer can be

**FIGURE 6.70**

A. An *n*-channel MOSFET. B. $I_D(V_{GS}, V_{DS})$ curves for the *n*-channel MOSFET. C. Drain current vs V_{GS} for saturated channel operation. Note that when the gate is shorted to the drain, $V_{GS} \equiv 0$ and $i_D \equiv I_{DSS}$.

electrodeposited or sprayed on the gate [Briand, 2001]. POLFET sensors are noisy, with relatively high e_{ns} s (cf. Section 3.1). They also respond to temperature changes and stray ions in their gas input, all bad sources of uncertainty when making what are essentially dc measurements. In another embodiment of a gas sensing ChemFET, a porous (gas permeable) metal gate is vapor deposited on a thin catalyst layer overlying the silicon oxide insulating the FET's channel. Gas penetrates the porous metal and undergoes some redox reaction in the catalyst, liberating electrons, protons or ions which affect the gate channel potential and hence the I_D . Thus MOSFET gas phase sensors have comparatively poor sensitivity, generally in the ppm [Nagle *et al*, 1998]. ChemFETs do offer the advantage of being able to be made by IC fabrication processes, however. To cancel dc drift in I_D , they can be used differentially, one device not being exposed to the gas analyte. It appears that gas sensing ChemFETs have not been the sensor of choice in commercial e-nose products [Nagle *et al*, 1998].

The fourth class of VOC analyte sensor we will examine is the surface acoustic wave (SAW) sensor. Figure 6.71 illustrates a simple SAW device. The SAW device is built on a thin piezoelectric material slab, usually quartz or LiNdO₃ [Pallàs-Areny and Webster, 2001]. A surface area between the electrodes is coated by a polymer film or other chemical having affinity for the VOC analyte. Electrodes at one end of the slab are excited by an oscillator, generally in the range of hundreds of MHz. The excitation induces ultrasonic mechanical vibrations in the piezo-material, which propagate on its surface as Rayleigh (surface) waves, not unlike waves on water. To prevent wave reflections at the two ends of the piezo slab, an absorbing material is deposited, as shown in Figure 6.71. When the Rayleigh waves reach the electrodes at the far end of the coated piezo slab,

**FIGURE 6.71**

Top view of a dual SAW device. The phase delay of the reference SAW device is independent of the analyte.

a second set of electrodes picks up a vibration induced voltage. The Rayleigh surface waves propagate with a finite velocity. When the analyte molecules are absorbed by the polymer, its mass is increased. The mass increase causes a fractional increase of the propagating Rayleigh wave's velocity. For low values of analyte concentration, this increase can be written as:

$$v(M) = v_o + \alpha(M/A) = v_o(1 + \alpha(M/A)/v_o) \quad (6.149)$$

or,

$$v([C]) = v_o(1 + \beta[C]/v_o) \quad (6.150)$$

where v_o is the surface wave propagation velocity in the absence of absorbed analyte, α and β are constants, (M/A) is the mass/area of the analyte absorbed on the surface, and $[C]$ is steady state analyte concentration over the membrane. Thus the SAW device wave propagation velocity increases with absorbed analyte mass, or analyte concentration around the SAW device. Thus, a vibration wavefront launched from the excitation end will arrive at the pickup end $\delta t = L/v$ later. Thus for CW excitation, there will be a phase lag between output and input given by:

$$\theta = 2\pi f_o L/v \text{ radians} \quad (6.151)$$

where f_o is the oscillator frequency, L is the path length of the surface acoustic wave and v is the velocity of propagation. The v is affected by the density of the absorbent film covering the slab surface and the temperature of the film and slab. Assuming that temperature of the SAW gas sensor is held constant, we can rewrite the phase shift to include analyte concentration as:

$$\theta = \frac{2\pi f_o L}{v_o(1 + \beta[C]/v_o)} \simeq \frac{2\pi f_o L}{v_o^2}(1 - \beta[C]) \quad (6.152)$$

Thus, the sensor's phase shift decreases with increasing analyte concentration, [C] is:

$$\Delta\theta = -\frac{2\pi f_0 L}{v_o^2} \beta[C] \quad (6.153)$$

Figure 6.71 also illustrates a possible way to configure two SAW sensors to measure $\Delta\theta$ substantially independent of temperature. One sensor is exposed to the gas phase analyte, the other not. Both are at the same temperature. Note that to maximize phase shift sensitivity, we need a long L , a high f_0 and a low v_o .

Another way of exploiting the increase on SAW velocity, caused by the increased mass of the polymer layer, is to use the SAW sensor's propagation lag, $\delta_t = L/v$ to make a feedback oscillator. One can then detect small changes in the oscillation frequency as a function of [C]. δ_t can be viewed as a transport lag:

$$\delta_t = \frac{L}{v_o(1 + \beta[C]/v_o)} \cong (L/v_o)(1 - \beta[C]/v_o) \quad (6.154)$$

Figure 6.72 illustrates a simple negative feedback oscillator circuit. The amplifier has the transfer function, $A_v(s) = -K_{vo}/(s + a)$. K_{SAW} is the volt/volt gain from input electrodes to output electrodes of the SAW device. The oscillator's loop gain transfer function is thus:

$$A_L(s) = \frac{-K_{vo} K_{SAW} e^{-s\delta t}}{s + a} \quad (6.155)$$

Use of the root locus technique angle criterion for instability [Northrop, 2000] gives us an expression for the frequency of oscillation. In general, the root locus plot is based on complex solutions of the vector equation:

$$1 - [\mathbf{A}_L(s)] \equiv 0 \quad (6.156)$$

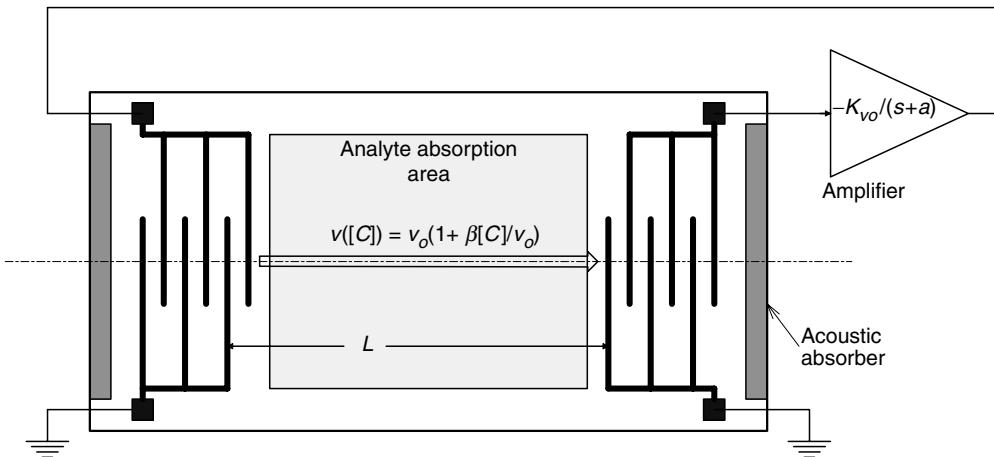


FIGURE 6.72

Schematic of a negative feedback SAW delay oscillator.

The system oscillates when the gain $K_{vo}K_{SAW}$ is high enough for two root locus branches cross the $j\omega$ axis at $s = \pm j\omega_o$. This is where the vector $s = \sigma + j\omega = j\omega_o$. To find ω_o , we must solve the equation above based on the RL *angle criterion*. In general:

$$-\mathbf{A}_L(j\omega_o) = -1 \quad (6.157)$$

The angle criterion is:

$$\angle -\mathbf{A}_L(j\omega_o) = -\pi \quad (6.158)$$

Thus:

$$-\omega_o\delta_t - \tan^{-1}(\omega_o/a) = -\pi \quad (6.159)$$

Equation 6.159 is a transcendental equation, whose solution gives the SAW delay oscillator's oscillation frequency. Note that the oscillation frequency, ω_o , is a function of the SAW propagation delay, δ_t . There is no explicit solution to equation 6.159, but we can linearize it by assuming $(\omega_o/a)^2 \ll 1$, so that the arctangent function can be approximated by $\tan^{-1}(x) \cong x$ radians. Thus:

$$\begin{aligned} \omega_o\delta_t + \omega_o/a &\cong \pi \\ \downarrow \\ \omega_o &\cong \frac{\pi}{\delta_t + 1/a} = \frac{a\pi}{a(L/v_o)(1 - \beta[C]/v_o) + 1} \\ &= \frac{a\pi}{[aL/v_o + 1] - aL\beta[C]/v_o^2} \\ &\cong \frac{a\pi}{[aL/v_o + 1]} \left[1 + \frac{aL\beta[C]/v_o^2}{[aL/v_o + 1]} \right] = \omega_c + \Delta\omega \end{aligned} \quad (6.160)$$

Thus the incremental change in oscillator frequency due to the analyte absorption, $\Delta\omega$, can be written:

$$\Delta\omega = \frac{-\pi a^2 L \beta[C]}{[aL + v_o]^2} \text{ r/s} \quad (6.161)$$

The fractional change in oscillator frequency due to analyte absorption is:

$$\frac{\Delta\omega}{\omega_c} = \frac{-aL\beta[C]}{v_o(aL + v_o)} \quad (6.162)$$

ω_c is the oscillator frequency in the absence of analyte ($[C] = 0$).

Unfortunately, in polymer membrane based SAW sensors, there is a trade-off between analyte sensitivity and response speed to transient changes in analyte concentration. A thin polymer film rapidly saturates with analyte molecules in the gas having a fixed VOC analyte concentration. Since it is thin, its total mass increase will be smaller than for a thick film. The thick film takes longer to saturate because of diffusion

time. It also takes longer to lose the analyte into pure carrier gas (air). Maximum absorption sensitivity in thin films is in the range of femtograms ($\times 10^{-15}$) of analyte. Detectors for CO, HCl, H₂, H₂S, NH₃, NO₂, SO₂, various VOCs and organophosphorous compounds (insecticides) have been developed using SAW sensor arrays [Staples, 1999].

A fifth class of VOC analyte sensor is the quartz resonator. An analyte binding surface film is deposited on a quartz crystal which is part of an oscillator. When analyte molecules are either absorbed or chemically attached on the film, its mass increases and this raises the frequency of oscillation of the oscillator. Mass increases on the order of 1 pg can be detected. A sensor of this type is also called a quartz crystal microbalance (QCM). If a liter of air has 1 pg of methane in it, this translates to a concentration of 1.4 ppb (1.4×10^{-9}) [Nagle *et al*, 1998]. QCM research is being carried out at IBM's Zurich Research Lab.

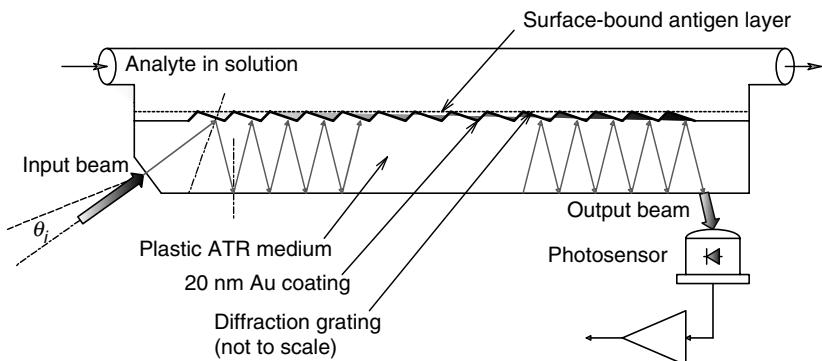
A sixth class of VOC analyte sensor for e-nose applications is Surface Plasmon Resonance (SPR). SPR is a quantum optic-based analytical method that makes use of surface binding of an analyte. SPR is a relatively new analytical tool which is just now being investigated for e-nose applications. SPR allows rapid, specific determination of the concentration of a variety of environmentally, medically and biologically important analytes, either in liquid or gas phase. For example, specific bacteria, antigens, antibodies, theophylline, caffeine, gases, pesticides, explosives, controlled substances (opioid drugs), etc., have been sensed with SPR in solution or as vapors. Threshold sensitivities for certain analytes have been reported as low as 0.01 ppb (10^{-11}). SPR technology is relatively simple and inexpensive to implement compared with traditional analytical systems such as HPLC, mass spectrometry, IR spectrometry and gas chromatography, and SPR instruments lend themselves well to field measurements. Since this section is about e-noses, the analytes considered are largely gas phase and are derived from a variety of sources including drugs, explosives, environmental contaminants, as well as from medical fluids such as blood, urine, feces, bacteria, cancers, etc.

SPR is a quantum optical phenomenon which occurs when a beam of monochromatic, linear polarized light (LPL) is reflected off a thin metal film which has been vapor deposited on one side of a glass prism, or when a beam of LPL is incident at a critical angle on a gold coated diffraction grating. SPR is described in detail in Section 7.6.6.

The reader should be aware that the design of SPR devices for chemical analysis is a rapidly growing field. One alternative configuration of the SPR sensor places the analyte solution over the surface of a plastic diffraction grating whose surface has been vapor deposited with gold, silver or aluminum (gold is generally preferred because of its chemical stability). The plastic top of the grating acts as an attenuated total reflection (ATR) prism, where light reflected from the grating at the point at which the beam strikes it initially, is reflected back many times to the grating surface. Such a grating SPR design was proposed by Simon (1998). Figure 6.73 shows a side view of Simon's 'long range', SPR grating system. SPR occurs at a critical beam input angle, θ_i , giving a minimum of output light intensity when the coupling component of the monochromatic, TM-polarized, input beam wave vector satisfies the relation:

$$(2\pi/\lambda)\sin(\theta_i) + (2\pi/b) = (2\pi/\lambda)\sqrt{(\kappa_a\kappa_m)/(\kappa_a + \kappa_m)} \quad (6.163)$$

Here, b is the grating constant, κ_a is the dielectric constant of the analyte and κ_m is the magnitude of the complex dielectric constant of the metal film. A very thin layer of polymer with high affinity to the analyte to be detected is chemically bound to the 20 nm gold film over the grating surface. The analyte binds to the polymer molecules,

**FIGURE 6.73**

A surface plasmon resonance system after Simon (1998). The sensitivity is enhanced by lengthening the effective lightpath with an ATR prism and causing multiple interactions with the metal film on the grating which is in intimate contact with the analyte.

changing κ_a at the gold-polymer interface. This change of dielectric constant changes the interfacial refractive index and 'retunes' the SPR to a new input angle, θ_i . Either by changing θ_i to retain a minimal light output, or by measuring the increase in light output due to $\Delta\kappa_a$, the number or density of bound analyte molecules can be quantified. Simon (1998) claimed that a refractive index change in the analyte of less than 1% causes a 0.5° shift in θ_i to renull the system.

Other embodiments of the grating SPR system place the analyte over the gold film covered grating. The monochromatic, TM polarized light is directed through a thin film containing the analyte onto the gold covered grating. Jory *et al* (1995) reported on an exquisitely sensitive grating SPR system in which an acousto-optical tunable filter (AOTF) element was used to control the wavelength of the incident beam to a precision of 0.0005 nm. They used their system to measure the concentration of NO₂ gas in N₂ (a gas phase analyte). By depositing a thin layer of phthalocyanine over the gold coating, a wavelength shift of -0.004 nm renulled the system from zero concentration when 0.01 ppm NO₂ in N₂ was applied. They claimed that the sensitivity of their system allowed detection of changes in the refractive index of the gas of 1×10^{-6} .

Capan *et al* (2003) reported on a Kretschmann type SPR system in which a thin film of polymethyl methacrylate (PMMA) over the gold film was found to respond most strongly to benzene vapors and less strongly to toluene, methyl benzene and *m*-xylene vapors. Threshold sensitivity was about 100 ppm (10^{-4}).

In a 1999 review of SPR sensing, Homola *et al* describe the use of polyethylene glycol films in sensing the vapors of hydrocarbon, aldehydes and alcohols. Chlorinated hydrocarbons could be sensed by SPR by their absorption in polyfluoroalkylsiloxane films, and tetrachloroethane could be resolved by absorption in a polydimethylsiloxane film. Teflon films have been used to sense aromatic hydrocarbon vapors. Still other coatings have been reported to be selective for other vapor analytes (e.g. polyaniline was used to sense NO₂ and H₂S, bromocresol purple was used to sense NH₃). When the SPR active metal film coating itself was palladium, which has affinity to hydrogen, it responded to molecular hydrogen, H₂.

SPR gas sensors of the Kretschmann design are apparently more expensive than the conductive polymer film and metal oxide sensors described above and their sensitivities are generally on the order of ppm of analytes, rather than ppb, with some exceptions. Practically any vapor analyte that can react with a chemical film bound to the gold film surface on a prism or grating with a strong affinity can be sensed by SPR.

The reaction must cause a change in the refractive index or the permittivity at the gold surface in order to affect the SPR conditions (reflected intensity, λ and/or θ_i).

The problem of quick, efficient, SPR sensor regeneration remains to be solved, however. Once the bound surface reactant has combined with the analyte, the analyte or analyte complex must be totally removed from the bound reactant film before the next measurement can be made. That is, the surface reactant layer must be renewed or rejuvenated by some means which can be physical (e.g. heat) or chemical (a purging solvent that does not affect the ligand film over the gold).

The seventh class of gas sensing system includes traditional chemical analytical systems that are generally not hand carryable. These include:

1. The quadrupole mass spectrometer (QPMS) is lighter and more portable than a conventional MS that uses a magnetic field to separate ions. A QPMS produces a mass spectrogram with peaks representing ions broken off the analyte. Its peaks have a unique pattern for a given analyte [Northrop, 2002].
2. All sorts of spectrophotometers that measure the absorption or transmission of light through a sample chamber containing the analyte gas. Light wavelengths from FIR through to UV can be used to establish an analyte's unique signature spectrogram. Spectrophotometry gives less certain results when presented with mixtures of analytes.
3. Gas chromatography exploits the physical property of vapor phase analytes to propagate through a capillary adsorption column at different rates. The peaks in a gas chromatogram are thus from the individual analytes in a mixture, separated in time. The temporal separation of the peaks, given the column's temperature are signatures of the analytes, the peak areas are proportional to the concentrations. Standard analyte samples are used to calibrate a gas chromatograph. The type of detector used in a gas chromatograph determines its sensitivity and the class of analyte it is best suited for. For example, the minimum detectable amount (MDA) of chlorinated hydrocarbons using an electron capture detector is ca. 10 fg/ml (10^{-14} g/ml). A flame ionization detector's MDA is ca. 1 pg/ml (10^{-12} g/ml) for hydrocarbons [Northrop, 2002]. Due to its simplicity and sensitivity, efforts are underway to miniaturize GC columns and sensors.
4. Various fluorescent techniques for gas sensing work in several ways. In one, a fluorescent probe compound that has an affinity for the vapor analyte is used. Fluorescence is induced by shining short wavelength light (blue or UVA) on the probe compound. As the analyte reacts with the fluorescent probe substance, it can quench the probe's natural fluorescence (dims it proportional to the analyte concentration), or, the bound analyte molecules can cause a shift in the fluorescent emission spectrum, or increase the intensity of the fluorescence [Levitsky *et al*, 2001]. A colorimetric approach described by Rakow and Suslick (2000) uses an array of metalloporphyrin dyes whose colors change in response to analyte binding. These sensors permit the visual identification of ligating vapors (alcohols, amines, ethers, phosphines, phosphites, thioethers, thiols, arenes, halocarbons and ketones. Sensitivity extends in to the single ppm (10^{-6}) range. At the stage of development reported, the dye dots appear to be a qualitative means of analysis. However, intensity and wavelength variables should be able to yield quantitative data as well.

In summary, most commercial e-noses make use of metal oxide semiconductor sensors or the conductive polymer sensors.

6.11.5 Feature Extraction in E-Noses

A central problem in using arrays of gas sensors having different sensitivities to a group (mixture) of analytes is finding an algorithm that will yield a reliable estimate of the analyte concentrations in the mixture. In general, the fractional change in resistance of the k th sensor in an array of N sensors can be written, assuming linearity of response for low concentrations, as:

$$\frac{\delta R_k}{R_{ok}} = \sum_{i=1}^N \sigma_{ki} C_i, \quad k = 1, 2, \dots, N. \quad (6.164)$$

where σ_{ki} is the k th sensor's response sensitivity to the i th analyte in the mixture and C_i is the concentration of the i th analyte in the mixture. From the N equations of the form of equation 6.164, we can use Cramer's rule to solve for $\{C_1, C_2 \dots C_N\}$. In their review paper on Cross-Reactive Chemical Sensor Arrays, Albert *et al*, 2000 describe and cite references for various linear and nonlinear algorithms which have been used to estimate $[C_i]_N$ from MOS and CP sensor arrays, as well as other gas sensors with cross sensitivities. These include cluster analysis (CA), transformed cluster analysis (TCA), principal component analysis (PCA), partial least squares (PLS) and various architectures of artificial neural networks (ANNs). The ANN approach offers the advantage in that it can be 'trained' to detect desired analytes. Keller *et al* (1995) cite the early use of ANNs in e-nose design. Ouellette (1999) also discusses the use of ANNs with e-nose sensor arrays. The Cyranose® 320 e-nose described by Hobbs (2000), which uses an array of 32 CP sensors, uses PCA for detection of outliers and data reduction, along with three supervised algorithms for model building and predictions. These algorithms are K-nearest neighbor (KNN), K-means and Canonical Discriminant Analysis (CDA).

It is way beyond the scope of this text to treat the details of the data processing algorithms cited above used to extract concentration data from sensor array outputs.

6.11.6 Summary

The development of sensitive, portable e-noses (sensors and data processing algorithms) continues to be an active area in chemical forensic science. Applications of e-noses which drive development funding are the need for portable sensors that will detect drugs and explosives at least with the sensitivity of the legendary canine nose. Other applications lie in the food industry. The detection of bacteria, spoiled or contaminated food are important for public health. Being able to objectively control aromas and flavors from lot-to-lot of a foodstuff (e.g. coffee, cheese) is also important. In the area of environmental science, there is a need to detect contaminants in water and soil at the ppb level and better yet, in the field. Currently, samples are taken back to a lab for analysis by gas chromatography/mass spectrometry. These include insecticides and industrial chemicals. In forensic science, e-noses can find use in sensing buried corpses, a role currently occupied by dogs.

6.12 Mechano-Optical Sensors

In this section, we discuss certain sensor mechanisms, excluding fiber optics, in which mechanical parameters such as displacement and velocity cause changes in the intensity

of transmitted or reflected light, either directly, or as the result of induced changes in the linear polarization angle of the output light, or from optical interference. In the simplest case, shaft angle is sensed with an optical encoder disk. Shaft velocity can be found by differentiating the position signal.

Another means of measuring angular velocity is through the use of the Sagnac effect, discussed below. The displacement of interference fringes is proportional to the angular velocity. The Sagnac effect forms the basis for modern optical gyroscopes, which are discussed in the following chapter. While their primary response mode is to angular velocity, gyroscopes can be combined with electromechanical feedback systems in order to sense angular position. Such systems are covered in Chapter 7. The linear velocity of a fluid medium can be measured optically by laser Doppler velocimetry (LDV), the basics of which are described below.

6.12.1 Optical Coding Disks

Optical coding disks are used to convert analog shaft rotation angle to a digital word proportional to the angle. Optical coding disks may be subdivided into two categories—Incremental Position Encoders and Absolute Position Encoders. In both types of sensors, light is either transmitted to or blocked from a linear phototransistor array, depending on the angle of the shaft to which the photoetched coding disk is attached. Thus, the coding disk output is digital and may be input directly into a computer controlling the shaft angle or recording its value. Figure 6.74A shows a three-bit, absolute position encoding disk etched for a straight binary output code and B) illustrates a gray code etched disk. Practical resolution of an absolute position disk is 14 bits. The MSB (2^0) track is the innermost track; higher order bits are encoded radially outward, with the LSB (2^{N-1}) lying on the outside edge of the disk. The arc length of the smallest (outer) sectors can be shown to be given by:

$$l \cong R \tan(360/2^{N+1}) \quad (6.165)$$

Thus for $N=14$ and $R=4$ in, the radius to the outer track, the resolvable sector length l is 0.000767 in or $19.5\text{ }\mu\text{m}$. This also implies that one LSB of shaft revolution is $\theta = 360/2^{N+1} = 1.099 \times 10^{-4}$ degrees. To obtain finer resolution, one only has to connect a

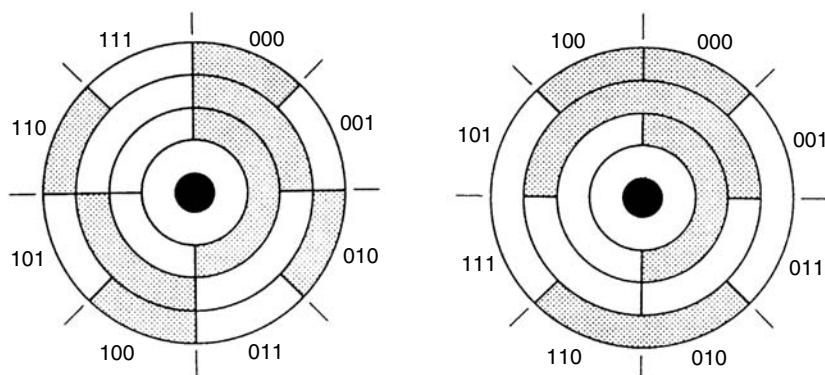


FIGURE 6.74

A. A three-bit, straight binary, absolute position shaft coding disk. B. A three-bit, gray code disk.

second coding disk to the primary shaft through a step-up gear ratio of $2K$, where K is 1, 2, 3, etc., and use the outer K tracks of the second disk as the K LSBs.

Incremental coding disks are far simpler in their construction. On their outer edges, they have photoetched two simple, light/dark square wave patterns, one of which is 90° out of phase with the other. Depending on the application, 100–1000 or 1024 cycles may be used. Once a reference angle is established, total shaft angle can be obtained by counting the output pulses with up/down counters, such as the 74LS192 or 74LS193. Comparison of the phase of the two square wave outputs of the incremental coding disk gives rotation direction and sets the up/down control on the counters.

It should be pointed out that in the case of absolute position encoders, angular velocity information can be obtained from the digital signals from optical coding disks by simple digital differentiation, and by simply counting the pulses per unit time for the output of an incremental disk, or doing digital differentiation on the position counts stored in the up/down counters.

Optical mask techniques, similar to those used in angle coding disks, can be used to realize direct linear analog displacement to digital coding sensors. Linear, charge coupled photodetector arrays can also be used with an opaque mask and lens optics to measure linear displacements with a precision of the order of 1/1024. However, the output of such a CCD system is basically analog and the array element number at which the light/dark transition occurs gives the desired displacement.

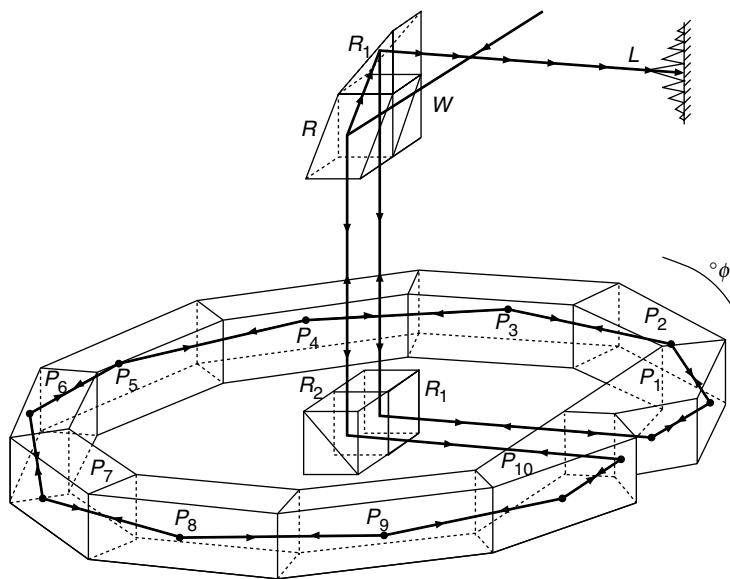
Yet another method of optical detection of linear displacement, described in Barney (Section 7.4.1, 1985), makes use of the moving interference fringes caused by the superposition of two linear Moiré, patterns. An optical detection scheme that uses the relative motion of the fringe pattern to obtain direction and displacement information is employed.

6.12.2 Sagnac Effect Sensing of Angular Velocity

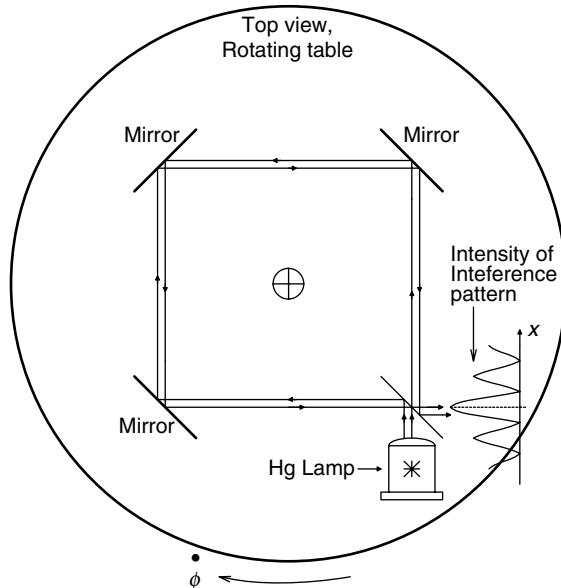
The Sagnac effect is the principle underlying the modern fiber optic gyroscope. In this section, we will describe the basic Sagnac effect. Fiber optic gyros will be discussed in detail in the next chapter. We stress that the Sagnac effect occurs whether the light traverses a closed path in air or in a glass optical fiber loop.

The Sagnac effect was first observed in 1911 by Harress. Harress constructed a polygonal ring interferometer using glass prisms, diagrammed in Figure 6.75. His purpose was to measure the dispersion properties of glasses. When the ring was rotated, a shift in the interference fringe pattern was noted. Harress assumed that this shift was due to Fresnel-Fizeau drag. This drag effect is observed when light propagates through a linearly moving optical medium.

Sagnac performed his experiments in 1914. A mercury green line source was collimated and directed at a beam splitter (a half-silvered mirror at 45°). The two beams were directed in opposite but colinear paths in air around an optical table, as diagrammed in Figure 6.76. The half silvered mirror served to direct the beams to a screen where the incident beams produced interference fringes. Sagnac's loop area was 866 cm^2 and he rotated his interferometer at 2 revolutions/s. Sagnac observed a fringe shift of 0.07 fringes as the entire optical table containing mercury lamp, lenses, mirrors and screen was rotated about an axis perpendicular to the table. Sagnac was able to show that the fringe shift was not due to Fresnel-Fizeau drag and he derived a relation for the fractional fringe shift given by equation 6.166. In 1926, Pogany repeated Sagnac's experiment with improved apparatus in which $\lambda_0 = 546 \text{ nm}$, $A = 1178 \text{ cm}^2$ and $\dot{\phi} = 157.4 \text{ r/s}$. He measured $\Delta z = 0.906$; the theoretical value was 0.900. Pogany repeated his experiment in 1928 and obtained results within 1% of theoretical values.

**FIGURE 6.75**

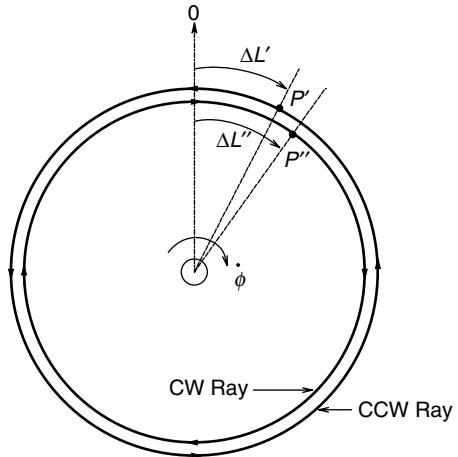
Harress' ring interferometer (a Sagnac system).

**FIGURE 6.76**

Top view of Sagnac's ring interferometer in which fringe shifts are proportional to the angular velocity of the apparatus around the center of the table.

From the early work on the Sagnac effect, four principles emerged:

1. Equation 6.166 applies for a one loop system
2. The fringe shift does not depend on the shape of the closed light path, only on its area
3. The fringe shift does not depend on the axis of rotation; it does, however, depend on the path tilt, θ
4. The fringe shift does not depend on the propagation medium

**FIGURE 6.77**

A single OF ring relevant to the derivation of the Sagnac gyro equation.

The Sagnac equation for fringe shift is:

$$\Delta z = [4\dot{\phi}A \cos(\theta)]/(\lambda_o c) \quad (6.166)$$

where Δz is the fractional fringe shift, λ_o is the wavelength of light, A is the area of the light path, θ is the angle the rotation axis makes with the perpendicular to the plane of the light path, c is the velocity of light around the path, and $\dot{\phi}$ is the angular velocity of path rotation around an axis at an angle θ to a perpendicular to its plane, in radians/sec.

It is easy to derive equation 6.166 above if we consider a circular light path, rather than a rectangular one. Such a circular path, shown in Figure 6.77, could be the limiting result of using a very large number of prisms or mirrors, as shown in Figure 6.75 (Harress' polygonal approximation), or of using a single circular turn of a glass optical fiber. Assume that at $t=0$, the input/output point of the ring is at 0° . First, we assume that the ring is stationary, or $\dot{\phi}=0$. In this case, it takes $\tau_o = 2\pi R/c$ for a light wave to circle the ring, either clockwise or counterclockwise. Now, let the ring rotate clockwise at a constant angular velocity, $\dot{\phi}$. In the time, τ_{cw} , a fixed point on the ring rotates an arc length $\Delta L''$ to point P'' . This may be expressed as:

$$\tau_{cw} = \Delta L''/R\dot{\phi} \quad (6.167)$$

In the same time, the clockwise lightwave travels

$$\tau_{cw} = (2\pi R + \Delta L'')/c \quad (6.168)$$

Obviously, $\tau_{cw} > \tau_o$. A wave propagating counterclockwise takes less time to reach point P' sic:

$$\tau_{ccw} = (2\pi R - \Delta L')/c = \Delta L'/R\dot{\phi} \quad (6.169)$$

The fractional fringe shift, Δz , is given by:

$$\Delta z = c\Delta\tau/\lambda_o \quad (6.170)$$

where $\Delta\tau$ is given by:

$$\Delta\tau = \tau_{cw} - \tau_{ccw} = (2\pi R + \Delta L'')/c - (2\pi R - \Delta L')/c = (\Delta L'' + \Delta L')/c \quad (6.171)$$

Now it is easy to show that:

$$\Delta L'' = 2\pi R^2 \dot{\phi}/(c + R\dot{\phi}) = (2\pi R^2 \dot{\phi}/c)/(1 + R\dot{\phi}/c) \approx 2\pi R^2 \dot{\phi}/c \quad (6.172)$$

Similarly,

$$\Delta L' = 2\pi R^2 \dot{\phi}/(c - R\dot{\phi}) = (2\pi R^2 \dot{\phi}/c)/(1 - R\dot{\phi}/c) \approx 2\pi R^2 \dot{\phi}/c \quad (6.173)$$

So, the fractional fringe shift can finally be written as:

$$\Delta z = (\Delta L'' + \Delta L')/\lambda_0 = 4\pi R^2 \dot{\phi}/(c\lambda_0) = 4A\dot{\phi}/(c\lambda_0) \quad (6.174)$$

which is the same as equation 6.166, with $\cos(\theta) = 1$.

As an example, the basic sensitivity of a one-turn Sagnac system can be calculated from equation 6.166, θ is assumed to be 0° :

$$\begin{aligned} \Delta z/\dot{\phi} &= 4A/(c\lambda_0) = 4 \times 1 \text{ m}^2 / (3 \times 10^8 \text{ m/s} \times 4 \times 10^{-7} \text{ m}) \\ &= 0.033 \text{ fringe/r/s} \end{aligned} \quad (6.175)$$

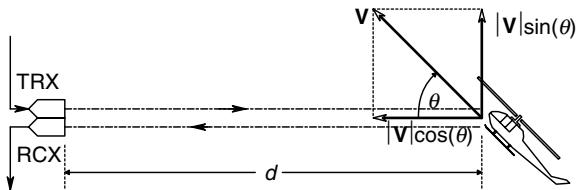
Clearly, sensitivity will be raised if multiple optical paths are used, a situation addressed in the discussion of fiber optic gyroscopes in the next chapter.

6.12.3 Laser Doppler Velocimetry

Laser Doppler velocimetry (LDV) provides a ‘no touch’ means of measuring the linear velocity of fluids (and particles). As the name suggests, LDV makes use of the Doppler effect at optical frequencies. In this section, we will review the basics of LDV and discuss certain key applications.

First, we describe the Doppler effect. We are all familiar with the Doppler effect on sound. A car moving towards us blows its horn. As it passes, there is a perceptible downward shift in the pitch of the horn. Johann Christian Doppler presented a paper, On the Colored Light of Double Stars and Some Other Heavenly Bodies, in 1842 before the Royal Bohemian Society of Learning. Doppler was Professor of elementary mathematics and practical geometry at the Prague State Technical Academy. He apparently got little recognition for his work, and died of consumption in 1854 at the age of 45. In 1844, a contemporary of Doppler, Buys Ballot, contested Doppler’s theory as an explanation for the color shift of binary stars. Ballot actually did an experiment using sound waves, where a trumpet player played a constant note while riding on a flatcar of a train moving at constant velocity. A musician with perfect pitch, standing at trackside, perceived the trumpet tone to be a half-tone sharp as the train approached, and a half-tone flat as it receded. In spite of this evidence, Ballot continued to object to Doppler’s theory. Ballot’s publications apparently served to discredit Doppler for a number of years.

To derive the Doppler effect for either electromagnetic or sound waves, we assume a moving, reflecting target and a stationary source/observer, as shown in Figure 6.78.

**FIGURE 6.78**

A simple Doppler system with stationary, colinear transmitter and receiver directed at a moving object.

Assume that the sinusoidal waves leaving the stationary transmitter (TRX) propagate at velocity, c , over a distance, d , to the target, T . The target is moving at velocity, \mathbf{V} , toward the source at an angle θ . Velocity \mathbf{V} can thus be resolved into a component parallel to the line connecting TRX and the reflecting target, and a component perpendicular to the TRX-T line. These components are $V \cos(\theta)$ and $V \sin(\theta)$, respectively. The reflected wave from T propagates back to the stationary receiving transducer, RCX, along path d . The receiving sensor output waveform can be written as:

$$V_o = B \sin[\omega_o t + \psi] \quad (6.176)$$

The transmitted radian frequency is ω_o and ψ represents the phase lag between the transmitted signal and the received signal. In general, the phase lag, ψ , is given by:

$$\psi = 2\pi 2d/\lambda = 2(2\pi)d/(c/f_o) = \omega_o 2d/c \text{ radians} \quad (6.177)$$

However, the distance $2d$ is changing because of the target velocity component along the line from the transducers to the target. Thus, the frequency of the received signal, ω_r , is the time derivative of its phase:

$$\omega_r = d[\omega_o t + \omega_o 2d/c]/dt = \omega_o(1 + [2/c]\dot{d}) = \omega_o(1 + [2V/c]\cos(\theta)) \quad (6.178)$$

The frequency, ω_D , is the Doppler frequency shift which contains the velocity information:

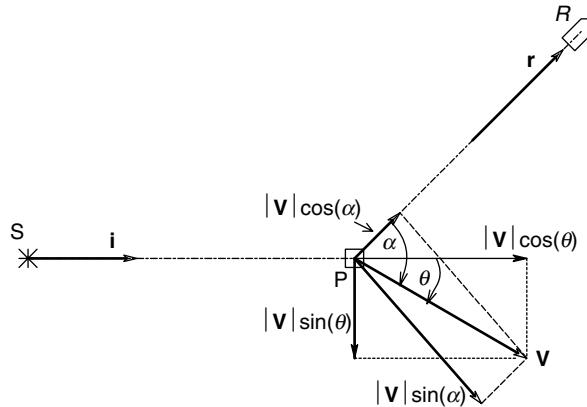
$$\omega_D = (\omega_o 2V/c) \cos(\theta) \text{ r/s} \quad (6.179)$$

Note that $\omega_r > \omega_o$, because in this example, the target is approaching the source/sensor. In LDV systems, the wave return path is seldom colinear with the path from the source to the reflecting object. A basic LDV system geometry is shown in Figure 6.79. The source, S , emits light with wavelength, λ_s , that propagates toward reflecting particle, P . Unit vector, \mathbf{i} , is directed from S to P . P moves with velocity, \mathbf{V} . It is assumed that $(|\mathbf{V}|/c)^2 \ll 1$, so that relativistic effects are negligible. If $\mathbf{V} = 0$, the number of wavefronts striking the particle per unit time is $f_s = c/\lambda_s$ (f_s for a HeNe laser is $3 \times 10^8 / 632.8 \times 10^{-9} = 4.741 \times 10^{14}$ Hz). The number of wavefronts hitting a moving particle, P is:

$$f_p = (c - \mathbf{V} \cdot \mathbf{i})/\lambda_s \text{ Hz} \quad (6.180)$$

Note that the vector dot product gives the particle velocity component parallel to \mathbf{i} . The wavelength, λ_p , apparent to the particle is:

$$\lambda_p = c/f_p = (\lambda_s c)/(c - \mathbf{V} \cdot \mathbf{i}) = (\lambda_s c)/[c - V \cos(\theta)] \quad (6.181)$$

**FIGURE 6.79**

Two dimensional geometry of a Doppler system used in LDV. Note that the transmitter (laser) (S) and the receiver (R) are at different angles to the object's velocity vector, \mathbf{V} .

Now the unit vector from the instantaneous particle position, P, to the stationary receiver, R, is \mathbf{r} . An observer at R sees a scattered wavelength, λ_r , from P:

$$\lambda_r = (c - \mathbf{V} \cdot \mathbf{r})/fp = [c - \mathbf{V} \cos(\theta)]/fp \quad (6.182)$$

Here again, $\mathbf{V} \cdot \mathbf{r}$ represents the velocity component of the particle in the \mathbf{r} direction. Hence, the frequency of the received, scattered radiation is:

$$f_r = c/\lambda_r \text{ Hz} \quad (6.183)$$

Substituting from above, we find:

$$f_r = c(c - \mathbf{V} \cdot \mathbf{i})/[\lambda_s(c - \mathbf{V} \cdot \mathbf{r})] \quad (6.184)$$

Now the observed Doppler frequency shift is:

$$f_D = f_r - f_s = (c/\lambda_s) \left[\frac{c - \mathbf{V} \cdot \mathbf{i}}{c - \mathbf{V} \cdot \mathbf{r}} - 1 \right] \quad (6.185)$$

For $\mathbf{V} \ll c$, equation 6.185 reduces to:

$$f_D \cong \mathbf{V} \cdot (\mathbf{r} - \mathbf{i})/\lambda_s = \mathbf{V}[\cos(\theta) - \cos(\alpha)]/\lambda_s \quad \text{Hz} \quad (6.186)$$

where θ is the angle between \mathbf{V} and \mathbf{i} , and α is the angle between \mathbf{V} and \mathbf{r} . Note that if the reflected radiation is directed back along \mathbf{i} to R coincident with S, the angle $\alpha = 180^\circ + \theta$. Since $\cos(180 + \theta) = -\cos(\theta)$, f_D given above reduces to the value given by equation 6.179.

Figure 6.80 shows the basic system first used for LDV by Yeh and Cummins (1964). A HeNe laser was used with $\lambda_s = 632.8 \text{ nm}$. A number of $0.5 \mu\text{m}$ diameter polystyrene spheres were used in water at a density of 1:30,000 by volume to obtain scattering. Sphere velocity is assumed to be equal to water velocity. The laser reference beam was frequency shifted by 30 MHz using a Bragg cell acousto-optic modulator. The measurement beam was directed parallel to the fluid velocity vector, \mathbf{V} , hence $\theta = 0^\circ$. The scattered, Doppler-shifted beam was taken off at, 30° relative to the angle of \mathbf{V} and passed through a half silvered mirror and directed to the photocathode of a photomultiplier tube (PMT), colinear with the frequency shifted reference beam. Optical mixing occurs at the surface of the PMT photocathode. Optical mixing is a process

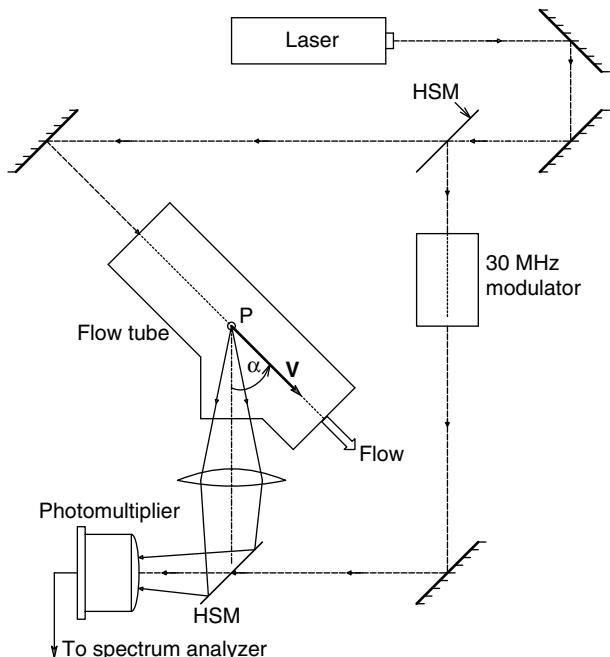
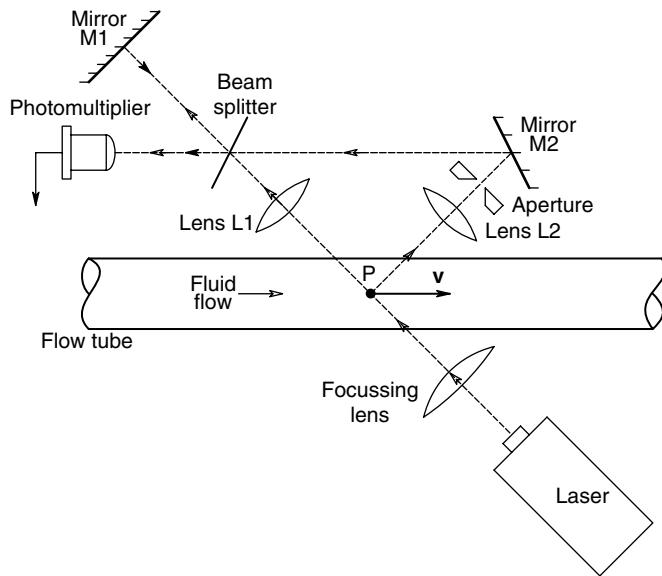
**FIGURE 6.80**

Diagram of the system used by Yeh and Cummins (1993) to first measure fluid velocity by Doppler shift. In their system, $\theta = 0^\circ$ and $\alpha = 30^\circ$.

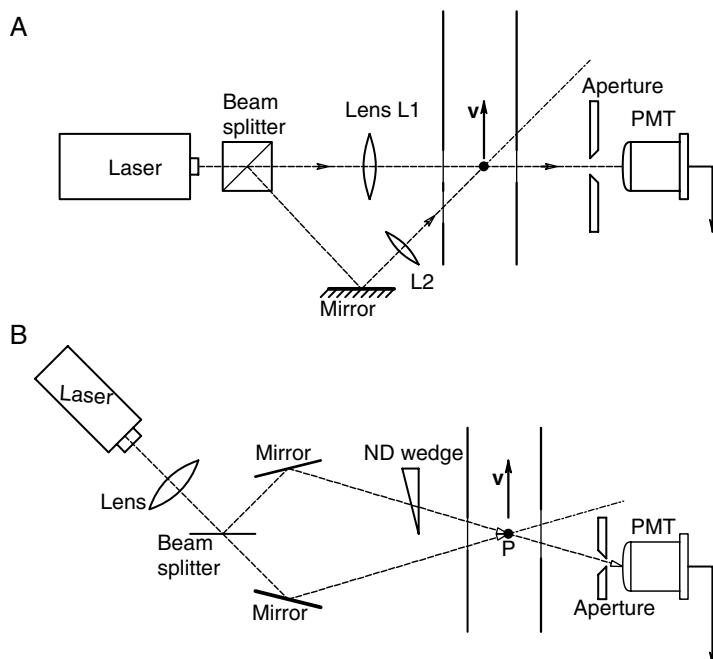
analogous to square law (heterodyne) detection in a radio, where two sine waves of different amplitudes and frequencies (f_s and f_r) are added together, then squared. This process produces sinusoidal terms at the PMT output with frequencies $2f_s$, $2f_r$, (f_s+f_r) , (f_r-f_s) and 0 (dc). Of interest in LDV is the difference frequency term. In the Yeh and Cummins system, $f'_s = f_s + 30\text{ MHz}$, so the difference term equals 30 MHz for $V=0$. The Doppler frequency shift, f_D , is added to the 30 MHz center frequency in the Yeh and Cummins system, which was used to measure the parabolic laminar flow profile in water. They could measure velocities as low as 0.007 cm/s, corresponding to a Doppler frequency shift of 17.5 Hz. The maximum velocity they measured was 0.05 cm/s, corresponding to a Doppler shift of 125 Hz. The minimum detectable f_D was 10 Hz, or one part in 10^{14} !

Many architectures for LDV systems have been described. See, for example, texts by Watriasiewicz and Rudd (1976) and Thompson and Stevenson (1979). Figure 6.81 illustrates a ‘single beam laser anemometer’ described by Foreman *et al* (1993). The laser is focused on a micro-region of the liquid whose velocity is to be measured. Part of the light is scattered and forms the signal beam and part is transmitted to form the reference beam. Through the use of lenses, slits, mirrors and a half silvered mirror, the signal and reference beams are combined at the surface of the PMT’s photocathode. This system was effectively used to measure both laminar flow in a liquid and the flow of air. No particles needed to be added to the liquid, but smoke particles were required to obtain adequate signal strength in the air measurements. Air velocities of up to 1000 fps were measured, corresponding to $f_D=33\text{ MHz}$. Foreman *et al* found that the signal energy due to scattering was several hundred times stronger for small scattering angles in the forward direction, rather than for light scattered back along the input path. It should be noted that the Foreman system design, which picks off the signal beam at a shallow forward angle, suffers from the disadvantage that the laser and the PMT are on opposite sides of the fluid flow path.

The dual beam LDV design, shown in Figure 6.82A, was first reported by Goldstein and Hagen (1993). Note that the laser output (reference) beam is perpendicular to V and

**FIGURE 6.81**

The LDV system of Foreman *et al* (1993). A so-called single-beam anemometer.

**FIGURE 6.82**

A. A dual-beam LDV system described by Goldstein and Hagen (1967). B. Another dual-beam LDV system, described by Goldstein and Kreid (1967).

passes directly to the PMT through aperture A. The signal beam irradiates the scattering particle at angle, θ . Some of the scattered signal light also passes through aperture A and mixes with the reference beam at the PMT photocathode. Observe that the dual beam LDV is similar to the single beam system, except that the laser and the PMT are interchanged. Goldstein and Kreid (1993) used this system to measure the air flow profile at the inlet of a square duct (Figure 6.82B).

There are many other considerations in the design of LDV systems which we will only touch upon; space considerations do not permit us to treat them in detail. First, we have noted that the detected LDV signal originates from the interaction of particles suspended in the moving fluid with the laser beams. Particle size, density, and concentration in the fluid medium are important considerations in LDV system design. Particles are generally assumed to be randomly distributed in the sampling volume. Under conditions of low Reynolds number flow (laminar flow, no turbulence), one can generally assume that the particle velocity is the same as the fluid velocity. However, in oscillating or turbulent flows, particles generally do not follow the fluid without some lag, creating measurement errors. As an example of the error caused by particle lag, consider the force on a smoke particle in moving air. From the Stokes drag formula [Watrasiewicz and Rudd (1976)], we can write the drag force on the particle in one direction as:

$$F_d = 3\eta d(v_p - v_f) \quad (6.187)$$

where η is the dynamic viscosity of the gas, d is the diameter of the smoke particle, v_p is its absolute velocity and v_f is the velocity of the fluid. The Newtonian equation of motion for a particle of density ρ is, thus:

$$F = ma = (\pi d^3 / 6) \rho \dot{v}_p = -3\eta d(v_p - v_f) \quad (6.188)$$

Equation 6.188 is an ordinary, first order, linear differential equation in v_p and if there is a step change in fluid velocity of size v_{f0} , then the ODE's solution is:

$$v_p(t) = v_{f0}[1 - \exp(-t/\tau)] \quad (6.189)$$

The particle time constant is

$$\tau = (\pi d^2 \rho) / (18 \eta) \text{ s} \quad (6.190)$$

This analysis implies that the smoke particles lag behind gas velocity changes with first order dynamics. The situation is actually far more complicated. A 1 μm diameter smoke particle with a density of 2.5 will have a τ of 20 μs , and will follow velocity fluctuations well at frequencies up to about 8 kHz. Some other particles used in LDV are silicone oil aerosols in air (density difference 900), magnesium oxide in flame (density difference 1.8×10^4), PVC spheres in water (density difference 1.54), polystyrene spheres in water (density difference 1). A 0.8 μm diameter particle with a relative density of 10^3 will follow 10 kHz velocity fluctuations with 1% error.

The calculation of the minimum laser power to achieve a given SNR in an LDV system is ordinarily quite complex, and we refer the interested reader to the many texts on LDV. Watrasiewicz and Rudd (1976) derived expressions for the SNR at the photosensor as a function of incident power. For example, a PMT photosensor will give a 40 dB SNR for 10 μW power incident on the detector. The same 10 μW incident power gives a 50 dB SNR for *pin* and avalanche photodiode sensors. At incident light, powers below 10^{-7} W , the PMT generally has a better SNR than the photodiodes.

LDV systems are used in a wide variety of applications, ranging from meteorology (studies of wind shear, clear air turbulence, tornadoes, etc.), studies of air flow in automotive streamlining, respiratory physiology, wind tunnel studies in aircraft design, investigation of laminar flows with recirculation and turbulent flow in ducts, studies of combustion dynamics in furnaces and gas turbines and investigations of blood flows.

6.13 Chapter Summary

In this chapter, we have examined various basic physical mechanisms underlying the operation of sensors and sensor systems used in making mechanical, chemical, physical and electrical measurements. In some cases, we have also described certain sensor applications in order to give a better understanding of the sensor mechanisms. Many sensors are seen to work through the QUM, causing changes in resistance or capacitance. Other sensors involve the generation of voltages through the separation or release of charges (electrons, holes, ions); these include piezoelectric, pyroelectric, Faraday and Hall effect sensors. Still other sensors make use of certain physical wave effects (e.g. the Doppler and the Sagnac effects).

Electrochemical sensors are seen to involve ion-specific chemical half-cells (batteries), polarographic electrochemical reactions, such as used in the Clark oxygen cell, and fuel cell (redox) reactions. The outputs of electrochemical sensors are voltages or currents proportional to the concentration of the QUM.

Radiation sensors were shown to be basically event indicating devices, where the radiation event generates ionized atoms or molecules, which are then detected by a variety of means—optical (by scintillation) or electrical (transient currents resulting from ionization in crystals or gasses). Photomultiplier tubes and channel photomultipliers were shown in Section 6.8 to convert EM radiation (NIR through X-rays) to electron currents, and are able to count single photons under special conditions.

In Section 6.11, we considered ‘artificial noses’ or manmade sensors, for detecting and quantifying vapor phase analytes, such as those emanating from controlled substances (e.g. heroin, cocaine, marijuana), from explosives (e.g. C4, Semtex), and from everyday substances such as meats, vegetables, perfumes, etc. The physiological basis for olfaction in vertebrates was summarized and the canine olfactory system sensitivity compared to manmade vapor phase sensors was given. Manmade e-nose sensors were described and the means for feature extraction in mixtures of scents was reviewed. The bottom line though, is dogs still beat e-noses.

Sagnac effect sensing of vehicle angular velocity in inertial space was derived and laser Doppler velocimetry was introduced in Section 6.12.

Our survey of sensor mechanisms in this chapter focused on the more important principles whereby physical quantities are converted to electrical signals. Our listing is by no means exhaustive. The next Chapter (7) deals with the various means by which we can measure specific physical quantities, using various types of transducers, devices and systems.

Problems

6.1 Two wire strain gauges are bonded to the outside of a SCUBA air tank in order to measure the circumferential strain when it is filled. With atmospheric pressure in the tank (zero strain), the circuit of Figure P6.1 is balanced so that $V_o = 0\text{ V}$, with all resistors $R = R_{27} = 300\ \Omega$.

- (A) Find an expression for $V_o = f(R, V_B, \Delta R)$ for the circuit. Assume the op-amp is ideal.
- (B) When the tank is pressurized to 2,500 psi, it is strained. Also, its temperature rises to 67°C due to adiabatic air compression. Calculate the tank’s strain, ε , at 2,500 psi

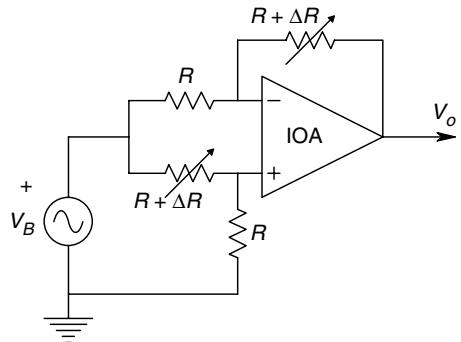


FIGURE P6.1

filling pressure. Assume $\Delta R = R_{27}[\varepsilon GF + \alpha(T - 27)]$, $V_o = -35.4$ mV. The excitation voltage is $V_B = 5$ V; the strain gauges have $GF = 2.9$ and $\alpha \text{ temp cos } = 1 \times 10^{-4}$.

- (C) The air in the tank is now rapidly exhausted to atmospheric pressure. Due to air expansion, the tank temperature falls below 27°C . The output voltage is now $V_o = 10$ mV. Find the tank's temperature.
- 6.2 A thin, intrinsic silicon slab is used as a photoconductor to sense 633 nm laser light intensity. The slab measures $0.5 \times 0.05 \times 0.001$ cm. The light hits the 0.5×0.05 cm surface. Electrodes are attached to the two 0.5×0.001 cm surfaces. One electrode goes to a -10 V dc supply, the other to the summing junction of an ideal op-amp, shown in Figure P6.2.

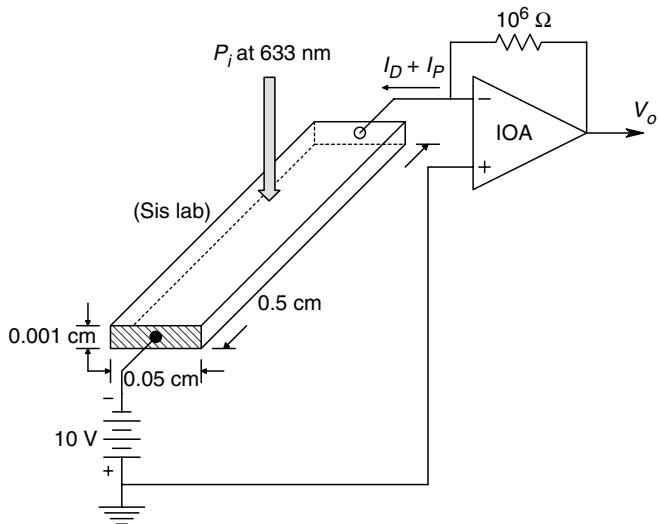


FIGURE P6.2

- (A) Calculate the dark current through the Si photoconductor. The resistivity at room temperature for Si is $\rho = 2.30 \times 10^3 \Omega \text{ cm}$ Find $V_{o(\text{dark})}$.
- (B) What absorbed power (radiant flux), P_i at 633 nm light in watts, on the photoconductor will give a photocurrent equal to the dark current? Assume $\mu_n = 1500 \text{ cm}^2/(\text{Vs})$, $\mu_p = 450 \text{ cm}^2/(\text{Vs})$, $\tau_p = 10^{-4}$, $\lambda = 6.33 \times 10^{-5} \text{ cm}$, $q = 1.60 \times 10^{-19} \text{ C}$, $\eta = 0.8$ (efficiency), $c = 3 \times 10^{10} \text{ cm/s}$, $h = 6.624 \times 10^{-34} \text{ Js}$ (Planck's

constant). What radiant intensity in W/cm^2 does this absorbed power correspond to? The photoconductor's photoconductance is given by:

$$G_p = \frac{q\eta\tau_p(\mu_p + \mu_n)}{L^2} \frac{[P_i\lambda]}{[hc]} \text{ S}$$

- (C) Show how you can modify the circuit so that V_o will be zero in the dark.

6.3 A pyroelectric IR sensor is connected in the short circuit current mode to an electrometer op-amp as shown in Figure P6.3. The sensor is irradiated with a rectangular pulse of Φ_o W IR power for T seconds. The Laplace transform of the input pulse is:

$$\Phi_i(s) = \Phi_o(1 - e^{-sT})/s$$

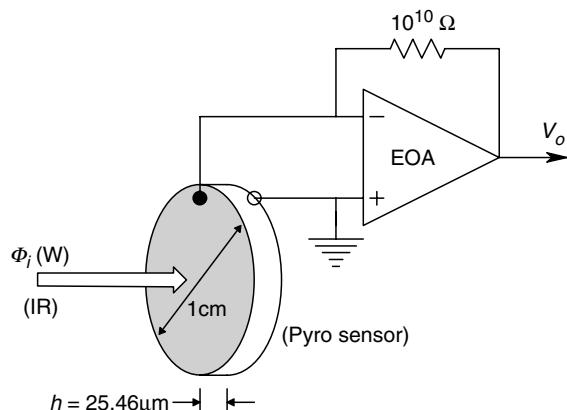


FIGURE P6.3

The sensor is 1 cm in diameter, $h = 25.46 \mu\text{m}$ (thickness), $K_p = 60 \mu\text{C}/(\text{m}^2 \text{ K})$ (pyroelectric constant), $c = 2.34 \times 10^6 \text{ J}/(\text{m}^3 \text{ K})$ (specific heat), $\Theta = 200 \text{ K/W}$ (thermal resistance).

- (A) Find the sensor's thermal time constant (note units).
 (B) Find an expression for $v_o(t)$. Plot and dimension $v_o(t)$. Let $\Phi_o = 100 \text{ nW}$, $T = 3 \text{ sec}$, $R_F = 10^{10} \Omega$.

6.4 A lead-zirconate-titanate (LZT) piezoelectric disk is used to make a 'crystal' microphone. The LZT disk has an effective area of 1 cm^2 , a compression piezo-strain coefficient of $d_{33} = 320 \text{ pCb/N}$, $\kappa = 1,500$ (dielectric constant), and $h = 3 \text{ mm}$ (thickness). The total leakage conductance of the sensor = 10^{-11} S .

- (A) Calculate the capacitance of the sensor, C_x . Assume a simple parallel plate capacitor model.
 (B) Draw the equivalent circuit of the microphone. Calculate the peak open circuit voltage across the microphone alone when a 1 kHz sinusoidal sound wave with peak pressure of 150 dynes/cm² impinges on the crystal.
 (C) Now the microphone is connected to the electrometer op-amp circuit as shown in Figure P6.4. Ten feet of coax are used which adds 300 pF in parallel to the sensor's C_x . Calculate the peak output voltage of the amplifier for the same sound input. Let $R_F = 10^{10} \Omega$, $C_F = 10^{-10} \text{ F}$.

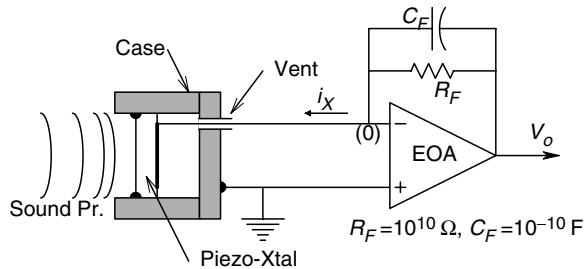


FIGURE P6.4

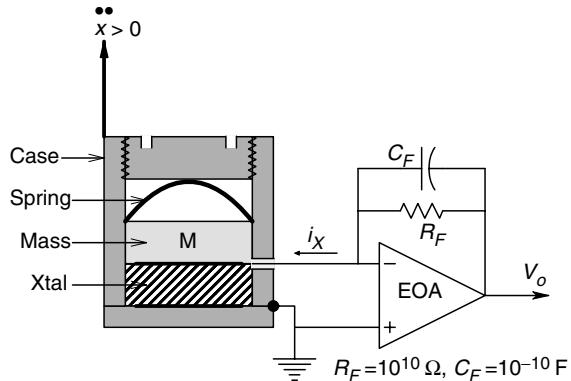


FIGURE P6.5

6.5 A piezoelectric accelerometer of the compression type has its output current conditioned by an electrometer op-amp charge amplifier as shown in Figure P6.5. The proof mass in the accelerometer is 10 grams, the piezo crystal's d coefficient is 140 pCb/N . The crystal's capacitance, $C_x = 500 \text{ pF}$, its leakage conductance $G_x = 10^{-12} \text{ S}$.

- (A) The accelerometer is given a vertical step of acceleration of 1 g. Find, sketch and dimension $v_o(t)$.
- (B) The piezoelectric accelerometer is attached to the tip of a helicopter rotor blade undergoing a static vibration test. At one part of the test, the blade tip vibrates in the steady state at 10 Hz in the vertical plane with a peak sinusoidal amplitude of 10 cm. Calculate the peak amplitude of $v_o(t)$ for this condition.

6.6 A piezoelectric crystal accelerometer is used to measure the horizontal deceleration of a vehicle in order to deploy its airbags (Figure P6.6). The airbags are to deploy if the deceleration exceeds $5g = 5 \times 9.8 \text{ m/s}^2$. A charge amplifier conditions the accelerometer current output. A comparator senses $v_o(t)$ over a limit, V_R , and triggers airbag inflation.

- (A) Draw the equivalent circuit of the accelerometer and amplifier.
- (B) Assume the vehicle is traveling at a velocity of 5.0 m/s (11.19 mph) when it hits the crash test barrier. Its velocity decreases linearly to zero in 100 ms. Graph and dimension the velocity $v(t)$, dv/dt and d^2v/dt^2 .
- (C) Calculate, plot and dimension the amplifier output, $v_o(t)$ for the crash. What must the value of the dc comparator reference voltage, V_R , be so that the comparator

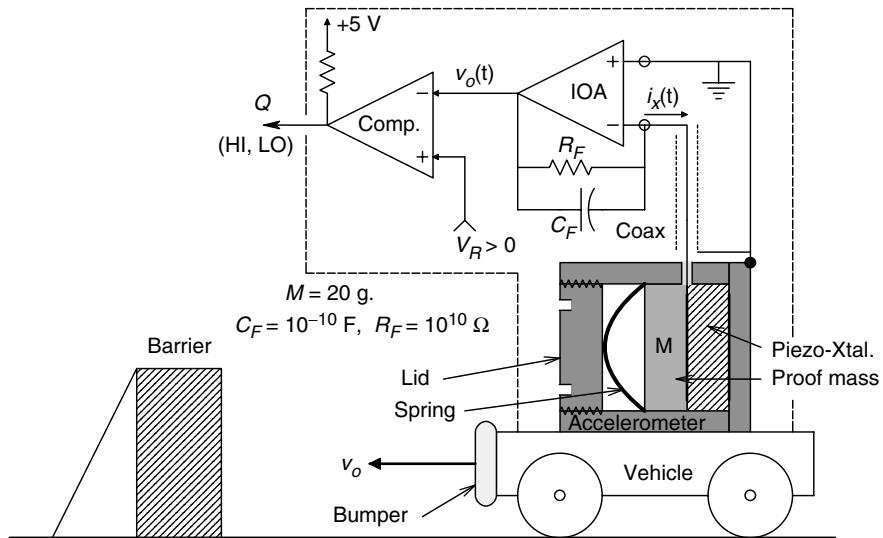


FIGURE P6.6

output goes LO and triggers the airbags for this crash. Assume that the crystal's polarity is such that the initial output, $v_o(0+)$, is positive. The crystal's $d = 150 \text{ pC/N}$ (Note the Laplace transform pair: $1/(s + a) \leftrightarrow (e^{-at})$).

6.7 A simple linear accelerometer is made by suspending a mass M in a case from a cantilever spring arm. Two strain gauges are cemented to the cantilever arm. The strain on the cantilever at the gauges is proportional to its relative displacement, $\varepsilon = K_k(x_1 - x_2)$. x_1 is the case displacement, x_2 is the mass displacement inside the case. The case is filled with oil to give the accelerometer viscous damping, D . A Wheatstone bridge circuit (Figure P6.7B) gives an output, $V_o = (V_1 - V_2)$, proportional to the strain, ε . For downward deflection of the cantilever, the top strain gauge is under tension, and the bottom gauge is under compression. Thus for the top gauge, $R_{\text{top}} = R_o(1 + \varepsilon GF)$ and for the bottom gauge, $R_{\text{bot}} = (1 - \varepsilon GF)$.

- (A) Write the Newtonian force balance ordinary differential equation for the accelerometer. Laplace transform it to find $\frac{X_2}{X_1}(s)$.

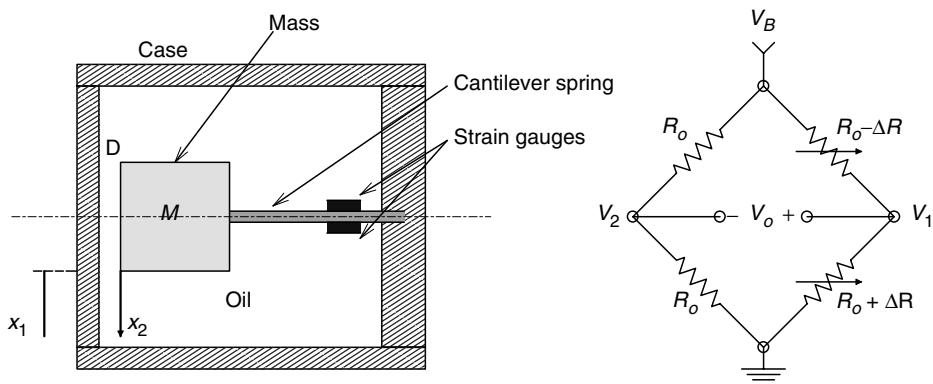


FIGURE P6.7

- (B) Derive an expression for the strain, $\varepsilon(s)$, as a function of $\ddot{X}_1(s)$.
 (C) Give an expression for $V_o(s)$ as a function of $\ddot{X}_1(s)$.
- 6.8 In the circuit of Figure P6.8, a silicon pin photodiode is used to sense the output of an optical system powered by a 532 nm (green) laser diode. The dark leakage current of the diode is 5 nA at 27°C, the op-amp is ideal.

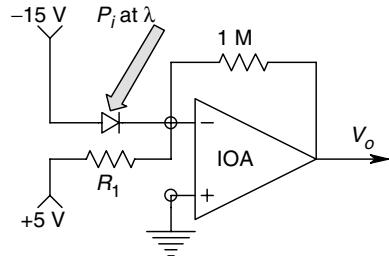


FIGURE P6.8

- (A) Find the R_1 value required to make $V_o = 0$ in the dark.
 (B) What optical power incident on the photodiode is required to produce $V_o = 1.0 \text{ V}$? Assume $\eta = 0.5$.
- 6.9 In the circuit of Figure P6.9, a silicon photodiode is operated in the zero current mode ($i_D = 0$). Let $I_{rs} = 1 \text{ nA}$, $V_T = 0.026 \text{ V}$, $I_P = 0.25 P_i$ (photocurrent at 512 nm wavelength).

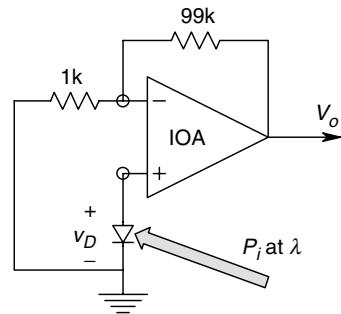


FIGURE P6.9

- (A) Sketch the general form of the $i_D = f(v_D, P_i)$ curves for the photodiode.
 (B) Derive an expression for $v_D = g(P_i)$.
 (C) Sketch and dimension a plot of V_o vs P_i .
- 6.10 In the circuit of Figure P6.10, a disk pyroelectric sensor is used to sense changes in LIR radiation. The exposed area is 1 cm in diameter, its thickness is $h = 25.46 \mu\text{m}$, its pyroelectric coefficient $K_p = 25 \mu\text{C}/(\text{m}^2 \text{K})$ its specific heat is $c = 2.5 \times 10^6 \text{ J/m}(\text{m}^3 \text{K})$ its thermal resistance is $\Theta = 20 \text{ K/W}$ and 50% of the incident IR radiation is absorbed by the sensor and 50% is reflected away.
- (A) Calculate the time constant of the sensor.
 (B) Assume a step input of LIR power to the sensor. What input power is required to produce a peak output voltage of 100 mV?

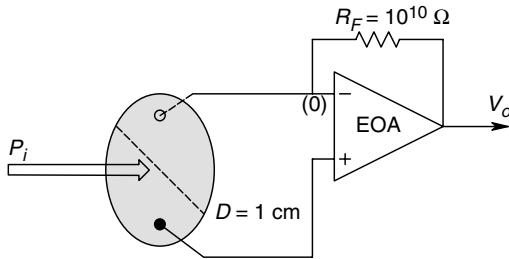


FIGURE P6.10

(C) Sketch and dimension $v_o(t)$, given the step input of (B).

- 6.11 In Figure P6.11 a Si *pin* photodiode is used to sense a 10 kHz, sinusoidally intensity modulated, 633 nm light beam. To improve output signal-to-noise ratio, the op-amp is followed by a bandpass filter centered at 10 kHz, and having a 1 kHz noise

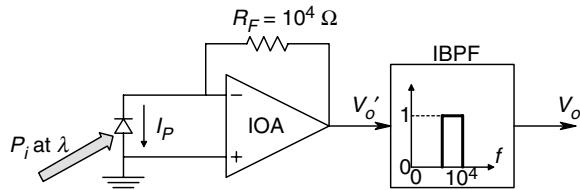


FIGURE P6.11

bandwidth. In the dark, the RMS output noise at V_o is $v_{on} = 0.5 \times 10^{-6}$ RMS V. Use $h = 6.624 \times 10^{-34}$ Js, $q = 1.60 \times 10^{-19}$ C, $c = 3 \times 10^8$ m/s, $\eta = 0.55$. The input light power is described by:

$$P_i(t) = 0.5 P_{ipk}[1 + \sin(2\pi 10^4 t)] \text{ W}$$

The photocurrent is given by:

$$I_P(t) = P_i(t)[\eta q / (hv)]$$

- (A) Give an expression for and sketch and dimension $V_o'(t)$.
 (B) Give an expression for and sketch and dimension $V_o(t)$.
 (C) Find the P_{ipk} value so that the RMS $V_o = 0.5 \times 10^{-6}$ V.
- 6.12 The feedback system illustrated in Figure P6.12 keeps the root mean square value of the ac current through the variable capacitor constant by changing the frequency at constant RMS V_c to compensate for changes in C . $C = C_o + \Delta C$. Instead of a VCO, a voltage-to-period converter (VPC) is used in which the output sinewave's period, T , is proportional to V_o . That is, $T = K_p V_o + b$ and $f_c = 1/T$ Hz. In the steady state, the integrator forces $V_e \rightarrow 0$, so the RMS $V_c = \text{RMS } V_1$.
- (A) Derive an expression for V_o at null.
 (B) What must b equal to so that $V_o \propto \Delta C/C_o$?

- 6.13 Figure P6.13 illustrates a circuit of a manually nulled, ΔC sensing circuit. The ac radian frequency is made $\omega = 1/(RC_o)$. The ideal op-amp makes an all-pass, phase shifting circuit. Its input is $V_1 \angle 0^\circ$, its output is $V_o = V_1 \angle \theta$. R_2 is manipulated to

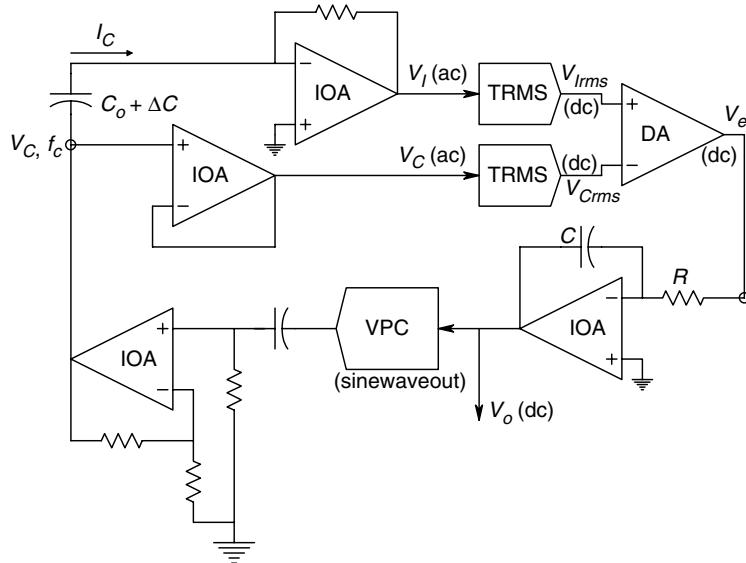


FIGURE P6.12

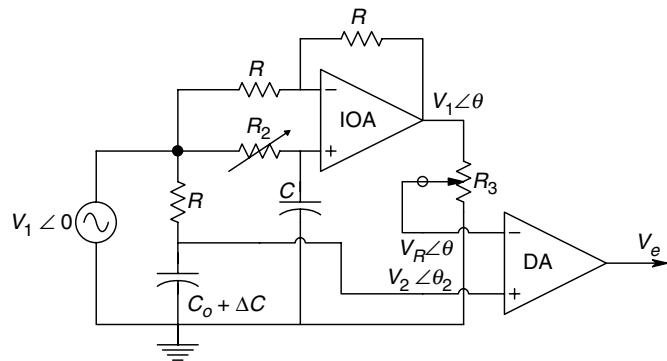


FIGURE P6.13

match the phase of \mathbf{V}_o to \mathbf{V}_2 . R_3 adjusts the amplitude of \mathbf{V}_o to match the amplitude of \mathbf{V}_2 .

- (A) Derive an expression for $\frac{V_o}{V_1}(j\omega)$.
- (B) Derive an expression for the magnitude of \mathbf{V}_e as a function of $\Delta C/C_o$ and other circuit parameters. Assume $\mathbf{V}_e = 0$ for $\Delta C/C_o = 0$ and $|\Delta C/C_o| \ll 1$ in the denominator of $|V_e|$.

6.14 A differential, commutated capacitor bridge (cf. Section 5.5.1.6) is used to convert the $\Delta C/C_o$ of a differential capacitive sensor to a dc output voltage (Figure P6.14). In the circuit, MOS switches periodically connect $C_o + \Delta C$ and $C_o - \Delta C$ to alternately discharge through the small resistors, R_d , or charge through the larger resistors, R_c . When $C_o + \Delta C$ is connected to R_d , $C_o - \Delta C$ is connected to R_c and vice versa. The bridge output voltage, $v_o(t)$, is the difference between $v_1(t)$ and $v_2(t)$. Assume each capacitor discharges completely through R_d in negligible time. Also, when connected to R_c , each capacitor charges slowly toward V_B . The switching clock has period T . Note that $R_d C_o \ll T/2 \ll R_c C_o$. Although the charging capacitor vol-

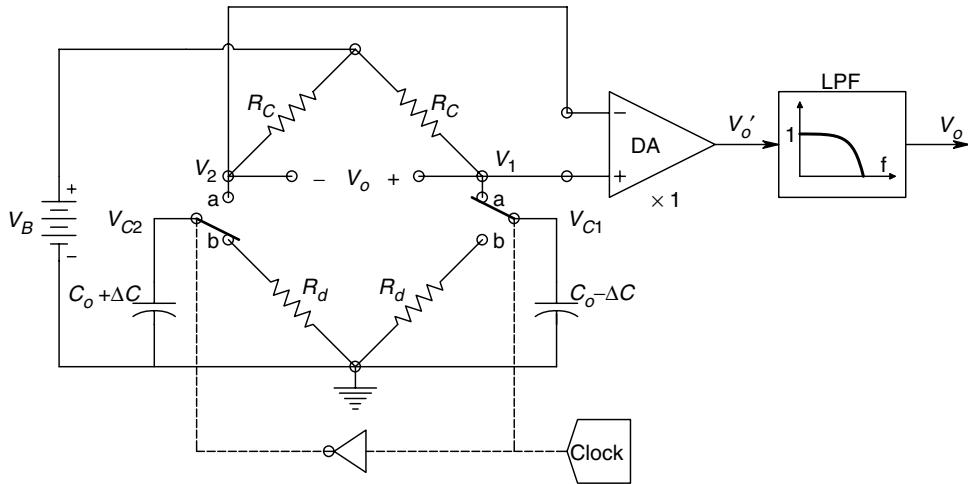


FIGURE P6.14

tage is an exponential, because $T/2 \ll R_c C_o$, it may be treated as a linear (triangular) waveform over $0 \leq t \leq T/2$. Thus you can use the approximation: $e^{-x} \approx 1 - x$.

- (A) Sketch and dimension the steady state for $v_{c1}(t)$, $v_{c2}(t)$, $v_1(t)$, $v_2(t)$ and $v_o(t)$ over several clock periods.
- (B) $v_o(t)$ is passed through a unity gain low-pass filter to give an output proportional to its average (dc) value. Find an expression for the average value of $v_o(t)$ in terms of V_B , bridge parameters, and $\Delta C/C_o$.

- 6.15 Figure P6.15 describes a self-nulling system used to measure small changes in a capacitance caused by the presence of foreign objects near the test capacitor's plates. With no foreign object present, $\Delta C \rightarrow 0$, V_3 is set to zero and the reference capacitor is set to C_o so that ΔI into IOA1's summing junction is zero. Thus V_1 and

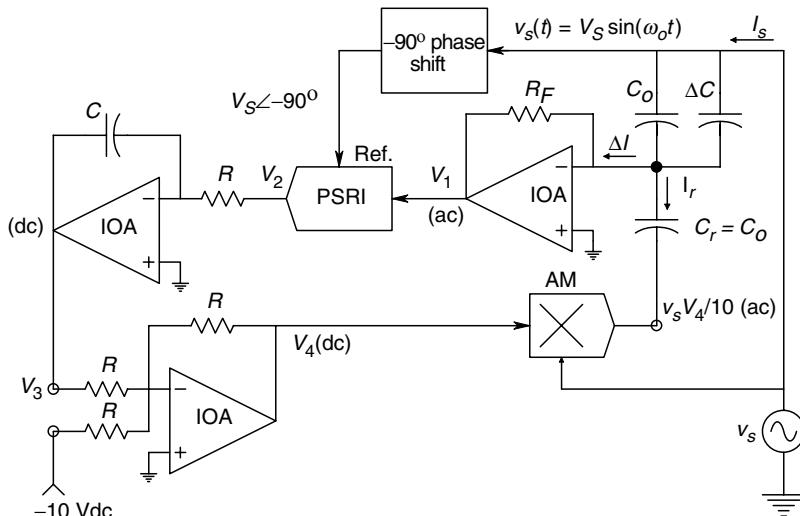


FIGURE P6.15

V_2 are zero. The output of the summer (IOA3) is $V_4 = +10$ V dc, so the multiplier output is $-v_s(t)$. In the figure, PSR=phase sensitive rectifier, AM=analog multiplier.

When a foreign object is present, it increases the sensing capacitor's capacitance by ΔC . Now $|I_s| > |I_r|$ and ΔI , V_1 (ac) and V_2 (dc) all increase. The integrator output V_3 (dc) goes negative by a small amount. This causes V_4 (dc) to increase, making v_m (ac) more negative, increasing $|I_r|$ until $\Delta I \rightarrow 0$ again.

- (A) Write a general algebraic expression for $\Delta I(j\omega)$ (system not at null).
- (B) Write a general expression for $v_1(t)$.
- (C) Sketch and dimension $v_2(t)$. Why is the -90° phase shift necessary for the PSR sync?
- (D) At null, $\Delta I \rightarrow 0$. Derive an expression for V_3 (dc) at null in terms of ΔC , C_o and other system parameters. (V_3 is the system output.)

- 6.16 In Figure P6.16, a capacitive sensor is used to make a low-pass filter. The radian frequency of the sinusoidal source is made $\omega = 1/(RC_o)$, $v_s(t) = V_s \sin(\omega t)$.

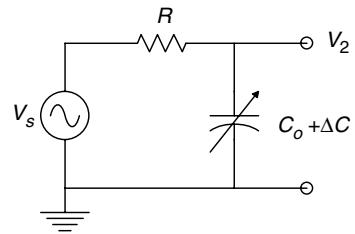


FIGURE P6.16

- (A) Give an expression for $\frac{V_2}{V_s}(j\omega)$ in time-constant form.
- (B) Find an expression for the transfer function's magnitude sensitivity, $S_{\Delta C}^{|V_2|}$, defined as $\frac{d|V_2|}{d\Delta C}$.
- (C) Find an expression for the transfer function's phase sensitivity, $S_{\Delta C}^\phi$, defined as $\frac{d\phi}{d\Delta C}$.

- 6.17 An ideal but noisy op-amp is used to condition the signal from a silicon photoconductor (PC) light sensor. Resistor R_D is used to compensate for the photoconductor's dark current, I_D . In the circuit, $e_{na} = 3 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 0.4 \text{ pA}/\sqrt{\text{Hz}}$, $R_F = 10^6 \Omega$, $R_D = 2.3 \times 10^5 \Omega$ and $V = 10$ Vdc. Since the light is chopped, the op-amp is followed by an (ideal) noiseless BPF with center frequency of $f_c = 400$ Hz, pass gain = 1 and bandwidth $B = 100$ Hz. See Figure P6.17.

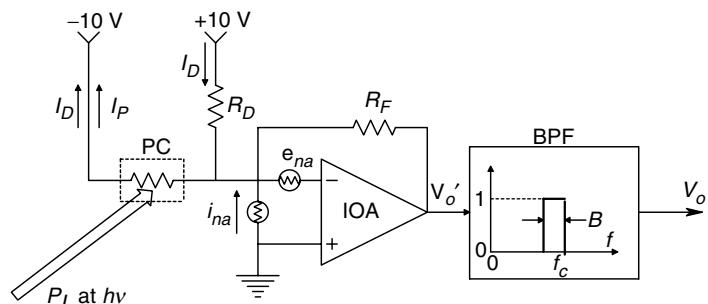


FIGURE P6.17

- (A) Derive an algebraic expression for, and calculate the numerical value of, the RMS noise output voltage, v_{on} , in the dark. Assume all resistors make thermal white noise and $T = 27^\circ\text{C}$.
- (B) Give an expression for V'_o as a function of the photocurrent, I_p . What is the PC's dc dark current.
- (C) The photocurrent may be shown to be given by $I_p = P_L \eta V \tau_p (\mu_p + \mu_n) / (l^2 h v)$ A, where $l = 0.01 \text{ cm}$, $h v = 1.2 \text{ eV}$, $\tau_p = 10^{-4} \text{ s}$, $(\mu_p + \mu_n) = 1,800 \text{ cm}^2/(\text{Vs})$, $\eta = 0.8$. What photon power, P_L , in watts must be absorbed by the PC in order that $V'_o = v_{on}$ from part A)? Find the ratio of the I_p at this P_L to I_D .
- 6.18 A square pyroelectric sensor is used to sense square wave chopper LIR light. The chopping has a 50% (ON/OFF) duty cycle. The circuit is shown in Figure P6.18. The sensor is 3 mm on a side and 10 μm thick. The specific heat is $c = 2.5 \times 10^6 \text{ J/m}^3 \text{ K}$. The pyroelectric constant is $K_p = 25 \mu\text{C/m}^2 \text{ K}$ and $\Theta = 20^\circ/\text{W}$.

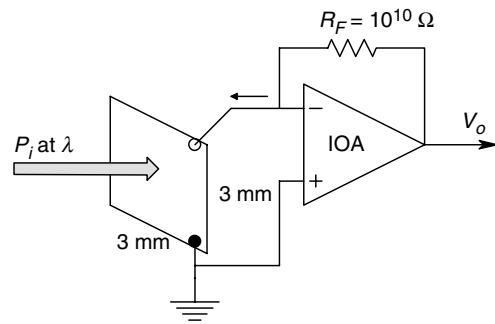


FIGURE P6.18

- (A) Calculate the sensor's thermal time constant, τ_{th} .
- (B) Find the P_i value required to give a 1 V peak V_o initially.
- (C) Assume the input square wave of peak height P_i found in (B) and period of 2 ms is applied to the sensor for a long enough time for the system to be in the steady state. Sketch and dimension $v_o(t)$ in the steady state, showing the maxima and minima.
- 6.19 A 1 lb weight is suspended from a 0.005 in diameter wire. The unstrained resistance of the wire, R_o , is 270Ω , and its unstrained length is $L_o = 36 \text{ in}$. The wire's gauge factor, GF, is 3.2.
- (A) Calculate the psi tensile stress on the loaded wire.
- (B) Assume Young's modulus is $3.2 \times 10^7 \text{ psi}$. Find the strain on the wire, and ΔR .
- (C) Assume a Wheatstone bridge incorporates the wire as one arm. The other arm resistances = R_o . The bridge is excited with 5 V dc. Find V_o due to the 1 lb load.
- 6.20 The Lion circuit is used to condition the output of differential capacitor sensors. It is shown in Figure P6.20. If we assume that the diodes are ideal, the waveform at the v'_o node can be represented by:

$$v'_o(t) = (V_S/2)[1 - \exp(-t/\tau_1)] \quad \text{for } V_S = +12 \text{ V.}$$

$$v'_o(t) = (-V_S/2)[1 - \exp(-t/\tau_2)] \quad \text{for } V_S = -12 \text{ V.}$$

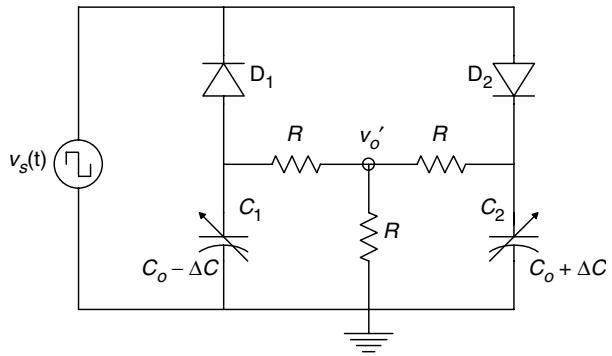


FIGURE P6.20

where $C_1 = C_o - \Delta C$, $C_2 = C_o + \Delta C$ and $\tau_1 = C_1 R (3/2)$, $\tau_2 = C_2 R (3/2)$. Also, we assume that τ_1 and $\tau_2 \gg T/2$ of the input squarewave, $v_s(t)$.

- (A) Derive an algebraic expression for the average $v_o'(t)$ in the steady state. Let $R = 10^6 \Omega$, $C_o = 30 \text{ pF}$, $T = 10^{-5} \text{ s}$, $V_S = 12 \text{ Vpk}$. Evaluate the system's sensitivity in dc V/($\Delta C/C_o$).
- (B) Use SPICE or MicroCap to simulate the Lion bridge output, $v_o'(t)$. Use a $\pm 12 \text{ V}$, 100 kHz squarewave input, 1N4148 diodes, let $\Delta C/C_o = 10 \text{ pF}/30 \text{ pF}$. Place a 10 fF capacitor across the v_o' node to ground (this prevents register overflow, and does not affect the results of the simulation). Examine 5 cycles of the output.

- 6.21 The RMS magnitude of a 60 Hz magnetic field vector in a small volume of 3-dimensional space is to be measured. Three, identical, orthogonal, Hall sensors are used, as shown in Figure P6.21A. The ac output voltage of each sensor is proportional to the perpendicular component of the \mathbf{B} field through that sensor. For example, $v_{Hy}(t) = K_H B_y(t) = K_H |B| \sin(377t) \sin(\beta)$. (The dc current through each Hall sensor is not shown.)

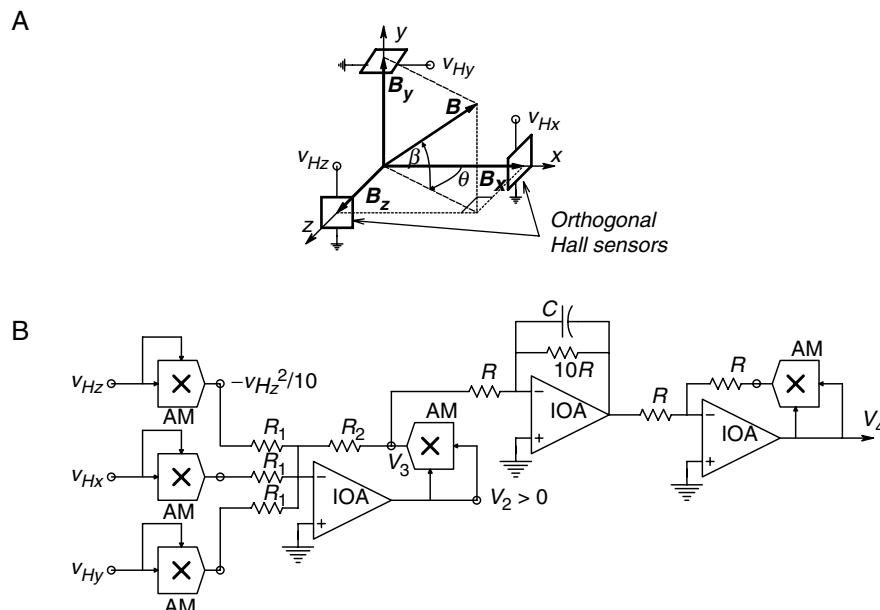


FIGURE P6.21

- (A) Give expressions for $v_{Hx}(t)$ and $v_{Hz}(t)$.
- (B) Consider the resolver circuit of Figure P6.21B. Derive an expression for $v_2(t)$, $v_3(t)$ and V_4 .
- 6.22 An analog multiplier and a low-pass filter are used to demodulate the output of a LVDT length sensor, as shown in Figure P6.22. The excitation voltage is $v_s(t) = V_S \sin(6283t)$. The output voltage is $v_o(t) = K_S x(t) \sin(6283t - \theta)$. The low-pass filter has unity dc gain and removes all signals of carrier frequency and above from $v_m(t)$.

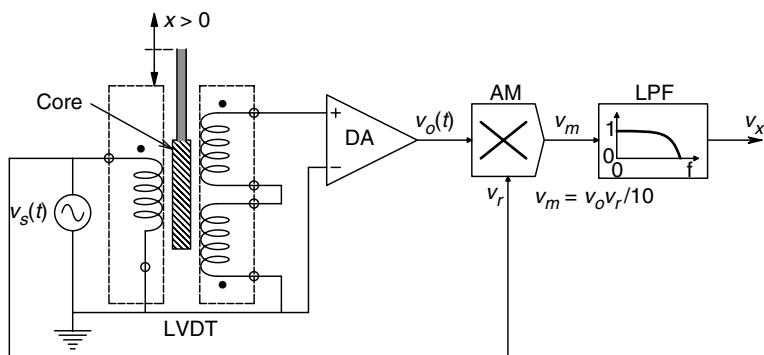


FIGURE P6.22

- (A) Derive an expression for $v_x(t)$.
- (B) Design a passive or active circuit that can be put in the v_r line to make the reference phase agree with the phase shift in $v_o(t)$. The peak value of $v_r'(t)$ should remain the same.

7

Applications of Sensors to Physical Measurements

7.1 Introduction

In this chapter, we examine various means available to measure certain physical quantities. In Chapter 6, we have reviewed various transducer mechanisms underlying the measurement system function, such as changes in resistance. (Many physical quantities cause resistive changes, such as temperature, strain (caused by force or pressure), light intensity, humidity, etc.) Our purpose in this chapter is to examine the most commonly employed sensor types used to measure a given quantity, such as temperature. The quantities which we shall consider here are—acceleration, velocity and position (angular and rectilinear), temperature, pressure, force, torque, gyroscopic measurement of angular velocity and position, the GPS system, and an introduction to spectrophotometry.

7.2 Measurement of Angular Acceleration, Velocity and Displacement

There are many applications where it is important to know information about the attitude of a vehicle in inertial space. In the case of aircraft, instrumentation of the roll, pitch and yaw are critical for the design of robust, closed loop, fly-by-wire autopilots and flight stabilization systems. Similar knowledge is important for the control of high speed hydrofoils, hovercraft and submarines. Drone surveillance aircraft, guided missiles and ‘smart’ bombs also require feedback on their roll, pitch and yaw to perform their tasks effectively. In addition to vehicle stabilization and guidance, measurement of angular acceleration, velocity and position are also important in the design of constrained mechanical systems which face variable loads and inputs, for example, automobile and truck suspension systems.

In this chapter, we describe some of the common sensor systems used to measure angular acceleration, angular velocity and angular position.

7.2.1 Angular Acceleration Measurement

In this section, we describe several means by which angular acceleration can be measured. One is tempted to obtain angular acceleration by simply differentiating the output of a rate gyroscope. While theoretically correct, this procedure will generally result in poor resolution and a poor SNR, because the process of differentiation enhances

the noise present in the angular velocity signal. It is generally better to measure angular acceleration directly, as described below.

7.2.1.1 Angular Acceleration Measurement with a Constrained Mechanical Gyro

The inner gimbals of a mechanical gyro (Section 7.2.2 discusses the development of the linear ODEs describing the behavior of mechanical gyroscopes) is restrained by a stiff, piezoelectric twister crystal as shown in Figure 7.1A. We may assume $\dot{\theta}$ and $\ddot{\theta}$ are zero, and the torsional moment for the inner gimbals is given by the gyro equation, 7.26a, as:

$$M_\theta = -H_r \dot{\phi} \quad (7.1)$$

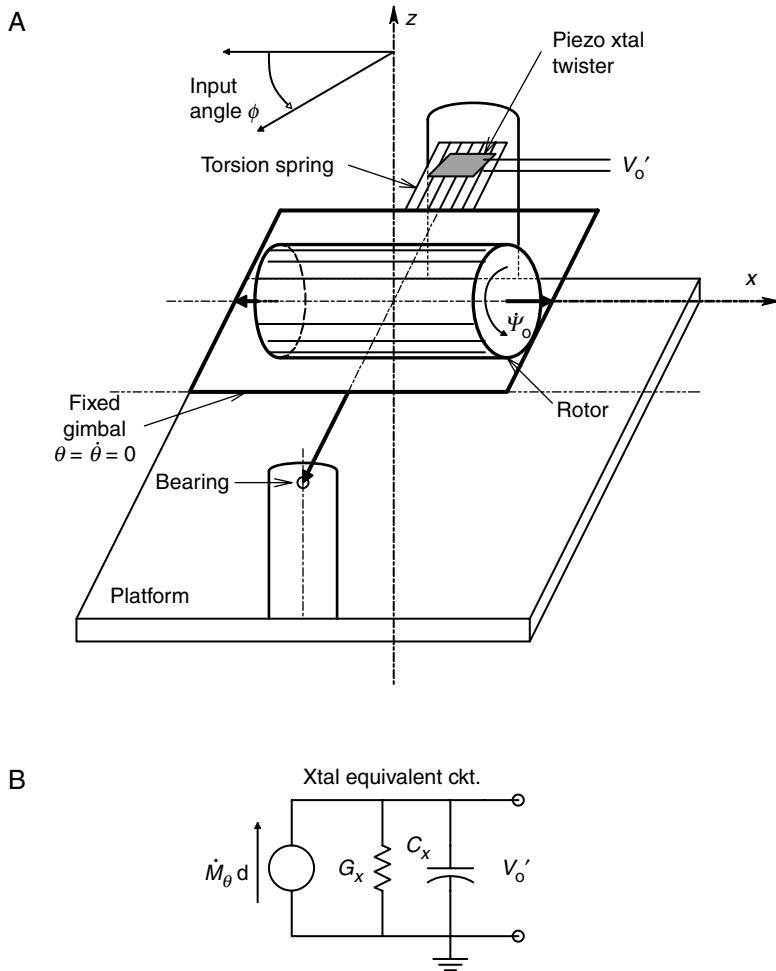


FIGURE 7.1

A. Simple mechanical gyroscope with restrained gimbal. Gimbal torque is proportional to the input angular velocity, $\dot{\phi}$. A piezoelectric 'twister' transducer senses the gimbal torque. B. Equivalent circuit of a piezoelectric transducer at mechanical frequencies well below the transducer's mechanical resonance frequency.

The twister piezoelectric transducer has an equivalent circuit given by Figure 7.1B at mechanical frequencies well below its mechanical resonance frequency. A net charge is displaced by the mechanical torque, $q = M_\theta d$, (d is a constant). Current is simply the rate of flow of charge, so the equivalent current source in the model for the transducer is $i_x = -H_r \ddot{\phi} d$ and the output voltage of the transducer is:

$$V_o(s) = \frac{H_r R_x d \ddot{\Phi}(s)}{(s C_x R_x + 1)} \quad (7.2)$$

Thus, the steady state output of the transducer is proportional to the angular acceleration of the gyro around the ϕ axis. If a charge amplifier is used (Section 2.6), the time constant in the output transfer function becomes $C_F R_F$, the feedback C and R of the charge amplifier. A far simpler means of measuring is shown below.

7.2.1.2 Simple Inertia Wheel-Spring-Dashpot Angular Accelerometer

Figure 7.2 illustrates a simple inertia wheel angular accelerometer in which the output voltage is proportional to the relative position of the wheel in the case, given by:

$$V_o = K_o(\phi_R - \phi_C) \quad (7.3)$$

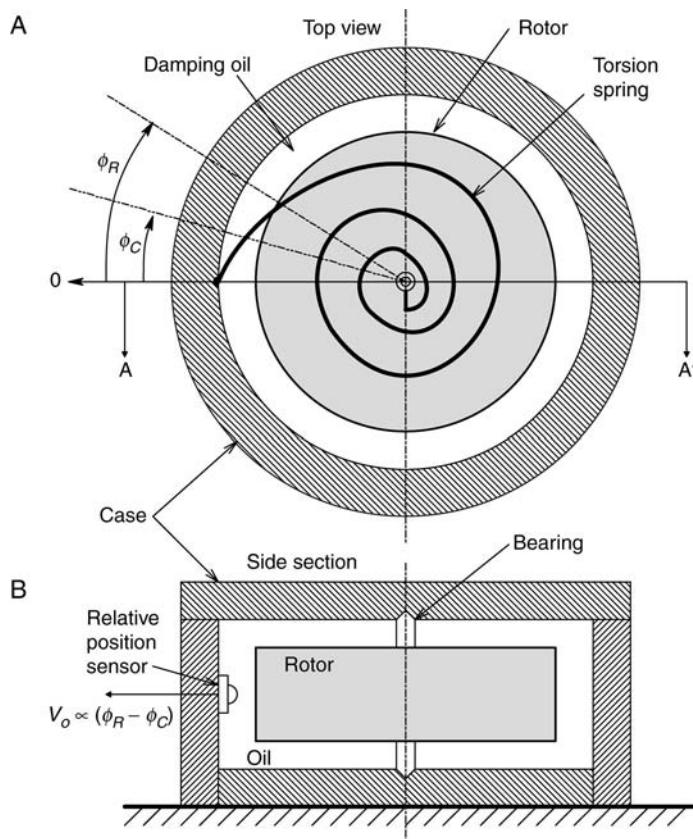


FIGURE 7.2

Two views of an inertia wheel angular accelerometer. A. Top view. Wheel position in the case can be sensed photooptically, or by a potentiometer. The inertia wheel is restrained by a helical spring. B. Side sectional view.

The torque equation is:

$$J\ddot{\phi}_R + B(\dot{\phi}_R - \dot{\phi}_C) + K_s(\phi_R - \phi_C) = 0 \quad (7.4)$$

Here, B is the viscous friction between the wheel and the case and K_s is the torsion spring constant. If we Laplace transform the ODE of equation 7.4 and solve for ϕ_R , then substitute into the expression for V_o , we obtain

$$\frac{V_o}{\Phi_{\ddot{C}}}(s) = \frac{K_o J / K_s}{s^2 J / K_s + sB / K_s + 1} \quad (7.5)$$

Thus, the steady state output of this simple angular accelerometer is $V_o = \ddot{\phi}_C K_o J / K_s$. Such an accelerometer should be well damped so that the rotor wheel will not oscillate. The accelerometer's damping factor is easily seen to be $\zeta = B/(2\sqrt{J K_s})$ and should be adjusted to about 0.5 by manipulating B , to obtain a rapid transient response without excessive overshoots. The natural frequency of this accelerometer is $\omega_n = \sqrt{K_s/J}$ r/s. Thus, extending the accelerometer's bandwidth decreases its sensitivity, a good example of a gain bandwidth trade-off. To see how it is possible to maintain high sensitivity and bandwidth, we will examine an active accelerometer design.

7.2.1.3 A Servo Angular Accelerometer

In this design, shown in Figure 7.3, a torque motor is used as an active spring. The torque motor has a flat torque speed characteristic over most of its speed range, the torque being

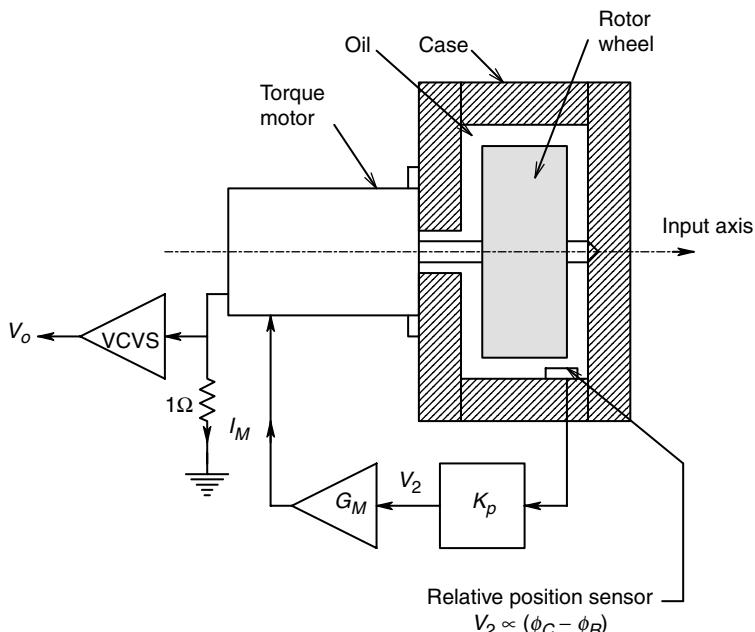


FIGURE 7.3

Side section through a servo angular accelerometer. The torque motor current, I_M , may be shown to be proportional to the angular acceleration. G_M = transconductance amplifier, K_p , K_o = gains.

proportional to the motor's input current, I_M . The accelerometer output voltage is made proportional to I_M . The torque balance equation for the servo accelerometer is:

$$T_M = K_T I_M = J\ddot{\phi}_R + B(\dot{\phi}_R - \dot{\phi}_C) \quad (7.6)$$

Also, we have:

$$V_2 = K_P(\phi_C - \phi_R) \quad (7.7)$$

$$I_M = G_M V_2 \quad (7.8)$$

$$V_o = K_O I_M \quad (7.9)$$

Combining the equations above, it can be shown that:

$$\frac{V_o}{\Phi_C}(s) = \frac{K_O/K_T}{s^2 J/(K_T K_P G_M) + sB/(K_T K_P G_M) + 1} \quad (7.10)$$

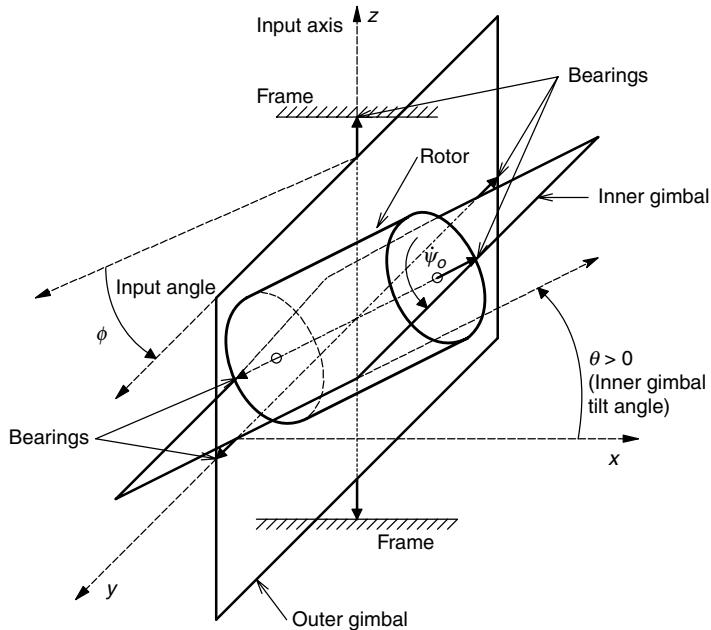
Now the natural frequency governing the high frequency response is set by the electrical parameters K_T , K_P and G_M . In this design, K_O can be made large to compensate for a high K_T , thus preserving sensitivity and bandwidth.

7.2.2 Angular Velocity Measurement with Mechanical Gyroscopes

The angular velocity of vehicles, such as that associated with the roll, pitch or yaw rate of aircraft, missiles, boats, submarines or spacecraft needs to be measured in order to control vehicle attitude in inertial space, and in the case of certain modern hypersonic aircraft, stabilize their flight under all conditions. The principal means of sensing angular velocity is by rate gyro. A rate gyro can use the traditional electromechanical spinning rotor, or be opto-electronic, using the Sagnac effect, introduced in Section 6.12.2, or use a variety of novel mechanisms, including an air jet, vibrating tuning fork, or a vibrating mass or inertia wheel. In this section, we will first discuss conventional mechanical gyroscopes and then describe the design and sensitivity limitations of fiber optic, Sagnac effect gyros.

The innards of a conventional, 2-gimbal, mechanical gyroscope are illustrated in Figure 7.4. The rotor is spun electrically at a high, constant angular velocity, typically as high as 24,000 RPM for 400 Hz excitation. The rotor is held in an inner frame, or gimbal, which is in turn suspended inside an outer gimbal. In early (e.g. 1950s), gyroscopes used in autopilots, etc., the angle between the inner and outer gimbals, θ , was sensed by a wiper on a wire wound resistor potentiometer. Resolution, therefore, was limited by the diameter of the resistance wire *vs* the area of the wiper. In our analysis which follows, we will assume infinite resolution in sensing the gyro tilt angle, θ .

Referring to Figure 7.4, we will define ϕ as the input angle. If the rotor is spinning in the direction shown with angular velocity ψ , a positive (CCW) increase in ϕ will result in the gyroscopic effect of having the inner gimbals and rotor tip up to the right, giving a $\theta > 0$ (CCW rotation). The detailed analysis of the dynamics of a mechanical gyro results in three coupled, highly nonlinear differential equations. Analysis proceeds smoothly if we define three, mutually orthogonal coordinates for the gyro system. These are in fact, the angular momentum axes of the rotor and inner gimbals. \mathbf{H}_1 is the angular momentum vector component perpendicular to the rotor spin axis and the plane of the inner gimbals. \mathbf{H}_2 is the angular momentum vector of the inner gimbals and rotor as they pivot on

**FIGURE 7.4**

Schematic diagram of a two gimbal, mechanical gyroscope. The input angle, ϕ , is a CCW rotation of the outer gimbal around the z axis.

the outer gimbals. \mathbf{H}_2 is in the plane of the inner gimbals and is perpendicular to the rotor spin axis and \mathbf{H}_1 . \mathbf{H}_3 is the angular momentum vector of the rotor, along the rotor axis and perpendicular to the other two momentum vectors. The direction of the angular momentum vectors is given by the right hand rule—the fingers curl in the direction of rotation and the thumb points in the vector direction.

The best way to describe the dynamics of a mechanical gyroscope is by the Euler-Lagrange equations [Aseltine, 1958; Cannon, 1967]. These equations can be derived from Newton's laws and provide a general means of describing the motion of complex mechanical systems having rotating or fixed coordinate systems. The Lagrange equations are of the form:

$$\frac{d}{dt} \left[\frac{\partial U}{\partial x_i} \right] - \frac{\partial U}{\partial x_j} = M_i, \quad j = 1, 2, 3, \dots N \quad (7.11)$$

where U is the total kinetic energy (KE) of the system, x_j is the j th general coordinate, and M_j is the j th generalized reaction force or torque. The differential work done in causing a small change in the coordinates can be written as:

$$dW = \sum_j M_j dx_j \quad j = 1, 2, 3, \dots N \quad (7.12)$$

In the case of the gyroscope shown in Figure 7.4, x_1 will be taken as the rotor angle, ψ , x_2 is defined as the inner gimbals' angle, θ and x_3 is the input angle, ϕ . The total kinetic energy of the gyro is the sum of the KEs associated with each orthogonal coordinate and can be written as:

$$U = (1/2)J_r[\dot{\psi} + \dot{\phi} \sin(\theta)]^2 + (1/2)J_1\dot{\theta}^2 + (1/2)J_1[\dot{\phi} \cos(\theta)]^2 \quad (7.13)$$

Here, J_r is the moment of inertia of the rotor around its spin axis and J_1 is the moment of inertia of the rotor and inner gimbals around the axes normal to the spin axis.

From equation 7.11, we can write the first nonlinear ODE for the system. Consider $j=1$, so $x_1 = \psi$. Thus it is easy to show that:

$$\frac{\partial U}{\partial \psi} = 0 \quad (7.14)$$

$$\frac{\partial U}{\partial \dot{\psi}} = J_r \dot{\psi} + J_r \dot{\phi} \sin(\theta) \quad (7.15)$$

$$\frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\psi}} \right] = J_r \ddot{\psi} + J_r \ddot{\phi} \sin(\theta) + J_r \dot{\phi} \dot{\theta} \cos(\theta) \quad (7.16)$$

Hence the Lagrange equation for the gyro rotor spin axis is:

$$M_\psi = J_r [\ddot{\psi} + \ddot{\phi} \sin(\theta) + \dot{\phi} \dot{\theta} \cos(\theta)] \quad (7.17)$$

The rotor friction moment, M_ψ , is usually taken as zero. The rotor is run at constant (synchronous) velocity, so that $\ddot{\psi} = 0$. Now, the second Lagrange equation is for $x_2 = \theta$ the inner gimbals' angle. In this case, we have:

$$\frac{\partial U}{\partial \theta} = J_r \dot{\psi} \dot{\phi} \cos(\theta) + J_r \dot{\phi}^2 \sin(\theta) \cos(\theta) - J_1 \dot{\phi}^2 \cos(\theta) \sin(\theta) \quad (7.18)$$

$$\frac{\partial U}{\partial \dot{\theta}} = J_1 \dot{\theta} \quad (7.19)$$

$$\frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\theta}} \right] = J_1 \ddot{\theta} \quad (7.20)$$

Hence, the Lagrange equation for the θ coordinate is:

$$M_\theta = J_1 \ddot{\theta} - J_r \dot{\psi} \dot{\phi} \cos(\theta) - J_r \dot{\phi}^2 (\cos \theta) + J_1 \dot{\phi}^2 \cos(\theta) \sin(\theta) \quad (7.21)$$

The third Lagrange equation is written for $x_3 = \phi$, the input angle.

$$\frac{\partial U}{\partial \phi} = 0 \quad (7.22)$$

$$\frac{\partial U}{\partial \dot{\phi}} = J_r \dot{\psi} \sin(\theta) + J_r \dot{\phi} \sin^2(\theta) + J_1 \dot{\phi} \cos^2(\theta) \quad (7.23)$$

$$\begin{aligned} \frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\phi}} \right] &= J_r \ddot{\psi} \sin(\theta) + J_r \dot{\psi} \dot{\theta} \cos(\theta) + J_r \ddot{\phi} \sin(\theta) \\ &\quad + 2J_r \dot{\phi} \dot{\theta} \sin(\theta) \cos(\theta) - 2J_1 \dot{\phi} \dot{\theta} \cos(\theta) \sin(\theta) + J_1 \ddot{\phi} \cos^2(\theta) \end{aligned} \quad (7.24)$$

The Lagrange equation for ϕ is just:

$$M_\phi = \frac{d}{dt} \left[\frac{\partial U}{\partial \dot{\phi}} \right] \quad (7.25)$$

To linearize the gyro ODEs, equations 7.21 and 7.25, we note again that $\ddot{\psi} = 0$ and we assume that the system operation is such that $|\theta| < 15^\circ$, so that $\sin(\theta) \cong \theta$ in radians. Also, $\cos(\theta) \cong 1$ and the terms containing products of the numerically small quantities, $\dot{\phi}$, $\ddot{\phi}$, $\dot{\theta}$ and $\ddot{\theta}$ are negligible. Under these approximations, we finally obtain the two simultaneous, linear ODEs which approximately describe the dynamic behavior of the gyro.

$$M_\theta = J_1 \ddot{\theta} - H_r \dot{\phi} \quad (7.26a)$$

$$M_\phi = H_r \dot{\theta} + J_1 \ddot{\phi} \quad (7.26b)$$

H_r is the angular momentum of the rapidly spinning rotor, $J_r \dot{\psi}$.

Under most conditions, the outer gimbals' angle, ϕ , is considered to be a rotation of the vehicle about a principal axis, such as its roll axis. The response of the roll gyro having its outer gimbals' axis pointing in the direction of the roll axis is considered to be the inner gimbals' angle, θ , or in some cases, the torque, M_θ . The linear gyro differential equations above can be Laplace transformed and written as:

$$M_\theta(s) = J_1 s^2 \Theta(s) - H_r s \Phi(s) \quad (7.27a)$$

$$M_\phi(s) = H_r s \Theta(s) + J_1 s^2 \Phi(s) \quad (7.27b)$$

We shall use these linear equations to describe the dynamics of the rate gyro, and later, the behavior of the rate integrating gyro (RIG).

7.2.2.1 The Mechanical Rate Gyro

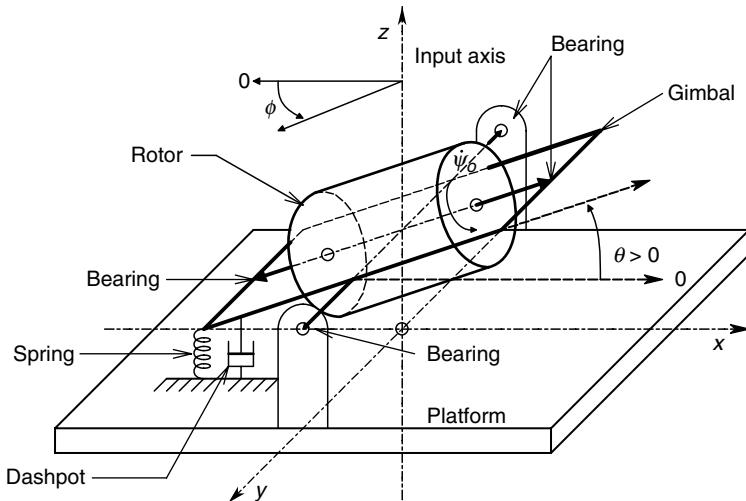
In the one-gimbal mechanical rate gyro, the gimbal's motion is restrained by a spring with torque constant, K_s , and a viscous damper (dashpot) with rate constant B , as shown in Figure 7.5. In this gyro, we have dispensed with the outer gimbal and assume that the gyro base rotates an angle ϕ as shown, around the input (z) axis. The rotor gimbal's moment is thus given by:

$$M_\theta = -\Theta(s)[Bs + K_s] \quad (7.28)$$

The gyro equations can now be written as:

$$0 = \Theta(s)[s^2 J_1 + sB + K_s] - H_r s \Phi(s) \quad (7.29a)$$

$$M_\phi = \Theta(s) s H_r + J_1 s^2 \Phi(s) \quad (7.29b)$$

**FIGURE 7.5**

Schematic diagram of a one gimbal, mechanical, rate gyro. In the steady state, the gimbal angle, $\theta = \dot{\phi}H_r/K_s$.

Using Cramer's rule, we can solve equations 7.29a and b for the relationship between θ and $\dot{\phi}$.

$$\frac{\Theta}{\dot{\Phi}}(s) = \frac{\begin{vmatrix} 0 & -sH_r \\ M_\phi & s^2J_1 \end{vmatrix}}{\begin{vmatrix} s^2J_1 + sB + K_s & 0 \\ sH_r & M \end{vmatrix}} = \frac{M_\phi sH_r}{M_\phi(s^2J_1 + sB + K_s)} = \frac{sH_r/K_s}{s^2J_1/K_s + sB/K_s + 1} \quad (7.30)$$

This transfer function is easily seen to be a relation between the gimbal's tilt angle, θ , and the angular velocity of the gyro around the input axis, $\dot{\phi}$:

$$\frac{\Theta}{\dot{\Phi}}(s) = \frac{H_r/K_s}{s^2J_1/K_s + sB/K_s + 1} \quad (7.31)$$

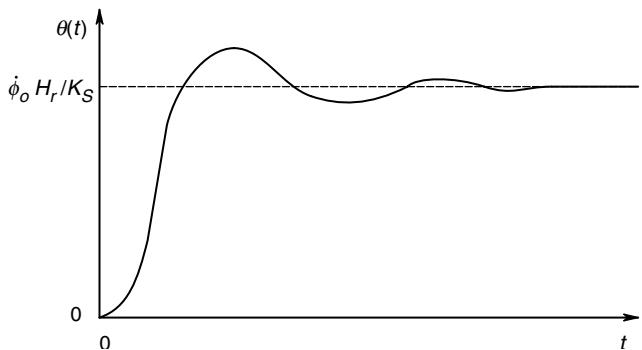
For a constant velocity input,

$$\dot{\Phi}(s) = \dot{\phi}_o/s \quad (7.32)$$

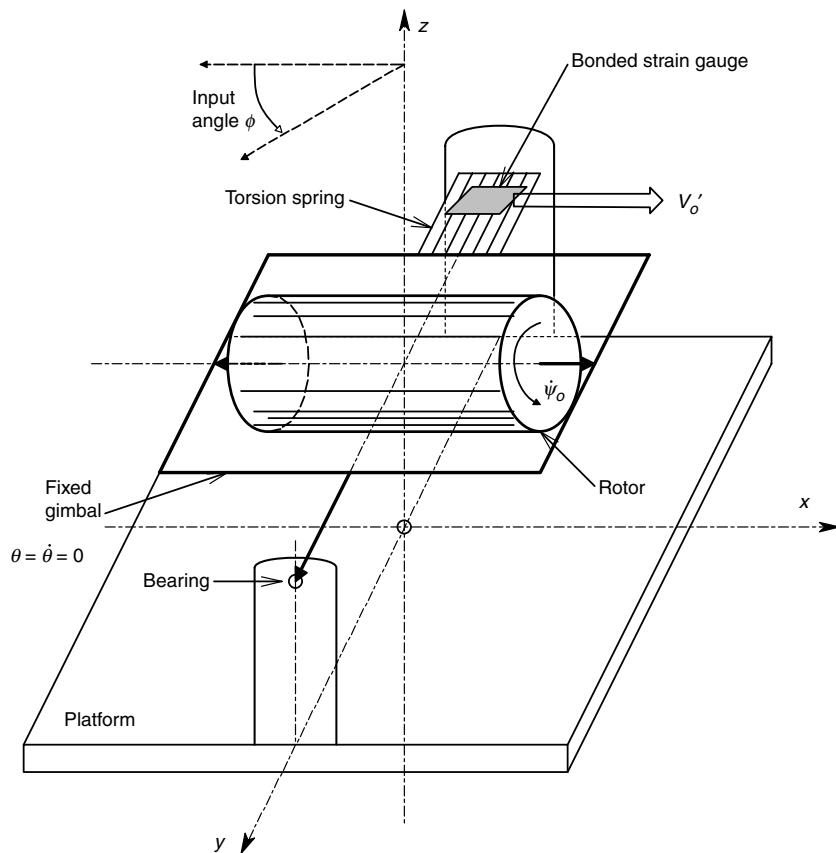
and the output angle is given by:

$$\Theta(s) = \frac{\dot{\phi}_o(H_r/K_s)}{s(s^2J_1/K_s + sB/K_s + 1)} \quad (7.33)$$

In the time domain, the response $\theta(t)$ is shown in Figure 7.6. It is a classic second order, low-pass system step response, with steady state value, $\theta_{ss} = \dot{\phi}_o H_r / K_s$. Thus, in response to a constant angular velocity input, the rate gyro rotor tips a steady state angle, θ_{ss} . θ can be measured by a low friction potentiometer wiper or by a frictionless optical coding disk.

**FIGURE 7.6**

Second order response of the rate gyro to a step input of angular velocity,
 $\phi(t) = \dot{\phi}_0 U(t)$.

**FIGURE 7.7**

A constrained gimbal rate gyro. The gimbal torque is measured by a strain gauge, and is proportional to $\dot{\phi}$. Compare this system to the constrained gyro angular accelerometer shown in Figure 7.1.

Another approach to rate gyro design is shown in Figure 7.7. Here, the rotor gimbal is constrained not to rotate by a very stiff torsion spring. The spring has strain gauges cemented to it, arranged so that the torque exerted on the spring produces an output voltage. Refer to the linearized gyro equations, equations 7.27a and b. For practical purposes, we assume that, $\ddot{\theta} = \dot{\theta} = 0$. Thus equation 7.27a reduces to:

$$M_\theta(s) = -H_r s \Phi(s) \quad (7.34)$$

Hence, the output of the strain gauge bridge attached to the torsion spring can be written as:

$$v_o(t) = K_B H_r \dot{\phi} \quad (7.35)$$

Thus, we see that the rotor gimbal does not need to move appreciably in order to sense the $\dot{\phi}$. Note that the $\dot{\phi}$ term is multiplied by a $\cos(\beta)$ term in the relations above when the gyro is rotated around some axis at an angle β with the input or z axis.

7.2.2.2 Sagnac Effect Fiber Optic Gyroscopes

In Section 6.9.2, we derived the relation for interference fringe shifts as a function of rotation rate for a Sagnac interferometer. The fractional fringe shift, Δz , (generally measured electro-optically) was shown to be related to the input rotation rate by equation 6.174. When we include the cosine term for off-axis rotation, the relation for Δz becomes:

$$\Delta z = \dot{\phi} \cos(\beta)[4A/c\lambda] \quad (7.36)$$

Equation 7.36 above describes the fractional fringe shift for a single optical circuit. Sensitivity of the Sagnac gyroscope can be increased by increasing the number of optical circuits—a convenient way of doing this is by using multiple turns of an optical fiber waveguide. A simple FO gyro is shown schematically in Figure 7.8. Since N turns of FO cable are used, the fringe shift equation is easily seen to become:

$$\Delta z = \dot{\phi} \cos(\beta)[4N A/c\lambda] = \dot{\phi} \cos(\beta)[2(N2\pi R)r/c\lambda] = \dot{\phi} \cos(\beta)[2L_F R/c\lambda] \quad (7.37)$$

where R is the radius of the FO coils, N is the number of turns, c is the speed of light on the optical fiber, λ is the wavelength of the light and L_F is the total linear length of the FO cable used. Practical considerations concerning the SNR limit the maximum number of turns used, or L_F . Light intensity from the laser or laser diode source is attenuated by the optical fiber as a function of path length. This attenuation is given by the well known relation,

$$P_o = P_i 10^{-\alpha L_F} \quad (7.38)$$

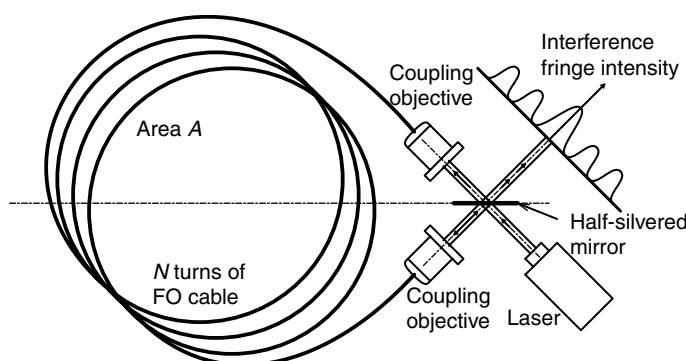


FIGURE 7.8

A simple, Sagnac effect, dc, fiber optic gyroscope.

Here, α is the attenuation constant, which is wavelength and fiber dependent, P_i is the input power in watts, P_o is the output power and L_F is the length of the fiber. The noise associated with the photodetector used to sense fringe shifts may be expressed in terms of an RMS uncertainty in the fringe shift, Δz_{RMS} . This noise is given by:

$$\Delta z_{RMS} = (1/2\pi) \sqrt{\frac{hv}{\eta P_o}} \sqrt{B} = (1/2\pi) \sqrt{\frac{hv}{\eta P_i 10^{-\alpha L_F}}} \sqrt{B} \quad (7.39)$$

Hence, the fringe shift RMS SNR can be written as:

$$\text{SNR} = \frac{\Delta z}{\Delta z_{RMS}} = \frac{\phi(2L_F R/c\lambda)}{(1/2\pi)\sqrt{(hvB/\eta)P_i 10^{-\alpha L_F}}} = \phi \frac{4\pi R \sqrt{P_i}}{c\lambda \sqrt{hvB/\eta}} (L_F 10^{-\alpha L_F})^{1/2} \quad (7.40)$$

where h is Planck's constant, v is the frequency of the light in Hz, B is the Hz noise bandwidth and η is the photodetector's quantum efficiency (typically 1/2). Regardless of ϕ , the SNR has a maximum with respect to L_F . This optimum L_F can be found by differentiating the right hand term in parentheses in equation 7.40 with respect to L_F and then setting the derivative equal to zero. This procedure yields:

$$L_{F(OPT)} = \frac{2}{\alpha \ln(10)} = (2/\alpha)(0.4343) \quad (7.41)$$

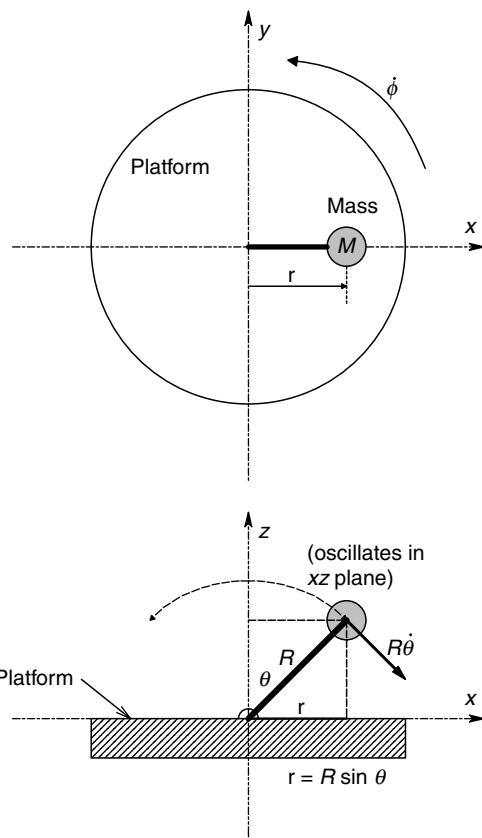
The optimum number of turns, N_{OPT} , can be found from:

$$N_{OPT} = INT[L_{F(OPT)}/2\pi R] = INT[0.13824/(\alpha R)] \quad (7.42)$$

Other embodiments of the Sagnac FO gyro have been developed to improve resolution by allowing more precise measurement of the fringe shift, and the carrier operation to avoid 1/f noise associated with dc and low frequency measurements. By introducing ac modulation of the light phase traversing the FO coil, the intensity of the interference fringes is also ac modulated, and fringe shift detection can be made more precise and noise free because photosensor outputs can be amplified out of the range of amplifier 1/f noise. The use of integrated coherent optical systems and single mode, polarization preserving optical fibers with attenuations (α) of less than 5 dB/km have permitted the design of Sagnac FO gyros with noise limited sensitivities of $> 10^{-2}$ degrees/hr using a 1 Hz noise bandwidth centered at the modulating frequency (10^{-2} degrees/hr is 5.56×10^{-6} degrees/s). By way of contrast, the earth's rotation rate is 4.167×10^{-3} degrees/s. Hence a FO gyro, with its input axis pointing north, can easily measure the earth's axial spin. Sagnac fiber optic gyros are the most sensitive angular rate sensors which have been developed.

7.2.2.3 The Vibrating Mass Rate Gyro

The first design we will consider is the single mass, vibrating mass rate gyro (VMRG). A schematic (top and side view) of the VMRG is shown in Figure 7.9. This particular

**FIGURE 7.9**

Top and side views of a vibrating mass rate gyro.

design is interesting because:

1. It illustrates the generality of the Lagrange method for analyzing complex electromechanical systems
2. By extending the single torsional, vibrating mass gyro to a paired structure in which the masses vibrate 180° out of phase, we can examine a crude model of the tuning fork rate gyro
3. The single VMRG forms a model of an ancient, biological rotational rate sensor system which evolved millions of years ago and acts to stabilize the flight of dipteran insects (flies, mosquitos).

Prehistoric insects had four wings like present day dragonflies. In the process of evolution, short bodied dipteran flies had their hind wings evolve into two (bilateral) organs known as halteres. A haltere has an inverted tenpin shaped structure, which is attached and pivoted at its small end from the insect's body below and behind each wing. It is basically a mass at the end of a rod which vibrates in one plane at the wing beat frequency when the insect is flying. Flies have many mechanoreceptor neurons in the elastic cuticle at the base of each haltere, which evidently sense strains in the cuticle resulting from the rotation rate induced gyro moments. The complex manner in which this sensory neural information is processed and used to control the flight of flies and mosquitoes remains to be described. However, we do know that if both halteres are removed, a fly is no longer capable of controlled flight.

To derive a mathematical description for the torques at the base of the vibrating mass, we use the well-known LaGrange method [Cannon, 1967] which is based on Newtonian mechanics. For a complex mechanical system we can write N equations of the form:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = M_k \quad (7.43)$$

where q_k is the k th, independent coordinate. The q_k can be rotational or translational. M_k is the k th moment (torque) if rotational dynamics are considered, or the k th force if the dynamics involve linear translation. T is the total kinetic energy of the system and U is the total potential energy of the system; $L \equiv T - U$. L is called the system's 'LaGrangian function' [Cannon, 1967].

Consider Figure 7.9 in which the linear flight velocity vector of a fly is along the x axis, and for simplicity, let the mass (haltere) vibrate only in the xz plane. In this example, the independent coordinates are θ , the angle of the vibrating mass in the xz plane with the z axis and ϕ , the input rotation around the z axis (yaw left is shown). The system's potential energy is $U = (1/2)K_\theta\theta^2 J$. K_θ is the spring constant of the torsion spring that restores the resting mass to align with the z axis. A torque motor (muscles) (not shown) keeps the mass vibrating sinusoidally in the xz plane, so that $\theta(t) = \theta_m \sin(\omega_m t)$. There are two components to the system's kinetic energy:

$$T = (1/2)\dot{J}_o\theta^2 + (1/2)M(r\dot{\phi})^2 \quad (7.44)$$

Here, J_o is the moment of inertia of the mass on the end of a rod of length R (haltere). It is well-known that $J_o = MR^2$. r is the projection of R in the xy plane. $r = R \sin(\theta)$. $(r\dot{\phi})$ is the tangential velocity of the mass M in the xy plane due to the input rotation rate, $\dot{\phi}$. Thus, the system's Lagrangian is:

$$\begin{aligned} L &= (1/2)\dot{J}_o\theta^2 + (1/2)M(r\dot{\phi})^2 - (1/2)K\theta^2 \\ &= (1/2)\dot{J}_o\theta^2 + (1/2)MR^2(1/2)[1 - \cos(2\theta)]\dot{\phi}^2 - (1/2)K\theta^2 \end{aligned} \quad (7.45)$$

And we have:

$$\frac{\partial L}{\partial \theta} = -(1/2)MR^2(1/2)[-2 \sin(2\theta)]\dot{\phi}^2 - K\theta = (1/2)MR^2 \sin(2\theta)\dot{\phi}^2 - K\theta \quad (7.46)$$

$$\frac{\partial L}{\partial \dot{\theta}} = J_o\dot{\theta} \quad (7.47)$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = J_o\ddot{\theta} \quad (7.48)$$

Hence, $M_\theta = J_o\ddot{\theta} + K\theta - (1/2)MR^2 \sin(2\theta)\dot{\phi}^2$ is the torque required to be produced by the muscles to maintain the haltere in simple harmonic motion.

Our interest is in the (instantaneous) torque around the pivot in the xy plane, M_ϕ , due to the input pitch angular velocity, $\dot{\phi}$. We can write:

$$\frac{\partial L}{\partial \dot{\phi}} = 0 \quad (7.49)$$

$$\frac{\partial L}{\partial \dot{\phi}} = M_r^2 \dot{\phi} = MR^2 \sin^2(\theta) \dot{\phi} = MR^2 \dot{\phi} (1/2)[1 - \cos(2\theta)] \quad (7.50)$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\phi}} \right) = MR^2 \ddot{\phi} (1/2) - MR^2 \ddot{\phi} (1/2) \cos(2\theta) + MR^2 \dot{\phi} \dot{\theta} \sin(2\theta) \quad (7.51)$$

Thus,

$$M_\phi = \ddot{\phi} (1/2) MR^2 - \ddot{\phi} (1/2) MR^2 \cos(2\theta) + (1/2) MR^2 \dot{\phi} \dot{\theta} \sin(2\theta) \quad (7.52)$$

Since $|2\theta_m| \gg 10^\circ$ in the haltere system, we cannot use the approximations $\sin(2\theta) \cong 2\theta$ and $\sin^2(\theta) \cong \theta^2$. The M_ϕ moment can be written:

$$\begin{aligned} M_\phi = & \ddot{\phi} (1/2) MR^2 - \ddot{\phi} (1/2) MR^2 \cos[2\theta_m \sin(\omega_m t)] \\ & + \dot{\phi} (1/2) MR^2 \theta_m \omega_m \cos(\omega_m t) \sin[2\theta_m \sin(\omega_m t)] \end{aligned} \quad (7.53)$$

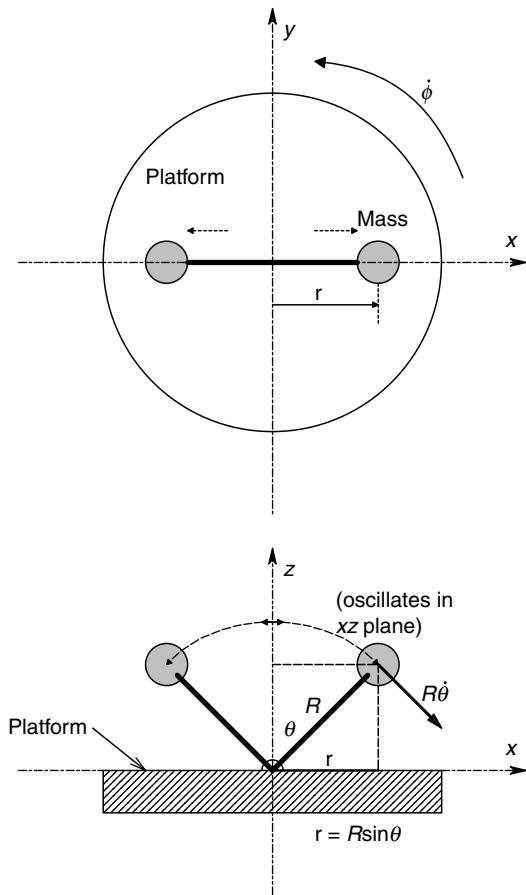
Thus the M_ϕ moment lies in the xy plane and contains a dc term proportional to $\ddot{\phi}$ and also double frequency terms in $\dot{\phi}$ and $\ddot{\phi}$. The desired angular velocity term is larger by an amount ω_m , which can be ≈ 900 r/s in mosquitoes.

Note that if a haltere were mounted so that it vibrated in the animal's xy plane, then it would appear to be responsive to angular acceleration and angular velocity due to roll and pitch.

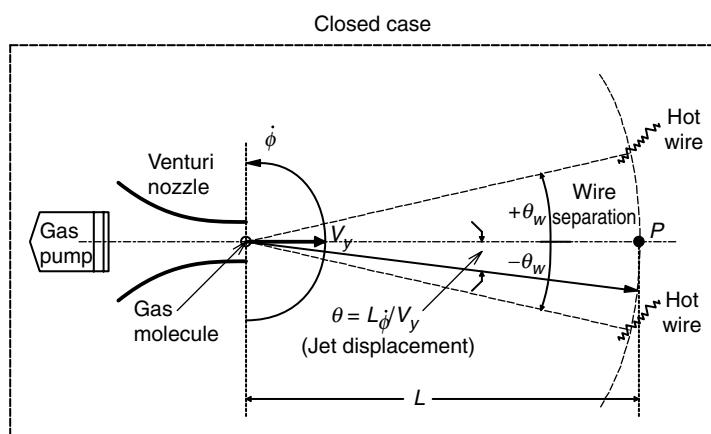
We now look at a gyro with Two, 180° Out of Phase Vibrating Masses. Figure 7.10 illustrates a side and top view of a haltere type VMRG with two, symmetrically vibrating masses. It can be easily shown that the second harmonic torque around the two combined pivots adds in phase for the two masses and twice the second harmonic M_ϕ is generated. A second advantage of this geometry is that the out-of-phase vibrations cancel the reaction forces in the x direction. By analogy, a tuning fork, whose arms are vibrating 180° out of phase, will experience a second harmonic torque around its handle when rotated around the handle axis at angular velocity, $\dot{\phi}$. Such a tuning fork design is the basis for the SystronDonner Gyrochip™ angular rate sensor.

7.2.2.4 The Humphrey Air Jet Gyro

This interesting angular rate sensor does not use a vibrating or rotating mass. However, its operation can be heuristically described in terms of simple Newtonian mechanics and the Coriolis effect. Its operation makes use of the fact that once a mass is in linear motion, it will continue to move in a straight line at constant velocity unless acted on by an external force. The geometry of the air jet gyro is shown schematically in Figure 7.11. For simplicity in the analysis, we will consider a single gas molecule mass and neglect its interactions with other gas molecules in the stream of gas emitted from the venturi, and with the sides of the chamber. The gas is forced through the venturi by a novel piezoelectric gas pump.

**FIGURE 7.10**

Top and side views of a two-mass, vibrating mass rate gyro. The masses vibrate at the same frequency, but 180° out of phase.

**FIGURE 7.11**

Top schematic view of a Humphrey air jet (Coriolis) gyro.

Assume a molecule leaves the nozzle at $t=0$ with velocity V_y in the y direction. If the input rotation rate, $\dot{\phi}$, is zero, the molecule will arrive at a point P midway between the hot wire sensors after $T=L/V_y$ seconds, assuming the velocity remains constant. L is the distance from the nozzle mouth to point P . Now if the input rotation is in the direction shown, the gyro's hot wires, which are used to sense the air jet's mean velocity,

will have rotated CCW by an angle, $\theta = \dot{\phi}T = \dot{\phi}L/V_y$ radians. The distance along the circumference of the circle of radius L is $\delta = L\theta = L^2\dot{\phi}/V_{yo}$ meters. Thus the gas molecule will strike more toward the lower hot wire instead of at P . The displacement of the molecule perpendicular to V_{yo} is called the Coriolis effect.

The distribution of gas molecule velocities in the jet of gas in the gyro is assumed to be modeled by a Gaussian distribution in angle at radius L . That is, at distance L , assuming zero angular velocity:

$$v_y(L, \theta, \dot{\phi} = 0) = V_y \exp[-\theta^2/(2\sigma^2)] \quad (7.54)$$

Thus, at $P(\theta = 0), v_y = V_y$. Assume that the hot wire velocity probes are separated by an angle of $\pm\theta_w$ around point P . Thus the lower wire at $\theta = -\theta_w$ will see a gas velocity of:

$$v_y(-\theta_w, \dot{\phi} = 0) = V_{yo} \exp[-(-\theta_w)^2/(2\sigma^2)] \quad (7.55)$$

The upper wire also sees a velocity of:

$$v_y(\theta_w, \dot{\phi} = 0) = V_{yo} \exp[-(+\theta_w)^2/(2\sigma^2)] \quad (7.56)$$

Now if $\dot{\phi} \geq 0$, the Coriolis effect will deflect the center of the jet by an angle $\theta_\phi = L\dot{\phi}/V_{yo}$. Now the gas velocity at the lower hot wire is:

$$v_y(-\theta_w, \dot{\phi}) = V_{yo} \exp[-(-\theta_w + L\dot{\phi}/V_y)^2/(2\sigma^2)] \text{ (lower wire)} \quad (7.57)$$

The gas velocity at the upper wire is lower than at the lower wire, therefore it is hotter and has a higher resistance.

$$v_y(+\theta_w, \dot{\phi}) = V_{yo} \exp[-(+\theta_w + L\dot{\phi}/V_y)^2/(2\sigma^2)] \text{ (upper wire)} \quad (7.58)$$

Let us now indulge in some approximations to linearize the system. We will use the approximation, $e^{\pm\gamma} \cong 1 + \gamma$, for $|\gamma \ll 1|$. Assume that $\theta_w^2 \gg 2\theta_w L\dot{\phi}/V_y \gg L^2\dot{\phi}^2/V_y^2$. Also, assume $L^2\dot{\phi}^2/V_y^2 \ll 1$. Thus, the gas velocity at the lower wire is:

$$\begin{aligned} v_y(-\theta_w, \dot{\phi}) &= V_y \exp[-\theta_w^2 - 2\theta_w L\dot{\phi}/V_y + L^2\dot{\phi}^2/V_y^2]/(2\sigma^2) \cong V_y \exp[-\theta_w^2/(2\sigma^2) + \theta_w L\dot{\phi}/(V_y \sigma^2)] \\ &= V_y \exp[-\theta_w^2/(2\sigma^2)] \exp[\theta_w L\dot{\phi}/(V_y \sigma^2)] \cong V_y \exp[-\theta_w^2/(2\sigma^2)][1 + \theta_w L\dot{\phi}/(V_y \sigma^2)] \end{aligned} \quad (7.59)$$

Using the same development, we find that the gas velocity at the lower probe is:

$$v_y(+\theta_w, \dot{\phi}) \cong V_y \exp[-\theta_w^2/(2\sigma^2)][1 - \theta_w L\dot{\phi}/(V_y \sigma^2)] \quad (7.60)$$

We now turn our attention to the circuit that provides the differential output from the hot wire anemometers in the Humphrey angular rate sensor. As you will see in Section 7.3.2.1, hot wire or hot film anemometers (HWAs) are widely used to measure gas velocity in a variety of applications, including subsonic vehicle aerodynamics, as well as heating and ventilating system design. The theory behind the HWA is simple.

When a wire is heated, its resistance increases. Also, when moving gas is directed at the wire, heat is removed, its temperature drops, and its resistance decreases. These phenomena can be described mathematically. The resistance increase can be modeled by:

$$R(\Delta T) = R_o[1 + \alpha\Delta T] \quad \text{abv1(7.61)}$$

Here R_o is the wire's resistance at reference temperature, T_o , ΔT is the temperature difference of the wire above the reference temperature caused by electrical (I^2R) heating, and α is the wire's alpha tempco (α is material and temperature dependent). ΔT can be modeled by:

$$\Delta T = P\Theta \quad \text{abv2(7.62)}$$

where P is the average electrical power dissipated in the wire ($P = I_o^2 R$) in watts and Θ is the thermal resistance of the wire in $^{\circ}\text{C}/\text{watt}$. Θ depends on the wire's geometry (shape, length, diameter, etc., and the heat conductivity and heat capacity of the gas surrounding the hot wire) and also on whether the gas surrounding the wire is moving or not. In moving gas, it has been found that:

$$\Theta \cong \Theta_o - \beta\sqrt{v} \quad \text{abv3(7.63)}$$

Here, Θ_o is the thermal resistance in still gas, β is a non-negative constant and v is the gas velocity at the wire. Using relations 7.61–63, we can solve for $R(v)$:

$$R(v) = \frac{R_o}{1 - \alpha I_o^2 R_o [\Theta_o - \beta\sqrt{v}]} \quad \Omega \quad (7.64)$$

Now, refer to the schematic shown in Figure 7.12. Each of the hot wires is heated by a dc current source, I_o . The voltage across each hot wire is simply $I_o R(v)$. Thus, the voltage across the top wire can be written as:

$$V_+ = \frac{I_o R_o}{(1 - \alpha I_o^2 R_o \Theta) + \alpha I_o^2 R_o \beta \sqrt{v_y(+\theta_w, \dot{\phi})}} \cong \frac{I_o R_o / (1 - \alpha I_o^2 R_o \Theta)}{1 + \frac{\alpha I_o^2 R_o \beta \sqrt{v_y(+\theta_w, \dot{\phi})}}{(1 - \alpha I_o^2 R_o \Theta)}} \quad (7.65)$$

$$V_+ \cong \frac{I_o R_o}{(1 - \alpha I_o^2 R_o \Theta)} \left[1 - \frac{\alpha I_o^2 R_o \beta \sqrt{v_y(+\theta_w, \dot{\phi})}}{(1 - \alpha I_o^2 R_o \Theta)} \right]$$

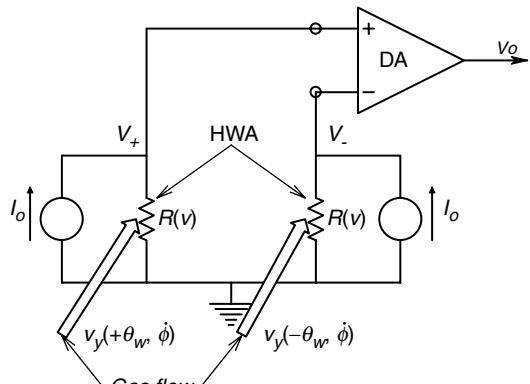


FIGURE 7.12

Simplified schematic of the hot-wire anemometers used in the air jet gyro.

Using a similar calculation, we find:

$$V_- \cong \frac{I_o R_o}{(1 - \alpha I_o^2 R_o \Theta)} \left\{ 1 - \frac{\alpha I_o^2 R_o \beta \sqrt{v_y(-\theta_w, \dot{\phi})}}{(1 - \alpha I_o^2 R_o \Theta)} \right\} \quad (7.66)$$

In the above calculations, we have assumed that:

$$\frac{\alpha I_o^2 R_o \beta \sqrt{v_y(\pm \theta_w, \dot{\phi})}}{(1 - \alpha I_o^2 R_o \Theta)} \ll 1 \quad (7.67)$$

and have used the approximation: $1/(1 + \varepsilon) \cong 1 - \varepsilon$. Next, we use the binomial theorem to write $\sqrt{1 \pm \varepsilon} \cong 1 \pm \varepsilon/2$. This latter relation is applied to equations 7.65 and 7.66.

$$\sqrt{v_y + (\pm \theta_w, \dot{\phi})} \cong \sqrt{v_y \exp[-\theta_w^2/(2\sigma^2)][1 \pm \theta_w L \dot{\phi}/(V_y 2\sigma^2)]} \quad (7.68)$$

When equation labelled Abv is substituted into the approximations for V_+ and V_- , V_o can be found from:

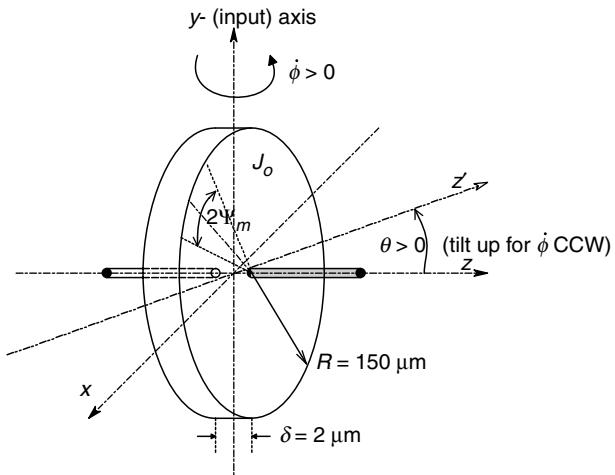
$$V_o = K_A(V_+ - V_-) = \dot{\phi} \frac{K_A I_o^3 R_o^2 \alpha \beta (\theta_w L / V_y \sigma^2) \sqrt{V_y \exp[-\theta_w^2/2\sigma^2]}}{(1 - \alpha I_o^2 R_o \Theta_o)^2} \quad (7.69)$$

Thus, it appears that within the bounds of the inequality assumptions made above, the steady state differential output of the pair of hot wire anemometers is proportional to the input rotation rate, $\dot{\phi}$. It should be noted that there are more sophisticated means of processing the outputs of HWAs. One such means, used for large signal changes, is to remove the \sqrt{v} proportionality by squaring. Another means uses feedback to automatically increase the I_o through the HWA to compensate for an increase in \sqrt{v} , thus forcing the HWA resistance to remain constant at $R(v=0)$.

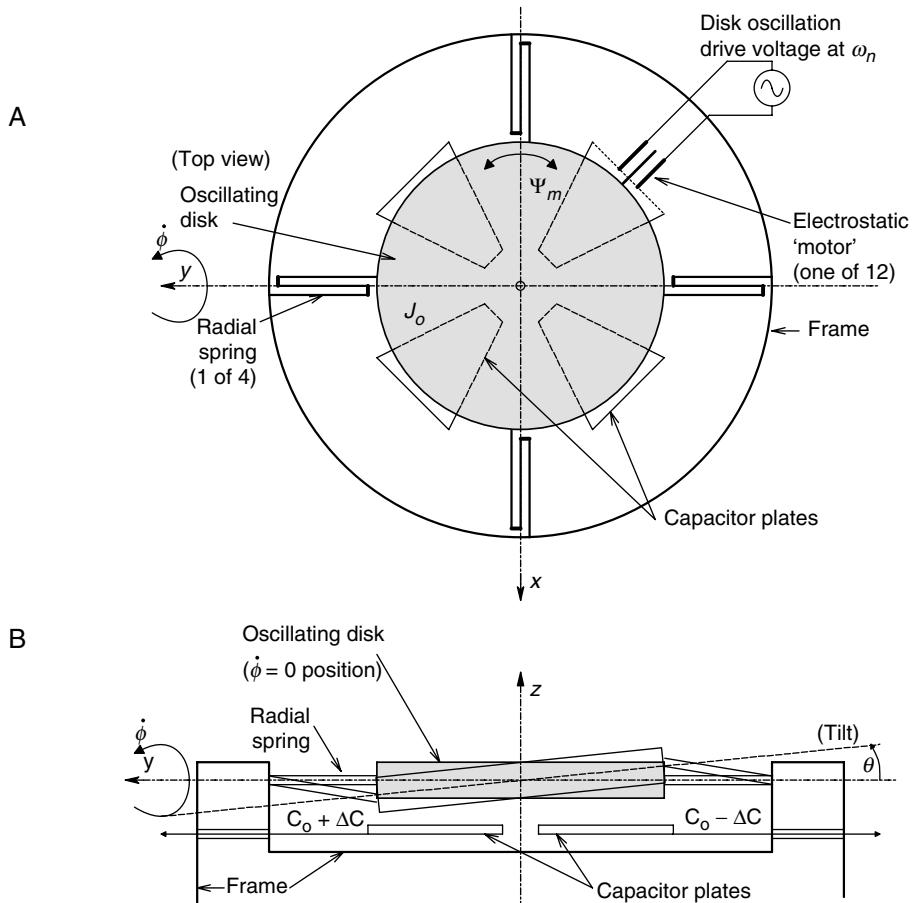
The Humphrey angular rate sensor has a built-in transport lag in its transfer function. This is simply the average time it takes a gas molecule to travel from the venturi nozzle to point P . In one Humphrey rate sensor, $L=0.4$ in and $V_y=80$ in/s. Thus, $T=5$ ms, the transport lag. If a step input of angular velocity is given at $t=0$, it takes 5 ms before the sensor responds. What will the steady state output voltage be if the same angular rate sensor is given a 200 Hz sinusoidal input? That is, $\phi(t) = \phi_0 \sin(2\pi 200t)$. It appears that the Humphrey rate sensor has a response bandwidth of ca. 0–25 Hz, which is adequate for application in most vehicles (subsonic aircraft, boats, submarines, etc.).

7.2.2.5 Oscillating Disk Rate Gyro

A micromachined, oscillating disk, rate gyro was developed at UC Berkeley by T. Juneau for his PhD research [Juneau and Pisano, 1996; Juneau, Pisano and Smith, 1997]. Figure 7.13 illustrates the basic geometry of this angular rate sensor. In this first example, consider that a thin disk is suspended on two torsion springs along the z axis. The disk is caused to oscillate at its resonant frequency, ω_m , around its spin (z) axis with a low amplitude, sinusoidal motion. One of 12 electrostatic ‘motors’ arranged around the circumference of the disk is shown at ‘2 o’clock’ in Figure 7.14A. Electrostatic attraction/repulsion between each electrode fixed to the disk and the stationary drive electrodes is used to

**FIGURE 7.13**

Simplified side view of a micromachined, oscillating disk rate gyro.

**FIGURE 7.14**

A. Top view of an oscillating disk rate gyro mass wheel. One of 12 electrostatic 'motors' is shown in the '2-o'clock' position on the disk edge. Four fixed capacitor plates are positioned under the disk to measure disk tilt due to gyro action. B. Side view of the oscillating disk gyro, showing disk tilt and the capacitor plates.

generate a periodic torque which maintains a constant amplitude disk oscillation around the z axis. The mechanical resonant frequency of the disk is:

$$\omega_m = \sqrt{K_\Psi/J_o} \text{ r/s} \quad (7.70)$$

where K_Ψ is the torsional spring constant of the disk and J_o is its moment of inertia. The induced oscillation can be modeled by:

$$\Psi(t) = \Psi_m \sin(\omega_m t) \quad (7.71)$$

Hence,

$$\dot{\Psi}(t) = \omega_m \Psi_m \cos(\omega_m t) \quad (7.72)$$

and

$$\ddot{\Psi}(t) = -\omega_m^2 \Psi_m \sin(\omega_m t) \quad (7.73)$$

The disk oscillation is of the order of 400 Hz.

It is observed that if the disk is rotated CCW at a constant velocity about its input (y) axis, a torque will be developed that tries to tilt the disk spin (z) axis up in the zy plane. This tilt, θ , due to the torque can be measured to sense $\dot{\phi}$.

The following mathematical development using the Lagrange equations illustrates how the oscillating disk rate gyro works: The total kinetic energy, T , of the oscillating disk is:

$$T = \frac{1}{2} J_o [\dot{\Psi} + \dot{\phi} \sin(\theta)]^2 + \frac{1}{2} J_1 [\dot{\theta}^2 + \dot{\phi}^2 \cos^2(\theta)] J \quad (7.74)$$

Note that J_o is the moment of inertia of the disk around the z (spin) axis and J_1 is its moment of inertia around the x and the y axes. The potential energy of the disk, U , is stored in the torsion springs on the z axis and the spring that restrains the tilt θ in the yz plane.

$$U = \frac{1}{2} K_\Psi \Psi^2 + \frac{1}{2} K_\theta \theta^2 \quad (7.75)$$

Now the Lagrangian function for the system is $L \equiv T - U$:

$$L = \frac{1}{2} J_o [\dot{\Psi}^2 + 2\dot{\Psi}\dot{\phi} \sin(\theta) + \dot{\phi}^2 \sin^2(\theta)] + \frac{1}{2} J_1 [\dot{\theta}^2 + \dot{\phi}^2 \cos^2(\theta)] - \frac{1}{2} [K_\Psi \Psi^2 + K_\theta \theta^2] \quad (7.76)$$

There are three Lagrangian coordinates (q_k) for the disk gyro:

$q_1 = \Psi$: Rotor \angle on the spin axis

$q_2 = \theta$: Rotor tilt $\angle \theta$ in yz plane

$q_3 = \phi$: Input \angle around y axis

1. For $q_1 = \Psi$

$$\frac{\partial L}{\partial \Psi} = -K_\Psi \Psi \quad (7.77a)$$

$$\frac{\partial L}{\partial \dot{\Psi}} = J_o [\dot{\Psi} + \dot{\phi} \sin(\theta)] \quad (7.77b)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\Psi}} \right) = J_o [\ddot{\Psi} + \ddot{\phi} \sin(\theta) + \dot{\phi} \dot{\theta} \cos(\theta)] \quad (7.77c)$$

Now the torque around the spin (z) axis is given by:

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\Psi}} \right) - \frac{\partial L}{\partial \Psi} = M_\Psi \quad (7.78)$$

We assume that the disk tilt is restrained by a very stiff spring, so that $\theta \rightarrow 0$ and $\dot{\theta} \rightarrow 0$. Thus, the torque required to oscillate the disk at constant amplitude is:

$$M_\Psi \cong J_o \ddot{\Psi} - K_\Psi \Psi \quad (7.79)$$

2. For $q_2 = \theta$

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \frac{1}{2} J_o \left[2 \dot{\Psi} \dot{\phi} \cos(\theta) + -\dot{\phi}^2 (-2 \sin(2\theta)) \frac{1}{2} + \frac{1}{2} J_1 \dot{\phi}^2 \frac{1}{2} (-2 \sin(2\theta)) \right] - K_\theta \theta \\ &= J_o \dot{\Psi} \dot{\phi} \cos(\theta) + \frac{1}{2} J_o \dot{\phi} \sin(2\theta) - \frac{1}{2} J_1 \dot{\phi}^2 \sin(2\theta) - K_\theta \theta \end{aligned} \quad (7.80a)$$

$$\frac{\partial L}{\partial \dot{\theta}} = 0 \quad (7.80b)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = 0 \quad (7.80c)$$

Thus,

$$M_\theta = \frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} = -J_o \dot{\Psi} \dot{\phi} + K_\theta \theta = -\dot{\phi} J_o [\omega_m \Psi_m \cos(\omega_m t)] + K_\theta \theta \quad (7.81)$$

Equation 7.81 is an expression for the torque on the tilt axis (i.e. the output torque). M_θ lies in the yz plane. Finally, we examine the torque around the third (input or y) axis:

3. $q_3 = \phi$

$$\frac{\partial L}{\partial \dot{\phi}} = 0 \quad (7.82a)$$

$$\frac{\partial L}{\partial \phi} = J_o \dot{\Psi} \sin(\theta) + J_o \dot{\phi} \frac{1}{2} [1 - \cos(2\theta)] + J_1 \dot{\phi} \frac{1}{2} [1 + \cos(2\theta)] \quad (7.82b)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) &= J_o \ddot{\Psi} \sin(\theta) + \frac{1}{2} J_o \ddot{\phi} - \frac{1}{2} J_o \ddot{\phi} \cos(2\theta) + \frac{1}{2} J_o \dot{\phi} \dot{\theta} 2 \sin(2\theta) + \frac{1}{2} J_1 \ddot{\phi} + \frac{1}{2} J_1 \ddot{\phi} \cos(2\theta) \\ &\quad - \frac{1}{2} J_1 \dot{\phi} \dot{\theta} 2 \sin(2\theta) \end{aligned} \quad (7.82c)$$

Now, assuming $\theta \rightarrow 0$ and $\dot{\theta} \rightarrow 0$, the expression for the input torque reduces to:

$$M_\phi = \frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) - \frac{\partial L}{\partial \phi} \cong J_1 \ddot{\phi} \quad (7.83)$$

Note that the rotation rate of Juneau's oscillating disk gyro can be sensed around the x axis as well as the y axis. When the x axis is the input axis, because of symmetry, the disk experiences a torque in the xz plane. Figure 7.14B illustrates schematically how the Juneau gyro is suspended, and how the small disk tilt is sensed. From equation 7.81, we find:

$$\theta = \dot{\phi} \frac{J_0 \omega_m \Psi_m \cos(\omega_m t)}{K_\theta} \quad (7.84)$$

K_θ is the stiffness (torsional spring constant) of the four radial suspension springs in the yz plane in Nm/radian. The actual small angular tilt, θ , is sensed by measuring the change in capacitance using two pairs of capacitor plates located under the disk—the plates of a pair are opposite each other. Each pair of plates can be connected to a DeSauty bridge type circuit (the DeSauty bridge was described in Section 5.5.1.4). Note from equation 7.84 that $\theta \propto \dot{\phi}$ and $\Delta C \propto \theta$ and therefore, after phase sensitive demodulation of the bridge output, $V_o \propto \dot{\phi}$. Two orthogonal pairs of capacitor plates are required to resolve both disk tilt θ due to rotation $\dot{\phi}$ around the y axis, but also orthogonal disk tilt due to rotation $\dot{\alpha}$ around the x axis (not shown). The DeSauty bridge drive voltage must have a frequency, $\omega_b \gg \omega_m$. The ac [$\cos(\omega_m t)$] term in the bridge output is again phase sensitive rectified to yield a dc $V'_o \propto \dot{\phi}$. As we commented above, the Juneau oscillating disk gyro can be used to sense two angle rates simultaneously (e.g. roll and pitch) [Juneau and Pisano, 1996]. Thus the outputs of the two capacitance bridges must be functionally combined to resolve the exact disk tilt vector which is the result of simultaneous $\dot{\phi}$ and $\dot{\alpha}$ inputs. It is left as a Chapter exercise to derive an expression for the tilt due to $\dot{\alpha}$, and to design circuitry to resolve the net tilt. Juneau reported that his micromachined disk gyro prototype could resolve $\dot{\phi} \geq 0.1667^\circ/\text{s}$ in a 0–25 Hz bandwidth.

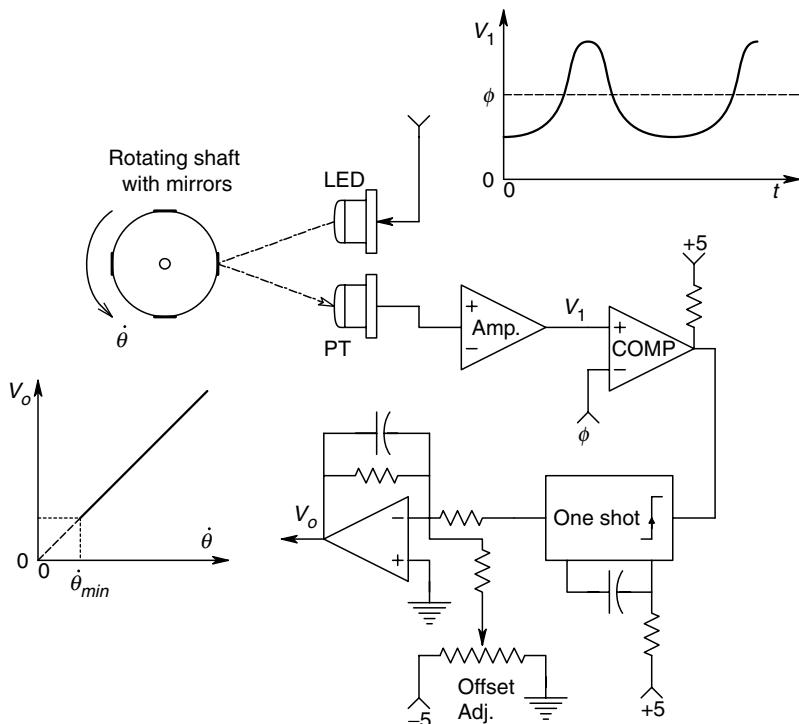
There has been much activity to develop sensitive, micromachined rate sensors since 1991, when Draper described his micromachined, double gimbal vibratory gyroscope. The interested reader should consult the review paper by Yazdi *et al* (1998) and the paper describing recent results in this R&D area by Acar and Shkel (2004). Clearly, inexpensive, 'chip' gyroscopes have application in vehicle stabilization (cars, trucks, planes, boats) and missile guidance and control. However, they generally lack the sensitivity required for inertial navigation.

7.2.3 Angular Velocity Measurement with Tachometers

A tachometer is a sensor/system which converts the angular velocity of a rotating shaft to an electrical output, generally a voltage. Tachometers are typically used to measure the angular velocity, generally in RPM, of rotating machines such as electric motors, generators, turbines, pumps, internal combustion engines, etc.

A commonly found form of tachometer is the dc, permanent magnet generator, in which the output voltage linearly follows the shaft velocity. Shaft velocity is read by a calibrated dc voltmeter. The response time of the dc generator tachometer is basically set by the mechanical natural frequency of the analog voltmeter. In the case of a digital voltmeter, the response time is limited by the low-pass filtering at the input analog-to-digital converter.

Other tachometers avoid the expense of the dc generator and use either optical or magnetic means to generate a train of pulses, whose repetition rate is proportional to the shaft speed. In the tachometer illustrated in Figure 7.15, a light beam is bounced off one

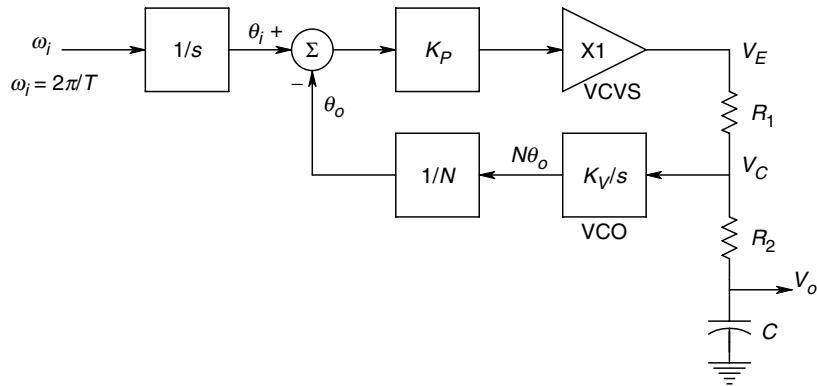
**FIGURE 7.15**

Schematic diagram of a simple, photoelectronic tachometer system. Amp. conditions the pulses from the phototransistor, PT. The conditioned pulses, V_1 , are made into standard width TTL pulses by the amplitude comparator COMP and the one-shot multivibrator. The op-amp low-pass filters the pulse train from the one-shot and adjusts the dc level. The dc output, V_o , is proportional to $\dot{\theta}$. However, at very low $\dot{\theta}$, the low frequency of the pulse train causes excessive ripple on V_o , setting a practical lower limit to $\dot{\theta}$ resolution, determined by the LPF time constant, RC .

or more reflecting strips cemented to the rotating shaft. Every time the shaft comes around, the light beam is reflected onto a photosensor, producing a train of pulses which is shaped by a comparator and one shot multivibrator. The average value of this conditioned pulse train is seen to be proportional to the shaft's RPMs. A low-pass filter averages the pulse train and an op-amp is used to correct for dc offset and set the output calibration, as shown in Figure 7.15.

An alternate means of generating a pulse train, whose rate is proportional to shaft speed, is to cement permanent magnets to the rotating shaft and place a digital Hall sensor in a fixed location near the path of the magnet. Every time the magnet comes close to the Hall sensor, its output changes state, generating the pulse train. The disadvantage of the pulse averaging tachometer is that the low-pass filter time constant needs to be long in order to suppress output voltage ripple at low RPM. The long time constant means a slow response to changes in RPM.

One way to get around the problem of a long filter time constant being required for ripple free output at low RPMs is to use a phase lock loop (PLL) to convert the rate of the pulse train to a dc output proportional to RPM. The block diagram of a PLL tachometer is shown in Figure 7.16. A 4046 CMOS PLL is used to convert input frequency to a proportional output voltage, V_o . In the block diagram, the first integrator is required to convert the input signal's frequency to phase, the quantity the PLL operates on. In the

**FIGURE 7.16**

Block diagram of a 4046 CMOS phase lock loop used as a tachometer. The (virtual) input integrator converts the input event rate (frequency, ω_i) to the phase input, θ_i , of the phase detector.

steady state, $\theta_i = \theta_o$ and it is easy to show that the dc output voltage is given by:

$$V_o = 2\pi f_i N / K_V \quad (7.85)$$

where N is a digital frequency divider ratio ($N = 1, 2, 4, \dots, 64, \dots$ etc.) and K_V is the PLL VCO's constant in (r/s)/V. Reduction of the block diagram yields the transfer function:

$$\frac{V_o}{f_i}(s) = \frac{2\pi N / K_V}{s^2 C(R_1 + R_2)N / (K_P K_V) + s(1 + CR_2 K_P K_V / N)N / K_P K_V + 1} \quad (7.86)$$

where K_P is the phase detector gain. The PLL tachometer is seen to be a second order system with a natural frequency of $\omega_n = \sqrt{K_P K_V / (C(R_1 + R_2)N)}$ r/s. In this design, the VCO runs at N times the input frequency at lock. This allows the measurement of very low frequencies. All VCOs have a minimum frequency at which they will oscillate, as well as a maximum frequency. The minimum and maximum VCO frequencies, divided by N , set the input range of this PLL tachometer.

The final means for the measurement of angular velocity which we will describe can be used with turbine wheels as well as rotating shafts. A system, known as an instantaneous pulse frequency demodulator (IPFD), operates on a pulse sequence obtained by the rotation of magnets attached to the turbine wheel or shaft passing the fixed Hall switch, or alternately, by the rotation of illuminated, alternating white and black markings past a fixed photosensor. The IPFD measures the period or time interval between each pulse pair in the sequence and then calculates the reciprocal of that interval. The reciprocal of the i th interval is defined as the i th element of instantaneous frequency of the pulse sequence, r_i , and is proportional to the angular velocity of the shaft or wheel. If, for example, each revolution produces 36 pulses, then 2400 RPM would produce 86,400 ppm, or 1440 pps. Since the IPFD is responsive to the reciprocal of the length of each period, any change in the speed of rotation from 2400 RPM will be sensed within about 0.7 ms. This is a considerably faster response time than is afforded by pulse averaging or PLL tachometers. The IPFD is thus a useful system to study subtle, short term changes in angular velocity. A block diagram for an analog IPFD tachometer is shown in

Figure 7.17A. The processing time (i.e. the time required to charge the capacitor to the maximum voltage) sets the maximum possible instantaneous frequency (shortest input period). The maximum input period (lowest input frequency) is determined by the lowest capacitor voltage at which the decay waveform follows the curve given by:

$$v_c(t) = \frac{C/\beta}{t + \tau_o} \quad (7.87)$$

V_o is the peak voltage that the capacitor is charged to at time $t=0$, when the voltage begins decaying hyperbolically. Time τ_o is the processing time between the $(i-1)$ th pulse and the beginning of the hyperbolic decay, $v_c(t)$, which is sampled at time t_i when the i th pulse occurs.

$$\beta \equiv C/(\tau_o V_o)$$

The hyperbolic decay of voltage from the capacitor occurs because the capacitor is discharged through a nonlinear resistor with the voltampere relation:

$$i_{nl} = \beta v_c^2 \quad (7.88)$$

Hence, the node equation for the capacitor discharge is:

$$\dot{v}_c + (\beta/C)v_c^2 = 0, \quad t \geq 0 \quad (7.89)$$

Eq. 7.89 is a form of Bernoulli's equation, which has a solution for the initial condition of $v_c = V_o$ at $t = 0$ given pulse, generating a stepwise voltage waveform, $q(t)$. The height of each step is by equation 7.87 above. The value of the hyperbola is sampled and held at the occurrence time of the $(i+1)$ th instantaneous frequency of the preceding pulse pair interval, and is proportional to RPM. In mathematical terms, the IPFD output can be written:

$$q(t) = k \sum_{i=1}^{\infty} r_i [U(t - t_i) - U(t - t_{i+1})] \quad (7.90)$$

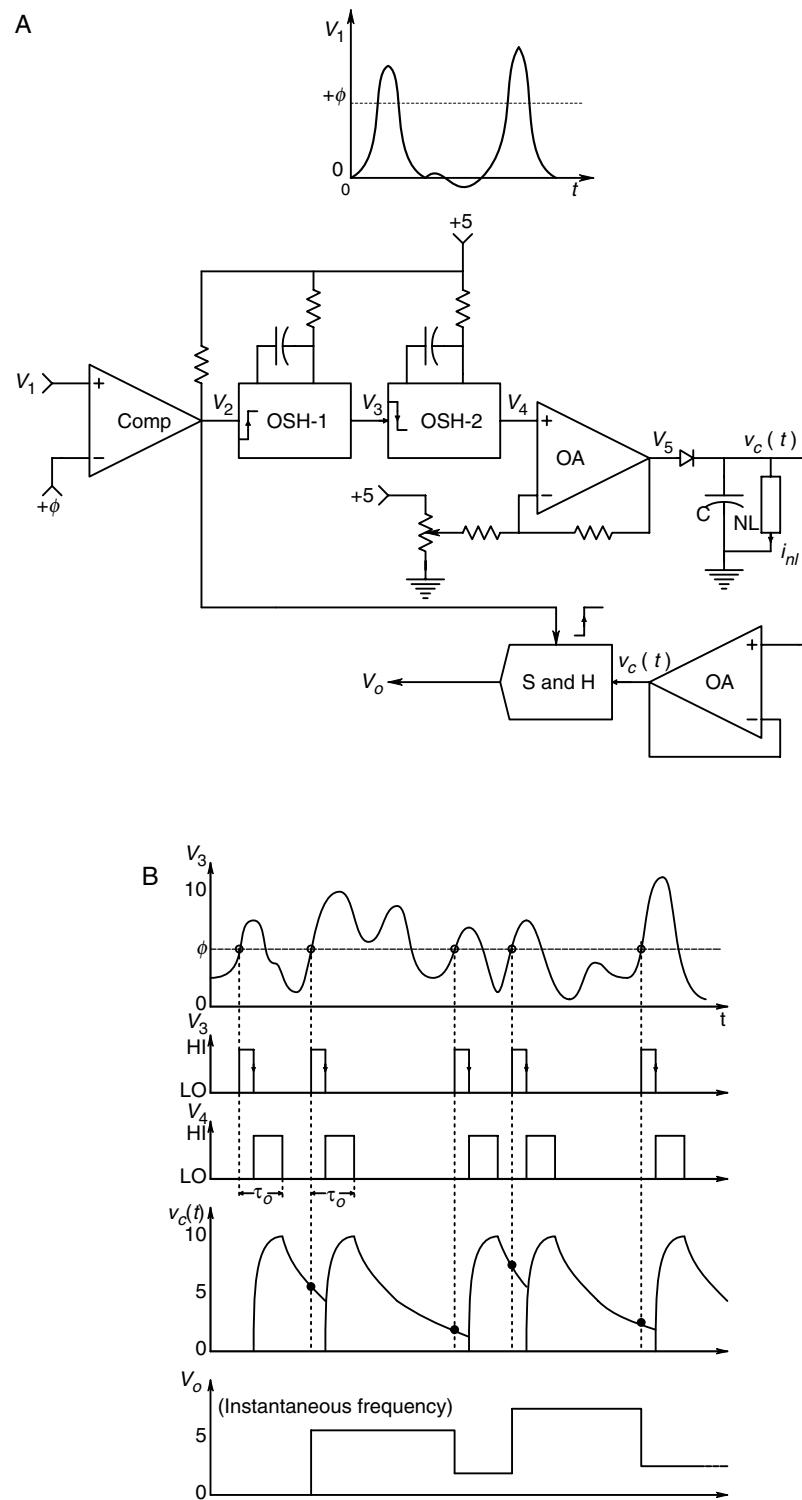
Here $U(t - a)$ is the unit step function, zero for $t < a$ and 1 for $t \geq a$. r_i is the i th element of instantaneous frequency, defined over $t_i \leq t < t_{i+1}$ as:

$$r_i \equiv 1/(t_i - t_{i-1}) \text{ s}^{-1} \quad (7.91)$$

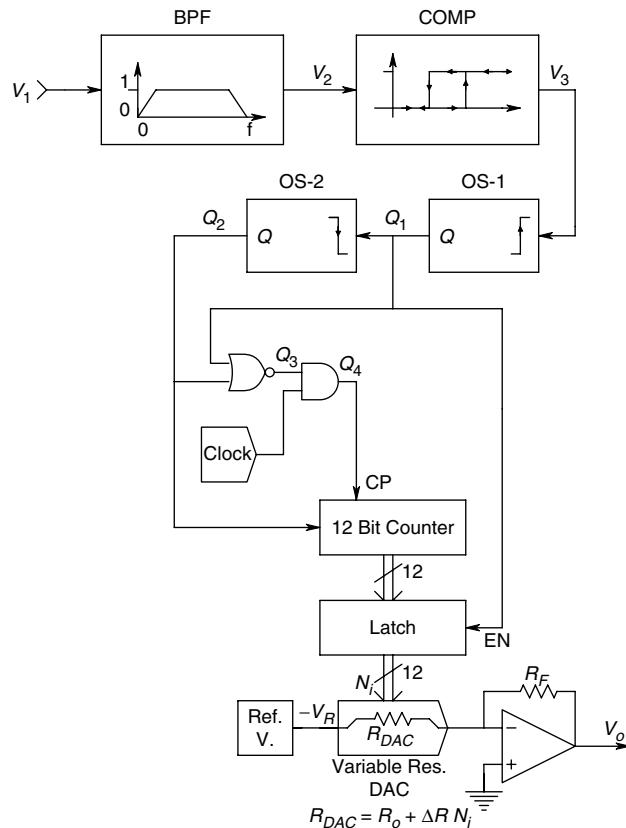
Figure 7.17B illustrates key waveforms in the operation of the analog IPFD.

Webster (1978) describes a digital IPFD which is used as a cardiotachometer (to measure heart rate). The Webster system has an analog output which is also given by equation 7.90 above. Illustrated in Figure 7.18, the reciprocal relation between interpulse interval and r_i is ingeniously generated by the use of a 12-bit DAC whose output resistance increases linearly with the size of its binary input, N_i . N_i is proportional to $(t_i - t_{i-1})$. A timing diagram for the Webster (1978) IPFD is shown in Figure 7.19. The op-amp's output is thus $V_o = V_R R_F / R_{DAC} = Kr_i$ over the interval (t_i, t_{i+1}) .

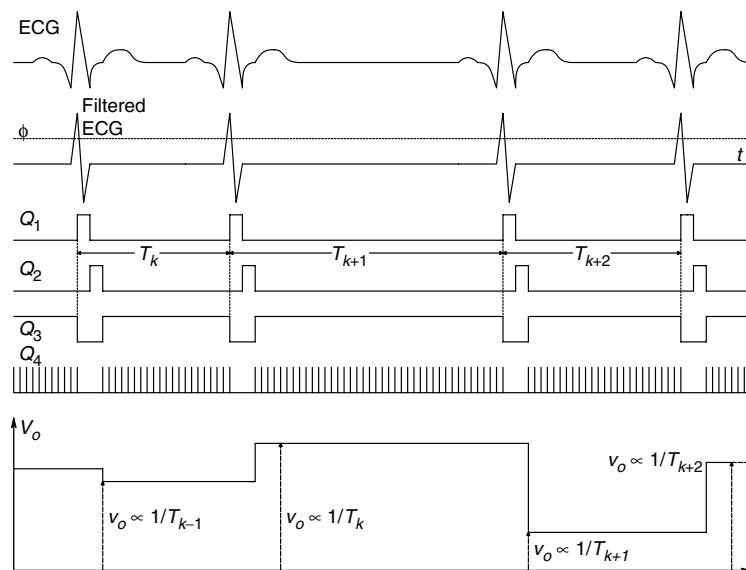
In summary, we have seen that tachometers can have a wide variety of embodiments, ranging from simple dc generators, whose output voltage is proportional to shaft angular velocity, to systems which generate pulse trains whose frequencies (and instantaneous

**FIGURE 7.17**

A. Schematic of an analog Instantaneous Pulse Frequency Demodulator (IPFD) designed by the author. B. Waveforms observed in the analog IPFD. Note that the output voltage appears stepwise, the height of a step being proportional to the reciprocal of the time between the two preceding pulses.

**FIGURE 7.18**

A digital IPFD (with analog output) described by Webster (1978).

**FIGURE 7.19**

Timing diagram for the Webster IPFD of Figure 7.18. Here, the IPFD is used to measure the beat-by-beat heart rate.

frequencies) are proportional to the angular velocity. The pulse trains can be processed by averaging, phase lock loops, or instantaneous pulse frequency demodulators, to obtain dc signals proportional to the angular velocity.

7.2.4 Angular Position Measurement with Gyroscopes

Modern inertial navigation systems and attitude control systems, such as those used by space vehicles, helicopters, aircraft, submarines, etc., need estimates of angular position for their operation. There are several simple modifications of mechanical and Sagnac FO gyroscopes which will allow them to sense angular position, rather than just angular velocity. These include the rate integrating gyro (RIG), and various servo platform gyroscope configurations.

The Rate Integrating Gyro (RIG) is the same as the rate gyro shown in Figure 7.5, except that the spring is removed. The linearized gyro equations now become

$$M_\theta = -B\dot{\theta} = J_1\ddot{\theta} - H_r\dot{\phi} \quad (7.92a)$$

$$M_\phi = H_r\dot{\theta} + J_1\ddot{\phi} \quad (7.92b)$$

Solution of equations 7.92a and b yields:

$$\frac{\Theta}{\Phi}(s) = \frac{H_r/J_1}{s + B/J_1} \quad (7.93)$$

Thus, in the steady state, a rotation ϕ_o around the input axis will produce an inner gimbals' deflection of $\phi_o H_r/J_1$ radians. The RIG is seen to be approximately a first order, low-pass sensor system with a time constant of J_1/B s.

Figure 7.20 illustrates the use of a feedback platform to obtain an output signal proportional to the steady state input rotation, ϕ_{io} . In this case, we use a servomotor driven platform to counter-rotate the RIG as the assembly rotates with the vehicle. The inner gimbals' angle, θ , is sensed by a potentiometer or optical sensor. The servo gyro equations are:

$$\phi_e = \phi_i - \phi_m \quad (7.94a)$$

$$M_\theta = 0 = B\dot{\theta} + J_1\ddot{\theta} - H_r\dot{\phi}_e \quad (7.94b)$$

$$M\phi = H_r\dot{\theta} + J_1\ddot{\phi}_e \quad (7.94c)$$

The platform position is driven by the θ signal through a proportional plus derivative compensated servomotor. The platform position may be related to θ by the relation:

$$\frac{\Phi_m}{\Theta}(s) = \frac{K_M K_A (\tau_c s + 1)}{s(\tau_M s + 1)} \quad (7.95)$$

The block diagram describing the servo RIG is shown in Figure 7.21. System output is proportional to the platform angle, ϕ_m . From the block diagram, we can write the

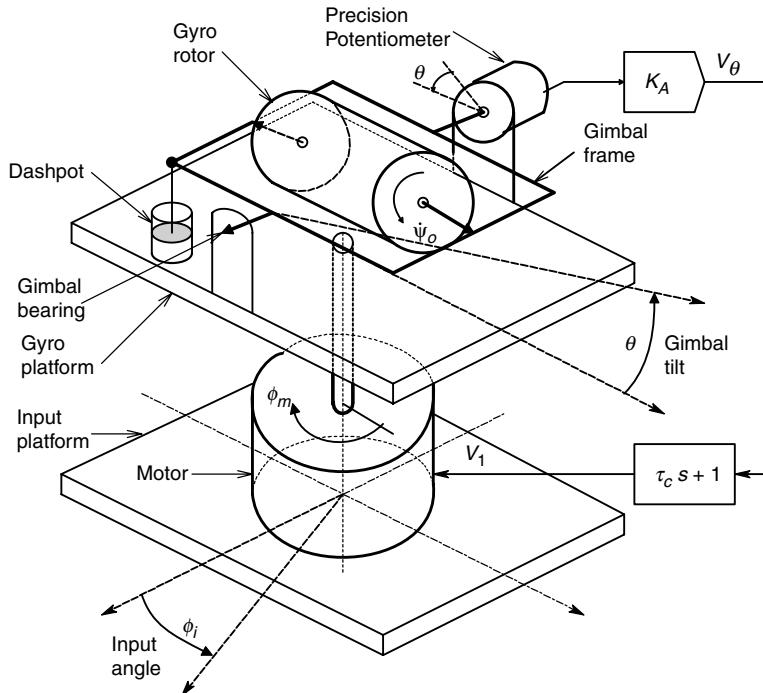
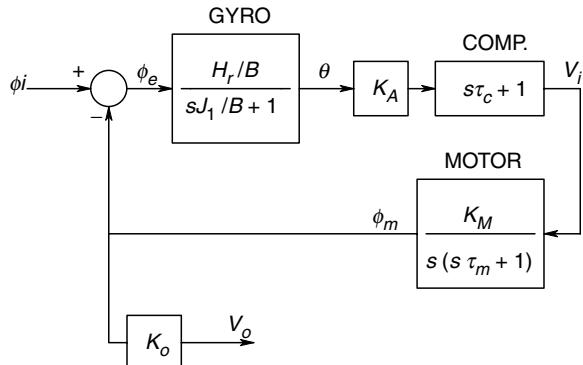
**FIGURE 7.20**

Diagram of a servo platform, rate integrating gyro. In the steady state, $\phi_m = \phi_i$. Proportional plus derivative feedback is needed for stability.

**FIGURE 7.21**

Block diagram describing the dynamics of the servo RIG of Figure 7.20.

servo RIG's transfer function:

$$\frac{V_o}{\Phi_i}(s) = \frac{K_o(\tau_C s + 1)}{s^3 J_1 \tau_M / (H_r K_A K_M) + s^2 (B \tau_M + J_1) / (H_r K_A K_M) + s (B + \tau_C H_r K_A K_M) / (H_r K_A K_M) + 1} \quad (7.96)$$

In the steady state, $V_o = K_o \Phi_i$. The advantage of the servo RIG configuration is that the sensitivity is set by the parameter, K_o , and the dynamic range is not limited by inner gimbal tilt angle, θ . In fact, it is easy to show that in the steady state, $\theta/\Phi_i \rightarrow 0$. The natural frequency of the servo RIG is seen to depend on the gains K_A and K_M .

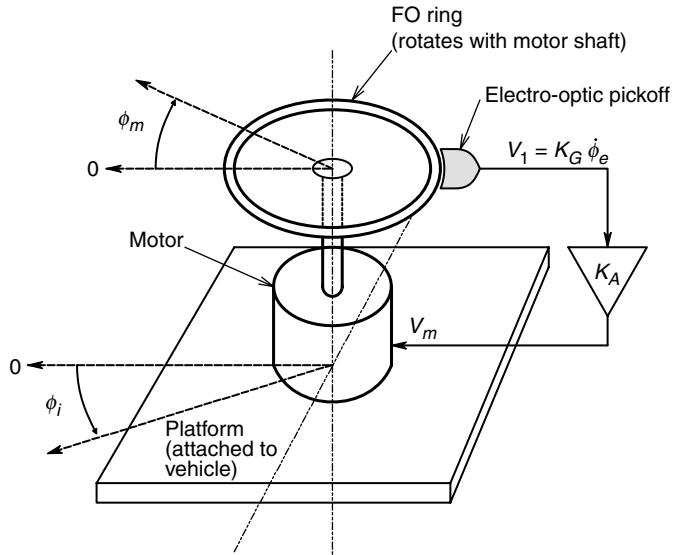
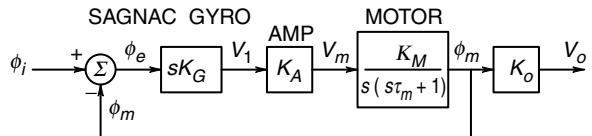
**FIGURE 7.22**

Diagram of a servo platform, fiber optic (Sagnac) gyro.

**FIGURE 7.23**

Block diagram describing the servo FO gyro of Figure 7.22.

A Sagnac FO gyro can also be placed on a servo platform to realize a RIG. Figure 7.22 illustrates such a system. As in the example above, the platform counter-rotates against the input angle so we have

$$\phi_e = \phi_i - \phi_m \quad (7.97)$$

The Sagnac FO gyro output can be summarized by:

$$V_1 = K_G \dot{\phi}_e \quad (7.98)$$

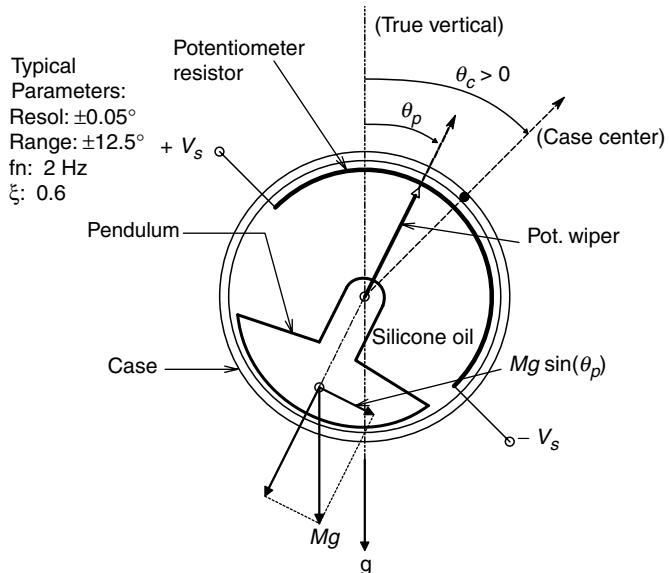
The Sagnac gyro output is further conditioned by an amplifier, K_A , whose output acts on the servomotor system. The net block diagram is shown in Figure 7.23. Again, the system output is proportional to the motor rotation, which can be measured with a potentiometer or a coding disk.

$$\frac{V_o}{\Phi_i}(s) = \frac{K_o}{s\tau_M/(K_G K_A K_M) + 1} \quad (7.99)$$

In the steady state, we find that $V_o = K_o \phi_{io}$, assuming $K_G K_M K_A \gg 1$, and the fringe shift is nulled to zero. Placing a gyroscope on a servo controlled platform is seen to be an effective means of increasing the gyro's dynamic range of response and its natural frequency. In designing such a system, it is clear that there is a trade-off between the considerable extra cost of the servo platform and the increase in performance obtained.

7.2.5 Angular Position Measurement with Clinometers

Gyroscopes, discussed in the previous two sections, were seen to operate either in a gravitational field, or in a microgravity environment, such as in an orbiting vehicle.

**FIGURE 7.24**

Section through a passive, pendulum clinometer. The axis around which the clinometer tilts is perpendicular to the plane of the paper.

Clinometers, on the other hand, require gravity to work. There are two classes of clinometer—pendulum and liquid filled (the liquid is either a dielectric or an electrolyte, depending on the design). Clinometers are used on ships to sense roll and pitch angle, and on heavy earthmoving machinery to sense tilt. Obviously, excessive tilt or roll can lead to capsizing, and loss of life and/or the equipment. A clinometer is considerably simpler in design than a mechanical gyroscope and consequently, is a far less expensive instrument. A schematic design of a pendulum clinometer is shown in Figure 7.24. The roll axis is perpendicular to the plane of the page. A pendulum of total mass, M , and moment of inertia, J , is suspended in the center of a cylindrical case. The case is filled with light oil for viscous damping, and the position of the clinometer pendulum relative to the case is measured by a potentiometer wiper with negligible friction. A torsion spring acts to restore the pendulum to the center position of the case. Summing the torques around the pivot point, we can write:

$$0 = -J\ddot{\theta}_P + B(\dot{\theta}_C - \dot{\theta}_P) + K_s(\theta_C - \theta_P) - MgR \sin(\theta_P) \quad (7.100)$$

This ODE can be linearized by assuming $|\theta_P| < 15^\circ$, so $\sin(\theta_P) \approx \theta_P$ (in radians). Laplace transforming and separating Θ terms, we can write:

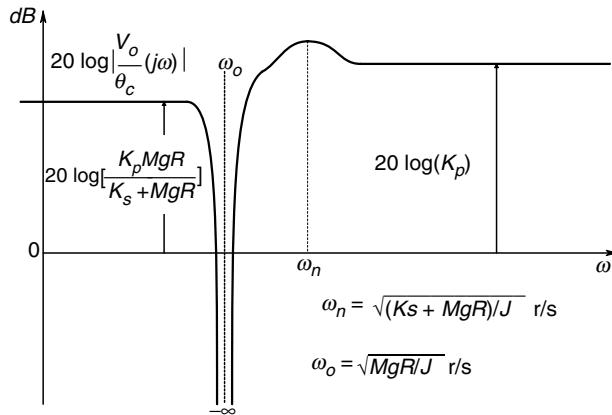
$$\frac{\Theta_P}{\Theta_C}(s) = \frac{(sB + K_s)}{s^2J + sB + (K_s + MgR)} \quad (7.101)$$

Now, the clinometer output is assumed to be given by:

$$V_o = K_o(\theta_C - \theta_P) \quad (7.102)$$

Substituting equation 7.101 in the relation for V_o , we finally obtain:

$$\frac{V_o}{\Theta_C}(s) = \frac{K_o(s^2J/MgR + 1)MgR}{[s^2J/(K_s + MgR) + sB/(K_s + MgR) + 1](K_s + MgR)} \quad (7.103)$$

**FIGURE 7.25**

Steady state frequency response of the pendulum clinometer, subject to a sinusoidal tilting of its case. Note that it is a mechanical notch filter.

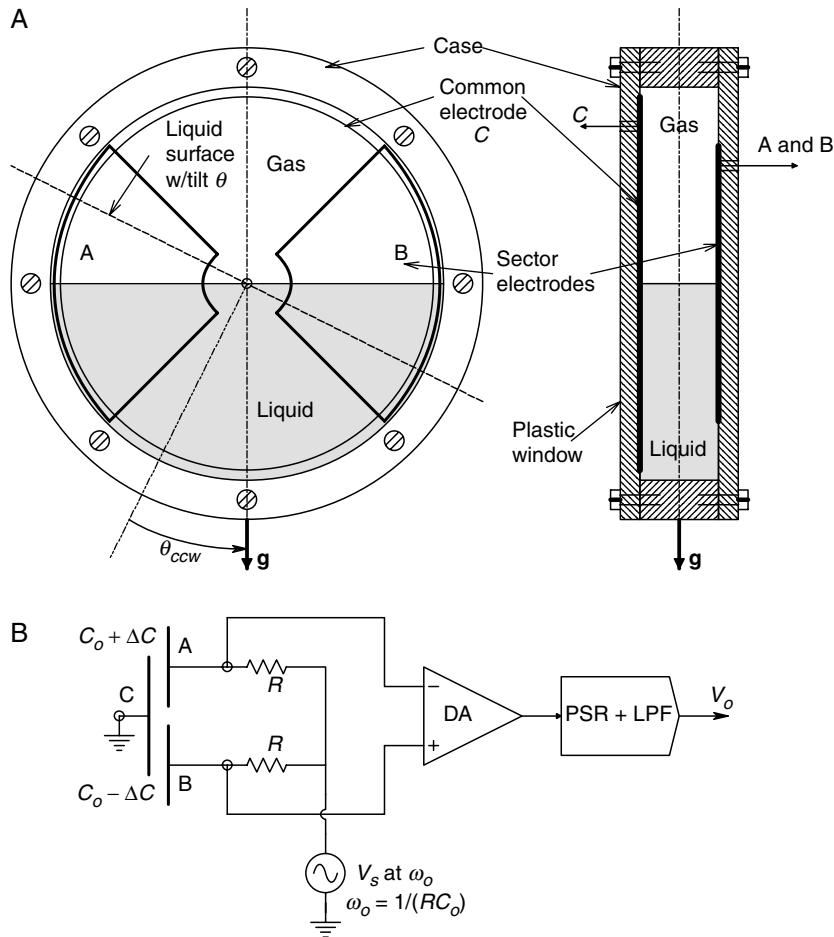
From equation 7.103, we see that the input/output transfer function of the pendulum clinometer is that of a mechanical notch filter. The transfer function has a pair of conjugate zeros on the $j\omega$ axis in the s plane at $\omega_0 = \sqrt{MgR/J}$ r/s and a pair of complex conjugate poles with $\omega_n = \sqrt{(K_s + MgR)/J}$ r/s. A typical steady state frequency response of the pendulum clinometer to $\theta_c(t) = \theta_0 \sin(\omega t)$ is shown in Figure 7.25. The response is flat from dc to just below the notch frequency, ω_0 . In order that the pendulum follow rapid changes in θ_c , $\omega_0 = \sqrt{MgR/J}$ should be made as large as possible. In the limiting case, the pendulum can be considered to be a mass point, M , at a distance R from the pivot. It is well known that such a pendulum has a moment of inertia given by $J = MR^2$. Thus the notch frequency of the pendulum is independent of its mass, and is given by:

$$\omega_0 = \sqrt{g/R} \text{ r/s} \quad (7.104)$$

Hence, a 10 cm pendulum in the earth's gravitational field will have an f_0 of about 1.6 Hz, a 2 cm pendulum has an f_0 of 3.5 Hz. The measurement of most vehicle roll angles is done under almost static conditions, so the dynamic performance of the clinometer angle sensor is generally not an issue. Even the roll rate of large ships is slow enough that most pendulum clinometers can follow the roll angle accurately.

A second class of clinometer has no moving mechanical parts. Figure 7.26 illustrates the simplified sectional drawing of a dielectric fluid filled, differential, capacitive type clinometer. This design principle is used by Schaevitz™ in their AccuStar® line of clinometers. A sealed, thin, cylindrical chamber is half filled with a dielectric liquid with a high dielectric constant. The top half is filled with a dry inert gas. A differential capacitor is formed by depositing a film of noble metal on glass or plastic plates. One electrode is a disk, the other two are sector shaped, so that when the dielectric liquid between the plates tilts, an equal area of the sector plates covered by the liquid is added or subtracted to the area covered at zero tilt. Since the area of a parallel plate capacitor is proportional to the area of the plates, it is easily seen that:

$$\Delta C = \frac{\kappa \epsilon_0 \Delta A}{d} \quad (7.105)$$

**FIGURE 7.26**

A. Front and side view of a dielectric liquid filled clinometer. The sector electrodes and metal back plate make a differential capacitor system. The differential capacitance is a linear function of the tilt angle, θ . B. Schematic of a DeSauly bridge used to sense $\Delta C \propto \theta$. V_o is proportional to θ .

κ is the filling liquid's dielectric constant and ϵ_0 is the permittivity of empty space, equal to $8.85 \times 10^{-12} \text{ C}^2/\text{Nm}^2$. From the electrode geometry, $\Delta A = k\theta$, so it is clear that $\Delta C = K_c\theta$.

The differential change in capacitance caused by tilting the clinometer case can be sensed by using a simple DeSauly bridge as shown in the schematic in Figure 7.26. From the schematic we can write:

$$V_o = K_A V_s [V_B - V_A] = K_A V_s \left[\frac{1/[j\omega(C_o - \Delta C)]}{R + 1/[(j\omega(C_o - \Delta C))]} - \frac{1/[j\omega(C_o + \Delta C)]}{R + 1/[(j\omega(C_o + \Delta C))]} \right] \quad (7.106)$$

Now $\omega \equiv 1/RC_o$ r/s, so:

$$V_o = K_A V_s \left[\frac{2j\Delta C/C_o}{[(j+1) + j\Delta C/C_o][(1+j) - j\Delta C/C_o]} \right] \cong K_A V_s \frac{\Delta C}{C_o} \quad (7.107)$$

Here, we assumed $(\Delta C/C_0)^2 \ll 2$. Thus, the bridge output voltage is proportional to the case tilt, θ . The AccuStar clinometer has a total range of $\pm 60^\circ$ and a linear range of $\pm 45^\circ$. Its resolution is 1 millidegree and its time constant is 0.3 sec, and its -3 dB bandwidth is 0.5 Hz. AccuStar offers analog voltage output, 16-bit serial digital output and a digital pulse width output model.

An interesting variation on the dielectric liquid AccuStar clinometers is the electrolyte liquid clinometer manufactured by HL-Planartechnik GmbH. This sensor apparently uses a conductive liquid instead of a dielectric, and operates as two differential arms of an ac powered Wheatstone bridge. The electrodes must be a noble metal such as gold or platinum to avoid corrosion. Their NS-5/P2 clinometer offers a $\pm 5^\circ$ range, with a resolution of 5×10^{-4} degrees. Output is in RS-232 digital format at 2400 or 9600 baud.

7.2.6 Angular Position Measurement of Shafts

We have already discussed the use of optical coding disks in Section 6.12.1, and the use of precision potentiometers in the measurement of shaft angles in Section 6.3.5. Another, magnetic sensor used in angular position sensing is the Inductosyn®, manufactured by Farrand Controls, Valhalla, NY. Inductosyns have been used in the inertial navigation systems, the periscopes and the fire control systems of Polaris, Poseidon and Trident submarines. They have also been used as position sensing transducers in the Remote Manipulator System Arm installed in the cargo hold of the US space shuttle. Physically, the rotational inductosyn consists of two opposing, flat, toroidal cores, each with a flat printed circuit coil with repeated, parallel hairpin turns attached to one flat surface of the flat toroidal cores. The length of one complete cycle of the hairpin coil is called the pitch, P . Inductive coupling between the two flat coils is used to measure the angular displacement between the two coils. An ac current, generally 2.5–100 kHz in frequency, is used as the excitation to one coil. The voltage induced in the pickup coil is given by:

$$v_i(t) = kV_s \cos(2\pi \theta/P) \sin(2\pi f_e t) \quad (7.108)$$

where k is the coupling coefficient, V_s is the input voltage, θ is the angular displacement between the inductosyn disks, P is the angular period of one hairpin turn and f_e is the frequency of the excitation voltage. If a second pickup winding is located adjacent to the first, but given an angular displacement of $P/4$ with respect to the first pickup coil, the voltage induced in the second coil is given by:

$$v'_i(t) = kV_s \sin(2\pi \theta/P) \sin(2\pi f_e t) \quad (7.109)$$

Only the amplitudes of v_i and v'_i change with θ/P , their phase and frequency remain the same. If the space quadrature windings of an inductosyn are excited by constant amplitude carriers in time quadrature, the resulting output signal has a constant amplitude, but its phase undergoes a 360° shift for an angular rotation of $\theta = P$. Inductosyns can be divided into absolute rotary sensors and incremental rotary sensors. Inductosyns that have only one pattern of coils, with a large number of electrical cycles, have a small absolute measurement span ($\theta \ll 360^\circ$) and are suitable for sensing incremental position data. Inductosyns used for absolute measurements, where θ can range up to 360° , have two coil patterns. In one approach, a coarse one cycle pattern is

combined with a fine multicycle pattern. The outputs of both patterns are digitized and combined in synchronization or correlation logic to provide an output that is both accurate and absolute. In the other approach, two multicycle patterns are used that differ by one cycle. The outputs of the two, $N/(N-1)$ patterns are digitized and subtracted. The result of this subtraction is digital coarse data, which is then combined with one of the digitized fine outputs to provide an output that is again accurate and absolute.

The available nominal outside diameters of inductosyn 'rings' range 2–42 inches. Limiting accuracies for Inductosyn diameters of 3, 4, 7 and 12 in are ± 2 , ± 1.5 , ± 1.0 and ± 0.5 seconds of arc, respectively. Initial alignment and concentricity of the inductosyn rings is essential in order to realize the high angular precision cited above. Care must also be taken to avoid electrostatic coupling and ground loops to avoid artifact errors in the operation of these sensors. Inductosyns are used with custom electronics packages to give a dc voltage output proportional to angular position between the coils. Since there is no physical contact required between the inductosyn stator and rotor, there are no brushes or wipers to wear out, similar to the case of optical coding disks.

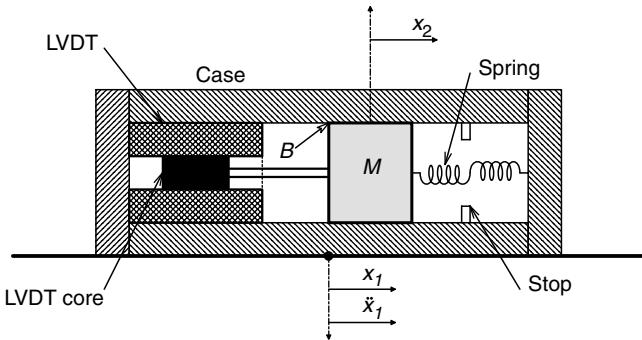
7.3 Measurement of Linear Acceleration, Velocity, Displacement and Position

In this section, we examine certain means of measuring the linear acceleration of structures and vehicles with sensors known as accelerometers, as well as certain techniques for measuring linear velocity and position. The output of a linear accelerometer can be integrated to obtain a velocity estimate, a second integration yields a position or deflection estimate. There are also other means for finding the velocity of particles, fluids, structures and vehicles, including the use of the Doppler effect with light, microwaves or sound. Knowing the precise position of a vehicle is really a navigation problem, and if done from within a closed vehicle such as a submerged submarine or an aircraft flying under IFR, is an inertial navigation problem. On the other hand, the relative position of small objects and structures can be measured with a wide variety of position sensors, including ones which do not physically touch the object. This latter class of no-touch position sensors include the use of ultrasound, microwaves and light in various schemes that utilize the propagation velocity of the waves, interferometry, or phase lock techniques.

7.3.1 Linear Accelerometers

7.3.1.1 A Basic, Newtonian Accelerometer

Figure 7.27 illustrated a basic, passive, moving mass accelerometer. A mass, M , is restrained to slide inside a cylindrical case. There is viscous damping, B , between the case and the mass, and a spring capable of compression or stretch is used to return the mass to its equilibrium position in the center of the case when there is no acceleration. The relative position between the moving mass and the case is measured by an LVDT. The LVDT's body is attached to the case, and the LVDT's moving core is attached to the movable mass. The case itself is firmly attached to the object whose acceleration is to be measured. A typical application of an accelerometer might be to measure the acceleration of a vibration table used to test the mechanical stability of components, or to measure the up/down acceleration of an automobile body while driving over a bumpy

**FIGURE 7.27**

Section through a passive, linear accelerometer. Mass position is sensed by the LVDT.

road. For a mass and spring accelerometer moving in the x direction, we may write, using Newton's law:

$$0 = M\ddot{x}_2 + B(\dot{x}_2 - \dot{x}_1) + K_S(x_2 - x_1) \quad (7.110)$$

Here, x_1 is the position of the case (and the structure whose acceleration we wish to measure), and x_2 is the position of the mass inside the case with respect to an external inertial reference point ($x=0$). This ODE may be Laplace transformed and put in the form of a transfer function:

$$\frac{X_2}{X_1}(s) = \frac{(sB/K_S) + 1}{(s^2M/K_S) + (sB/K_S) + 1} \quad (7.111)$$

Now, we assume that the output of the LVDT is demodulated and filtered to give a voltage proportional to the relative position of the mass in the accelerometer case.

$$V_o = K_o(X_1 - X_2) \quad (7.112)$$

Substitution of equation 7.111 into equation 7.112 gives us:

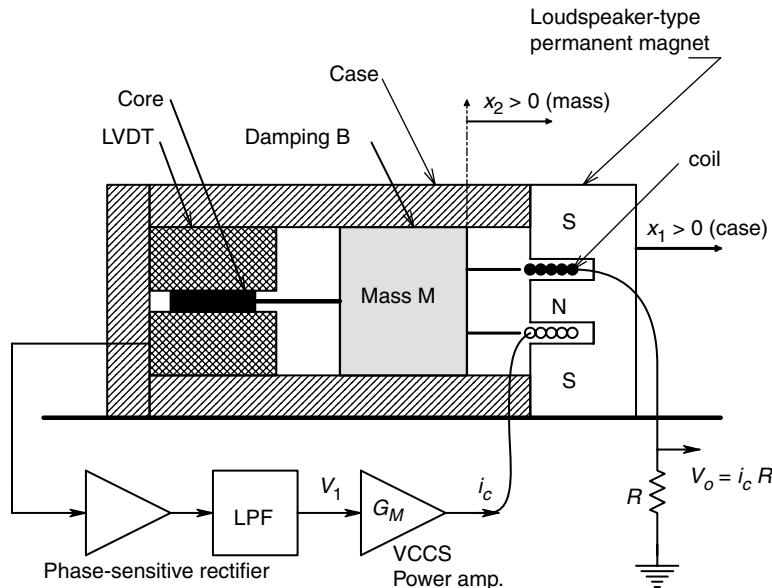
$$\frac{V_o}{X_1}(s) = \frac{K_o(M/K_S)}{(s^2M/K_S) + (sB/K_S) + 1} \quad (7.113)$$

Thus, for a constant acceleration, $V_o = \ddot{X}_1 K_o(M/K_S)$. Linearity of this simple mechanical accelerometer depends on spring linearity over the working range of displacement, as well as LVDT linearity. The natural frequency of this sensor is $f_n = (1/2\pi)\sqrt{K_S/M}$ Hz and the sensor's damping constant is given by $\zeta = B/\sqrt{MK_S}$.

If the magnitude of \ddot{x}_1 becomes too large, the mass will hit the stops in the case and the output voltage will saturate. On the other hand, if \ddot{x}_1 is very small, small amounts of static (coulomb) friction between the mass and case will prevent the mass from moving relative to the case, and there will be an output dead zone. The latter problem can be solved by lubrication, precision machining of the mass/case assembly, or by dithering the mass with an ac solenoidal magnetic field. The former problems can be solved by use of the servo accelerometer design, described below.

7.3.1.2 Servo Accelerometer Design

A servo accelerometer is shown in cross-section in Figure 7.28. We see that the same components are present as for the simple design described above, except the spring has been replaced with a force coil, not unlike the voice coil of a standard loudspeaker.

**FIGURE 7.28**

Section through an active, servo, linear accelerometer. The servo design allows the natural frequency to be raised while preserving sensitivity.

The force coil moves in the air gap of a strong, coaxial permanent magnet, also like that used for loudspeakers. The Newtonian force balance equation can be written:

$$F_C = K_F I_C = M \ddot{x}_2 + B(\dot{x}_2 - \dot{x}_1) \quad (7.114)$$

Here, we assume that the force produced by the coil and magnet is proportional to the current through the coil. In addition, the LVDT system output is $V_1 = K_D(x_1 - x_2)$ and the current is produced by a VCCS with transconductance, G_M . We assume further that the servo accelerometer output voltage is proportional to I_C . Combination of these relations after Laplace transforming yields:

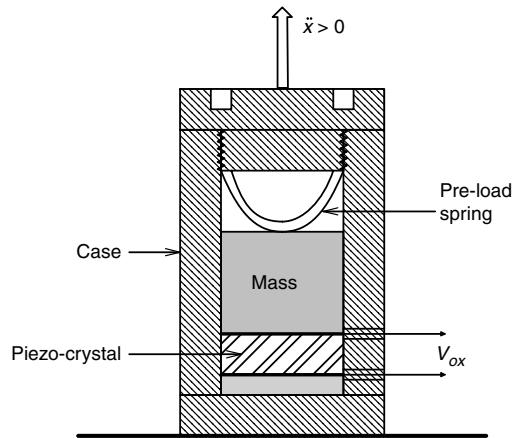
$$\frac{V_o}{X_1}(s) = \frac{K_o(M/K_F)}{s^2 M/(K_F G_M K_D) + sB/(K_F G_M K_D) + 1} \quad (7.115)$$

Now the system's natural frequency is $f_n = (1/2\pi)\sqrt{K_F G_M K_D/M}$ Hz and its damping factor is $\zeta = B/(2\sqrt{M K_F G_M K_D})$.

The feedback design of this active sensor allows the natural frequency to be increased using $G_M K_D$, without lowering sensitivity, as happens with the passive accelerometer. Of course, the complexity of the servo accelerometer design makes it more expensive than the simple moving mass design.

7.3.1.3 Piezoelectric Sensor Accelerometers

One way to eliminate the problem of coulomb friction which accompanies any moving mass design is to not have the mass move relative to the case. Figure 7.29 illustrates a single mass, compression type, linear accelerometer in which the mass is in contact with the surface of a piezoelectric crystal such as quartz, barium titanate, lead zirconatetitanate, etc. The mass is often pre-loaded with respect to the case with a very

**FIGURE 7.29**

A simple piezoelectric crystal accelerometer design shown in cross-section.

stiff spring. Acceleration of the case in the $\pm x$ direction causes a corresponding acceleration of the mass, M . Hence a force, $F_x = \pm M\ddot{x}$ is exerted on the piezoelectric material, which is compression sensitive. Thus, an equivalent current, $i_x = \pm Fd$, is generated in parallel with the below-resonance, $R_x - C_x$ parallel equivalent circuit for the piezo material. It is easy to show that:

$$V_o = \frac{\ddot{x}(s)M d R_x}{sR_x C_x + 1} \quad (7.116)$$

Thus for the accelerometer, we may write:

$$\frac{V_o}{X}(s) = \frac{s M d R_x}{sR_x C_x + 1} \quad (7.117)$$

Equation 7.117 is a simple high-pass transfer function. It tells us that the piezoelectric accelerometer will not measure constant or low frequency accelerations (such as due to gravity), but will work well for accelerations containing frequencies above $f_b = 1/(2\pi R_x C_x)$ Hz. Thus, piezo accelerometers are widely used for vibration testing, where the frequencies of the applied displacement are above f_b . If a charge amplifier is used, the break frequency is now set by the feedback R_F and C_F of the electrometer op-amp. The transfer function becomes:

$$\frac{V_o}{X}(s) = \frac{-s M d R_F}{s R_F C_F + 1} \quad (7.118)$$

The low break frequency, $1/(2\pi R_F C_F)$, will generally be lower than that for the piezo element alone, given the above. The high frequency gain is $-Md / C_F$ V/m/s².

Simple compression type, piezo accelerometers suffer from two drawbacks:

1. a high temperature sensitivity (temperature equivalent acceleration output caused by rapidly changing transducer temperatures)
2. sensitivity to base bending (under the piezo sensor) and off-axis accelerations (linear and angular) which cause the mass to exert force asymmetrically on the piezo material, producing erroneous outputs.

The Delta-Shear piezo accelerometer design developed by Brüel & Kjaer instruments, Inc., minimizes these problems.

Since there is no relative motion between the mass and the case, there is no coulomb friction to cause a dead zone, and the dynamic range of piezoelectric accelerometers can be an enormous 160 dB.

Due to noise and output saturation, amplifiers do not have this dynamic range and so, signal conditioning gains would have to be switched to accommodate an \ddot{x} which varies 160 dB. The mechanical resonance frequency of piezo accelerometers is orders of magnitude higher than that of moving mass and spring accelerometers. This difference is because the stiffness of the piezo material is orders of magnitude larger than that of the spring. Due to their high frequency bandwidth, piezo accelerometers are well suited for vibration testing applications in the kHz range.

7.3.1.4 Micromachined, IC Accelerometers

Figure 7.30 illustrates a simplified schematic top view of the Analog Devices ADXL50, monolithic, micromachined, linear accelerometer. The proof mass is constrained to only

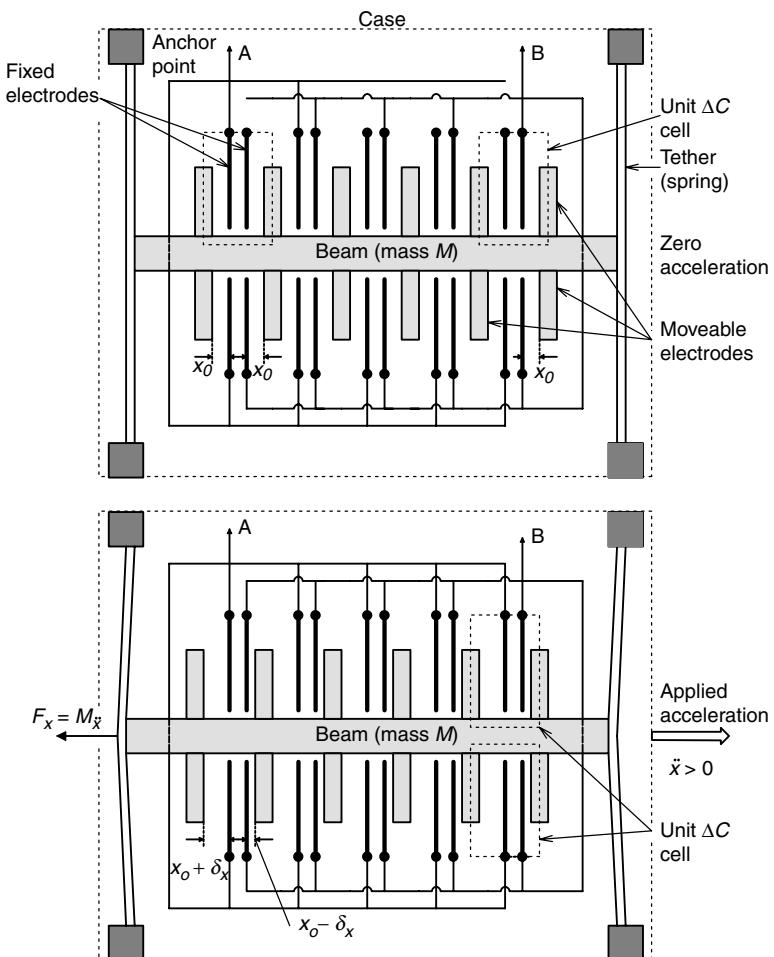


FIGURE 7.30

Simplified top view of the Analog Devices' ADXL50 micromachined, passive, linear accelerometer. Linear position of the linear bar proof mass is sensed by 42 pairs of differential capacitive sensing electrodes. A. Accelerometer with zero linear acceleration. B. Accelerometer is accelerating to the right.

move in the x direction. Micromachined ‘finger’ electrodes protrude from the two sides of the mass bar in the y direction. Between each pair of protruding fingers on each side is a pair of fixed electrodes. Under zero acceleration conditions, the mass bar is centered and the total capacitance between both the ‘A’ and ‘B’ fixed electrodes and the finger electrodes is C_o (in the ADXL50 accelerometer, there are 42 side fingers, 21 on a side). When the case (and mass) are given a constant acceleration to the right, the Newtonian force, $F = M\ddot{x}$, acts on the mass, moving a distance Δx so that the spring restoring force of the tethers equals the acceleration force. That is when:

$$F = M\ddot{x} = K_s\delta x \quad (7.119)$$

The displacement, δx , linearly increases the capacitance between the B electrodes and the fingers, and decreases the capacitance between the A electrodes and the fingers. The A electrodes are driven by a 1 MHz square wave, while the B electrodes are driven by the square wave’s compliment (180° phase shift). The A electrodes have a 3.4 Vdc mean and the B electrodes have a 0.2 V dc mean. The total C_o between each set of fixed electrodes and the mass fingers is 0.1 pF. At a full scale acceleration of 50 g, Δx is such that $\Delta C_{\max}/C_o = 0.1$. The mass of the movable bar and fingers is given as $0.1 \mu\text{g} = 10^{-10} \text{ kg}$, and the minimum resolvable δx is said to be $0.2 \text{ \AA} = 20 \text{ pm}$. The total area of the ADXL50 accelerometer is $500 \times 625 \mu\text{m}^2$. The output sensitivity is 19 mV/g. The ADXL50’s frequency response is low-pass and goes from dc to -3 dB at 1 kHz, and is down 20 dB at 10 kHz. Beam-tether mechanical resonance is at ca. 24 kHz. The equivalent acceleration noise input of the ADXL50 is 6.5 RMS $\mu\text{g}/\sqrt{\text{Hz}}$ noise bandwidth. A more recent Analog Devices’, micromachined, linear accelerometer model, the ADXL105, has a full scale input range of $\pm 5 \text{ g}$ and an input noise of 225 RMS $\mu\text{g}/\sqrt{\text{Hz}}$. The ADXL105 also has a signal bandwidth of 0–10 kHz (-3 dB).

Figure 7.31 shows a simplified schematic diagram of the electronic signal processing done on-board the ADXL50 accelerometer chip. The 42 beam ‘fingers’ are reduced to one equivalent finger and the fixed side electrodes, to a pair. The differential capacitor produces a double sideband, suppressed carrier (DSBSC) modulated signal at the

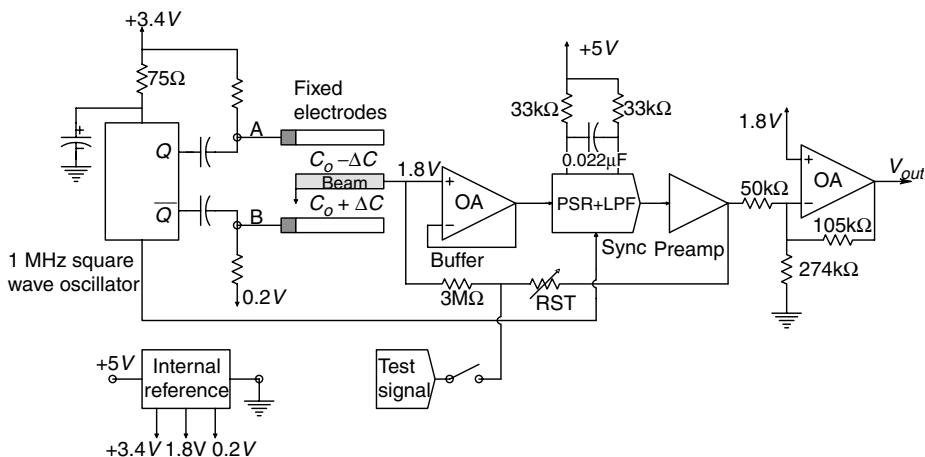


FIGURE 7.31

A simplified schematic diagram of the on-chip signal processing electronics of the ADXL50 series of micromachined linear accelerometers. There are actually 42 fixed electrode pairs and side beams in the system, $V_{out} \propto \ddot{x}$.

summing junction of the buffer amplifier. The amplitude of this signal is proportional to ΔC , hence $|\ddot{x}|$. The phase of the signal depends on the sign of ΔC , hence \dot{x} . The ADXL50 clock frequency is 1 MHz.

Honeywell also offers a line of micromachined linear accelerometers in their ASA line of sensors. Acceleration is sensed simply by a three-plate, differential capacitor. Fixed upper and lower silicon plates sandwich middle plate suspended from a spring, whose mass responds to the applied acceleration by deflecting and causing the desired $C_o + \Delta C$ and $C_o - \Delta C$, which is detected by a bridge and phase sensitive rectifier. The capacitive g sensor of the Honeywell ASA series is much simpler mechanically than the beam and finger architecture used by Analog Devices.

7.3.2 Linear Velocity Measurement

Measurement of linear velocity is subdivided into two classes—velocity measurements on solid or mechanical objects and velocity measurements of fluids—both from a stationary reference position. There appears to be no way to inertially measure vehicle velocity internally, except by integrating the accelerometer output. In some cases, the velocity observer is inside the moving object (e.g. an automobile, boat or airplane), in other cases, the observer is without, but in all cases, there must be an external reference point.

An obvious means of estimating the linear velocity of a mechanical (solid) object is to integrate the output of a linear accelerometer attached to it. This method works well when the SNR at the accelerometer output is high and the integrator does not drift. If the accelerometer output is integrated digitally, at least a 16-bit ADC is needed and the integration routine and sampling interval must be carefully specified.

Generally, other means of velocity measurement depend on the object whose velocity is being measured. We have already discussed the use of laser Doppler velocimetry to measure the velocity of fluids containing small, reflecting particles in Section 6.12.3. LDV can be used with any moving, reflecting surface to measure velocity. LDV has recently been adapted to work with an IR laser diode to measure vehicle velocity, competing with microwave Doppler radar speed detection. Not only is the IR laser beam highly focused, but it is invisible; the IR radar speed ‘gun’ has to be aimed with a telescopic sight. Within the limitations of propagation range, ultrasound can also be used to measure vehicle velocity in a Doppler system. A major problem with all Doppler systems for measuring the velocity of moving, reflecting objects is the necessity to know θ , the angle between the velocity vector and the line connecting the transmitting and receiving transducers with the moving object. This problem has been largely overcome by the crossbeam techniques developed by George and Lumley (1973) for LDV and by Fox (1978) for ultrasound velocimetry.

7.3.2.1 Measurement of Fluid Velocity

We have already described systems to measure fluid velocity by laser Doppler velocimetry, and by the Faraday effect. We now consider several other means.

7.3.2.1.1 Hot Wire and Hot Film Anemometers

The hot wire and hot film anemometer makes use of the fact that the resistance of a metallic wire or film changes with its temperature. The wire temperature, on the other hand, depends on the electric power dissipated in the wire (Joule’s law) and the rate at which the heat from the power is removed by the surrounding medium. The rate at which the heat is dissipated is a nonlinear, increasing function of the velocity of the fluid

surrounding the wire or film. Several representative hot wire and hot film probes made by TSI Inc. are illustrated in Figure 7.32.

Hot wire sensors are used in fluid mechanics research for gas velocity and turbulence measurements. They are generally very fine wires (e.g. 13 µm diameter, 1.3 mm length) stretched between two, gold-plated needle supports. Depending on the gas temperature and velocity range, materials used for the wires can be platinum-coated tungsten (tempco of 0.0042 Ω/°C), pure platinum (tempco of 0.003 Ω/°C) useful up to 300°C, and platinum-iridium alloy (tempco of 0.0009 Ω/°C) useful up to 750°C. Metal film sensors are less fragile. A high purity platinum film is bonded to a fused quartz substrate and then covered with alumina for gas velocity measurements, or with quartz for measuring the velocity of conducting liquids. Metal film sensors can be configured as wires (by bonding the platinum to quartz filaments with diameter as small as 0.001 in, or as wedges (the film is bonded to the end of a quartz 'screwdriver'), or as a flat strip on a flat quartz disk, or as a thin ring around the circumference of the tip of a quartz cone, etc. Film sensors can be used to about 300°C on high temperature probes.

All hot wire and hot film probes have minimum and maximum fluid velocities over which they work effectively. The minimum velocity detectable is set by the amount of free convection from the heated sensor. Velocities down to 0.15 m/s (0.5 ft/s) can be measured in air and down to 0.003 m/s (0.01 ft/s) in water. The maximum velocities which can be reliably measured are primarily limited by mechanical considerations (i.e. the strength of the fine wires). Special hot film sensors have been used in gases with supersonic velocities and in liquids with velocities up to 15 m/s.

Although most fluid velocity measurements are made with single probes or pairs of probes, true, three dimensional vector flow can now be measured with a special 3-wire probe developed by TSI, Inc. This probe is used with a special computer which solves 'Jorgenson's equations' for the 3-wire probe to give the exact, three dimensional velocity vector.

A heuristic, static analysis of a single, hot wire probe is based on several simplifying assumptions. First, we assume that the probe's resistance, R_W , is linearly related to its temperature. Mathematically, this may be expressed as:

$$R_W = R_o(1 + \alpha \Delta T) \quad (7.120)$$

where R_o is the probe's resistance at some reference temperature, T_o , and α is the first order, positive temperature coefficient of the metal wire. Second, we assume that the temperature rise of the probe is given by the heatsink equation:

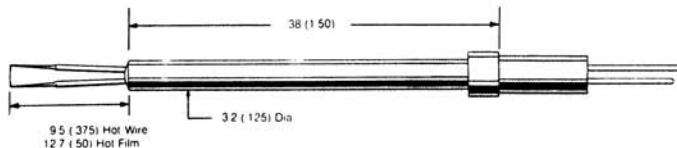
$$\Delta T = P \Omega \quad (7.121)$$

where P is the average electrical power dissipated in the probe, and Ω is the thermal resistance of the probe, in °C/W. Ω is determined by the geometry of the hot probe, the heat conducting properties of the surrounding medium, whether the medium is moving or not and the velocity of the medium relative to the probe's geometry. For the case where the fluid velocity is perpendicular to the axis of the wire, the thermal resistance is given approximately by:

$$\Omega = \Omega_o - \beta \sqrt{|\mathbf{v}|} \quad (7.122)$$

Here, Ω_o is the thermal resistance in still fluid, β is a geometry and fluid dependent constant, and $|\mathbf{v}|$ is the average magnitude of the fluid velocity, assumed perpendicular

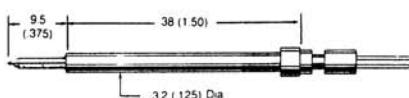
High Temperature Straight Probe - Model 1220



Designed for ambient temperature to 300°C. Suitable for one component velocity measurements.

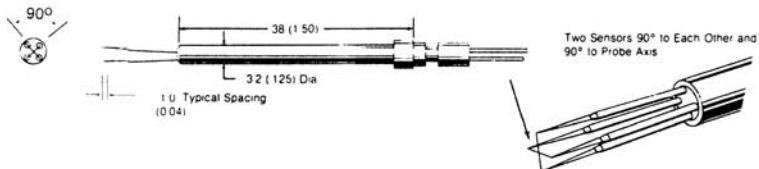
Straight Conical Probe - Model 1230

Sensor Detail



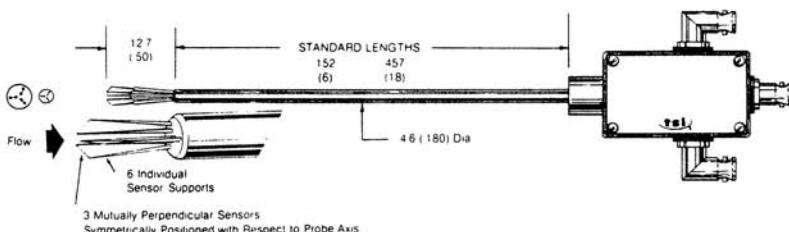
Rugged sensor design resists contamination and breakage. Conical probe shape is used widely in liquid flow measurements.

Standard "X" Probe - Cross Flow - Model 1240



Features two sensors 90° to each other and 90° to the probe axis. For measurements of two turbulence components, correlations and flow vectors in two dimensional flows. Ideal for measurements where space is limited.

3D Probe - Model 1294



Features three mutually perpendicular sensors symmetrically positioned with respect to probe axis. Recommended for high level turbulence (intensities 10%). Model 1146 (1147 for -60 sensor) shield included (see pg. 28).

FIGURE 7.32

Representative hot-wire and hot-film fluid velocity probes. (Source: Courtesy of TSI, Inc.)

to the axis of the wire. In still fluid, ΔT_o can be found from the heatsink equation:

$$\Delta T_o = P_o \Omega_o = I_o^2 R_o (1 + \alpha \Delta T_o) \Omega_o \quad (7.123)$$

Solving for ΔT_o , we obtain:

$$\Delta T_o = I_o^2 R_o \Omega_o / (1 - \alpha I_o^2 \Omega_o) \quad (7.124)$$

I_o is the RMS current through the probe resistor. This relation for ΔT_o can be substituted into the equation for resistance as a function of temperature, and we find:

$$R_{wo} = R_o / (1 - \alpha I_o^2 \Omega_o) \quad (7.125)$$

R_{wo} is the new, elevated, resistance above the ambient resistance of the hot wire probe in still fluid due to electrical power dissipation. If the fluid is moving, we may write:

$$R_w(I, v) = R_o / \left[1 - \alpha I^2 (\Omega_o - \beta \sqrt{|v|}) \right] \quad (7.126)$$

Figure 7.33 illustrates a simple, feedback bridge circuit in which the hot wire probe is maintained at nearly constant resistance and hence, operating temperature. In the circuit, we assume that $R_1 \gg R_w(I, v)$, so that the current through the probe is given by $I = V_B / R_1$. V_o is a dc source used to set R_{wo} . With the switch SW1 open, $V_B = V_S$ and $V_1 = V_{1o}$. Thus:

$$V_{1o} = V_S R_{wo} / R_1 = V_S R_o / [R_1 (1 - \alpha V_o^2 \Omega_o / R_1^2)] \quad (7.127)$$

V_2 is adjusted to have this same value, that is, $V_{2o} = V_{1o}$, so $V_3 = 0$. Exact analysis of the closed loop system is tedious and involves the algebraic solution of a cubic equation. It is useful to assume that the high gain of the amplifier, K_D , forces the closed loop $R_w(I, v)$ to equal the R_{wo} in still fluid. Thus:

$$\frac{R_o}{1 - \alpha V_S^2 \Omega_o / R_1^2} = \frac{R_o}{1 - \alpha V_B^2 (\Omega_o - \beta \sqrt{|v|}) / R_1^2} \quad (7.128)$$

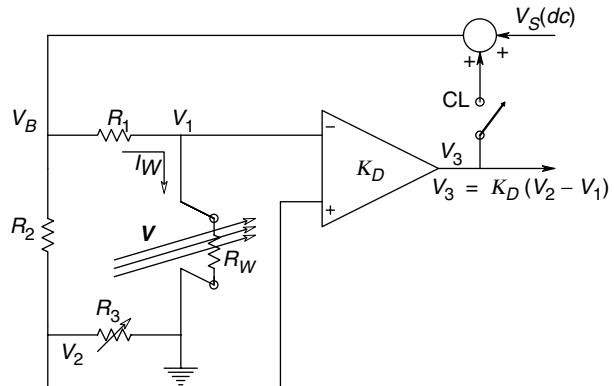


FIGURE 7.33

Schematic of a simple feedback bridge circuit useful with hot wire probes.

From equation 7.128 above, we find that:

$$V_B^2 = V_S^2 / (1 - \beta\sqrt{|v|}/\Omega_o) \quad (7.129)$$

If we assume that $\beta\sqrt{|v|}/\Omega_o \ll 1$, then it is easy to show that:

$$V_B = V_S(1 + \beta\sqrt{|v|}/2\Omega_o) \quad (7.130)$$

and finally,

$$V_3 = V_S\beta\sqrt{|v|}/2\Omega_o \quad (7.131)$$

Thus, the electrical output of this system for conditioning the hot wire anemometer response is proportional to the square root of the fluid velocity. Further linearization can be obtained by passing V_3 through a square law circuit. When fluid flows at an angle to the hot wire probe axis, a reduced response is obtained, the exact form of which depends on the probe geometry, including the supports.

7.3.2.2 Measuring Fluid Velocity and Flow—Other Methods

Fluid flow through a uniform conduit may be subdivided into two major categories—laminar flow, where the velocity vectors have a parabolic distribution, shown in Figure 7.34A, and turbulent flow, which is no longer characterized by a parabolic velocity profile and which no longer contains only parallel streamlines, but also eddies and whorls. The transition between laminar and turbulent flow occurs over a narrow range of (average) velocities. It has been found that the transition between laminar and turbulent flow occurs when a diagnostic variable called the Reynolds number exceeds a threshold value. The Reynolds number is given by

$$N_R = \bar{V} \rho D / \eta \quad (7.132)$$

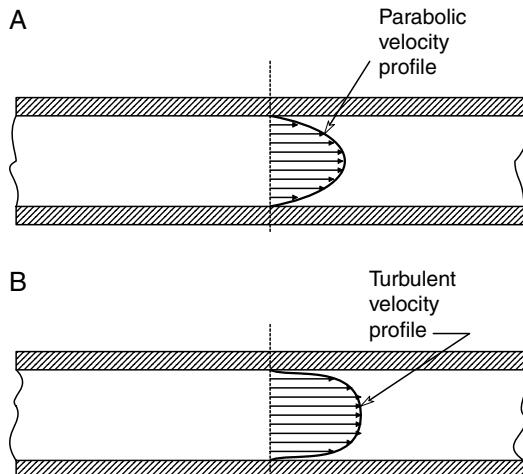


FIGURE 7.34

A. Axial section through a pipe carrying fluid under laminar flow conditions. Note the parabolic distribution of axial fluid velocity with near zero velocity at the pipe walls. B. This pipe is carrying fluid under turbulent flow conditions. Note that the axial velocity component is no longer parabolic with the pipe's radius.

where \bar{V} is the average velocity (LT^{-1}) in the pipe, ρ is the mass density (ML^{-3}) of the fluid, η is the absolute viscosity of the fluid ($ML^{-1}T^{-1}$) and D the internal diameter of the pipe (L). Combining the dimensions above, N_R is seen to be dimensionless. If English units are used (FPS system), a velocity greater than that which will produce N_R in the range 2000–2300, will result in turbulent flow. Needless to say, the mathematics describing turbulent flow are beyond the scope of this text, so we will concentrate on flow and velocity measurements under laminar conditions ($N_R < 2000$). Laminar flow velocity in a pipe with circular cross-section may be described by:

$$V(r) = \frac{(p_1 - p_2)}{4\eta L} (R^2 - r^2) \quad 0 \leq r \leq R \quad (7.133)$$

Here, $(p_1 - p_2)$ is the pressure difference across a length L of the pipe, R is the pipe's radius and r is the radial distance from the center of the pipe. The total volume flow across the cross-sectional area of the pipe may be found by integrating the parabolic velocity profile over the pipe's cross-sectional area.

$$Q = \int v(r)dA = \int_0^R V_m (1 - r^2/R^2) 2\pi r dr = V_m A / 2 \text{ m}^3/\text{s} \quad (7.134)$$

where V_m is the peak velocity, equal to $(p_1 - p_2)R^2 / (4\eta L)$ and A is the pipe's cross-sectional area, πR^2 . Thus, under laminar flow conditions, if we can measure the peak velocity (at the center of the pipe), we can calculate the net volume flow rate, Q .

One common means of measuring the volume flow of incompressible fluids (generally, liquids) is by obstruction meters [Beckwith and Buck, 1961]. These include venturi tubes, flow nozzles, and orifice nozzles. Figure 7.35, reprinted with permission from Beckwith and Buck (1961), illustrates these obstruction meters. What is actually measured is the pressure differential across the nozzle. Beckwith and Buck show that the volume flow through a venturi tube can be expressed as:

$$Q_v = CMA_2\sqrt{2g(P_1 - P_2)/\alpha} \text{ m}^3/\text{s} \quad (7.135)$$

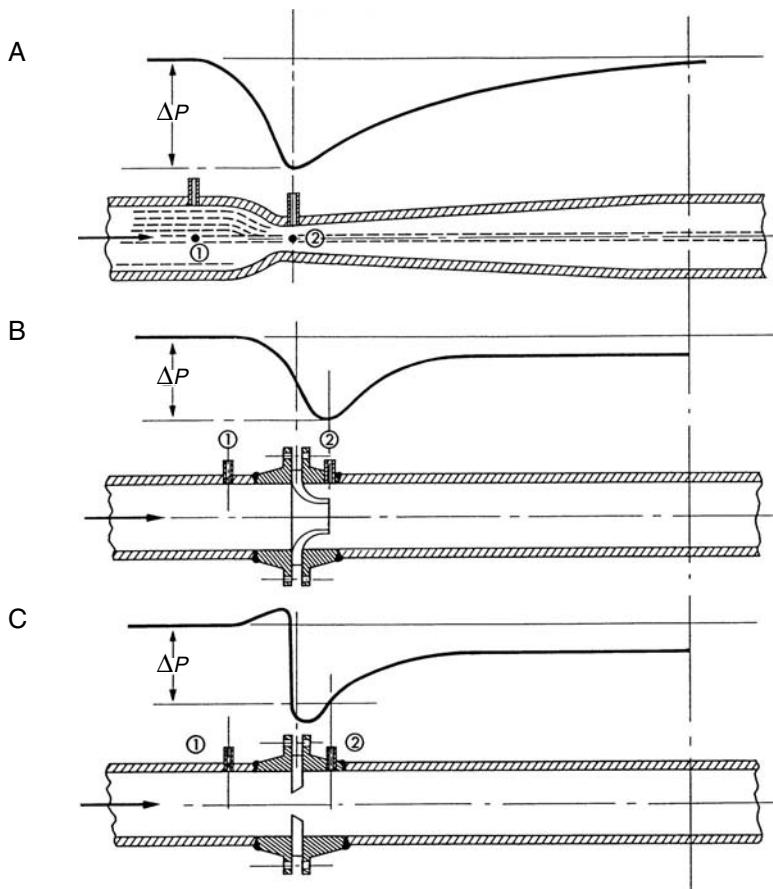
and the volume flow though nozzles and orifices is given by:

$$Q_n = KA_2\sqrt{2g(P_1 - P_2)/\alpha} \text{ m}^3/\text{s} \quad (7.136)$$

where C and K are constants which depend on, among other factors, the N_R and the geometry of the obstruction. M is given by:

$$M = 1/\sqrt{1 - (A_2/A_1)^2} \quad (7.137)$$

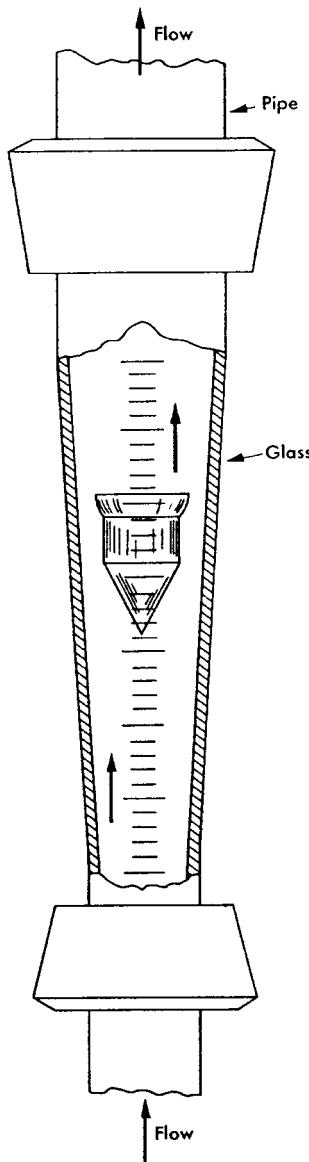
$K = CM$, g is the acceleration of gravity, α is the specific weight of the liquid, A_1 is the upstream pipe cross-sectional area and A_2 is the minimum area of the obstruction. Since the pressure difference, $P_1 - P_2$, is measured, it is clear that the actual volume flow is proportional to its square root. Due to the turbulence losses associated with obstruction flowmeters, there is always a steady state downstream pressure loss, in some cases up to 30% for an orifice flowmeter. There is significantly less steady state pressure loss with a venturi system because of its streamlining and reduced turbulence.

**FIGURE 7.35**

A. Section through a venturi nozzle. The volume flow through the nozzle is proportional to the square root of the pressure difference for incompressible liquids. B and C. Orifice flowmeters. The pressure drop across the orifice is proportional to flow. (Source: With permission, Beckwith and Buck, 1961)

Obstruction meters can also be used with compressible fluids (gases). The situation here is far more complex, because of the adiabatic compression and expansion of the gas as it passes through the obstruction. In this case, the volume flow is no longer a simple square root function of the pressure difference, power law terms involving the ratio of pressures across the obstruction are involved (*cf.* Beckwith and Buck, Section 12–15, 1961).

The fixed geometry obstruction meters suffer from the common problem that the pressure drop is proportional to the square of the flow rate. Thus, a very wide range differential pressure meter is required. If a large range of flow is to be measured, then accuracy may be low at the low end of the range. The rotameter, illustrated in Figure 7.36, is a flow measuring device of extreme simplicity. It consists of a vertical, glass or clear plastic tube whose inner diameter increases toward the top. Inside the tube is a ‘float’, which is pushed upward by the upward flowing fluid. The equilibrium position of the float in the tapered tube is determined by the balance between the downward gravity force and the upward pressure, viscous drag force and buoyant force. Beckwith and Buck give a complex relationship between volume flow of a non-compressible fluid and

**FIGURE 7.36**

Section through a typical rotameter flow meter. (Source: With permission, Beckwith and Buck, 1961)

the height of the float in the tapered tube:

$$Q = (\pi/4)[(D + by)^2 - d^2] \frac{\sqrt{2gv_f(\rho_f - \rho_w)}}{\sqrt{A_f \rho_w}} C \quad (7.138)$$

where D is the effective diameter of the tube at the height of the float, y . v_f is the volume of the float, b is the change in tube diameter (taper) with y , d is the maximum diameter of the float, A_f is the effective cross-sectional area of the float, ρ_f is the float's density, ρ_w is the liquid's density and C is the discharge coefficient.

Rotameters are used for gases as well as incompressible liquids. A major disadvantage in their use is that they cannot be used with opaque fluids or fluids with large amounts of particles in suspension. Rotameters must also be read out manually; they are ill-suited for electronic readout. Advantages of rotameters are that their flow scale is nearly linear and their pressure loss is constant over the entire flow range.

Another type of flow, or velocity meter, makes use of a free spinning turbine wheel in the pipe carrying the fluid. The flow or velocity is a function of the speed of the turbine, which can be measured by a tachometer generator which must be coaxial with the turbine in the pipe. Alternately, if small magnets are placed in the tips of the turbine's vanes, a Hall effect switch sensor can be used outside the pipe to measure the rotating vanes passing under it. The number of pulses/second is proportional to the turbine's angular velocity, which is in turn an increasing function of volume flow. Since no analytical expression exists for turbine velocity as a function of flow or velocity, turbine meters must be calibrated for a given application.

Pressure probes, which include Pitot tubes, are also useful in measuring fluid velocity in ducts or in free space. The Pitot static probe is used to estimate the relative airspeed of aircraft through the air. Here again, we see that the velocity is approximately proportional to the square root of the pressure difference for air. A differential pressure transducer may be used to effect readout. A typical Pitot tube geometry is illustrated in Figure 7.37. Pitot tubes lose calibration when the fluid flow is no longer coaxial to the probe axis. A yaw angle as small as 8° can produce a significant error in the calibration of the Pitot tube. The Kiel ram tube, shown in Figure 7.38, operates similarly to the Pitot tube, except that the reference pressure is not taken from the probe. The Kiel probe is evidently far more insensitive to off-axis air velocity, maintaining its calibration up to yaw angles of 40° [Beckwith and Buck, 1961].

In medicine and physiology, it is important to measure blood flow and velocity in various blood vessels and in the heart. One means of estimating blood flow is by the Faraday effect, already discussed in Section 6.4.6.1. Faraday flow probes have been designed which clamp on to the blood vessel in question. They do not actually contact the blood, but do require surgery to expose the vessel in question. Their use is obviously invasive and not without risk. Webster (1978) describes the designs of several Faraday blood flowmeters. It should be stressed that the Faraday system is responsive to the average velocity of the conducting fluid (blood) and that flow is a derived quantity based on knowledge of the internal diameter of the blood vessel.

The use of Doppler ultrasound, transmitted through the skin to the blood vessel of concern, offers a risk-free, non-invasive means to estimate blood velocity and flow (Webster, 1978, Section 8.4). All ultrasonic blood velocity/flow systems make use of the fact that blood contains erythrocytes or red cells suspended in it. Erythrocytes are typically found in densities of about 5×10^6 per mm^3 , and are biconcave disks about $8 \mu\text{m}$ in diameter, $2 \mu\text{m}$ thick at their edges and $1 \mu\text{m}$ thick at their centers. The velocities of individual erythrocytes follow the laminar flow stream lines of velocity in the larger blood vessels (arteries, veins). Ultrasound incident on the moving blood in a vessel is therefore reflected off myriads of tiny scatterers whose velocities range from zero to maximum, and whose cross-sectional areas vary randomly in time and space (they tumble as they move). Hence the return signal is composed of two major components: The largest component is at the carrier frequency and is reflected from all non-moving tissues in the beam path. The information carrying component is much smaller in energy and is made up of the superposition of many amplitudes and frequencies reflected from the ensonified, moving red cells. The exact nature of the Doppler return signal depends on how tightly the ultrasound beam can be focused in the interior of the vessel to discriminate individual regions of velocity.

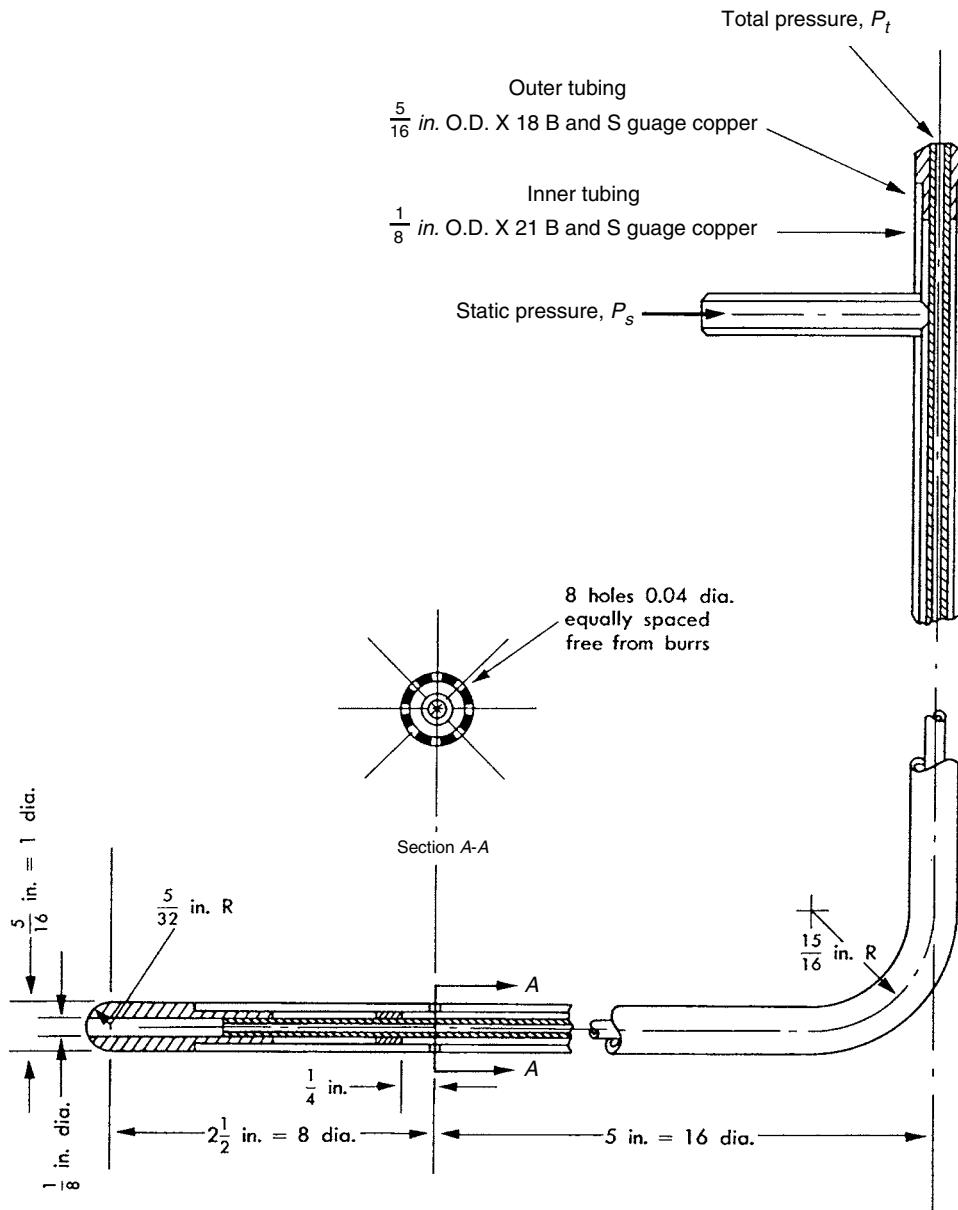
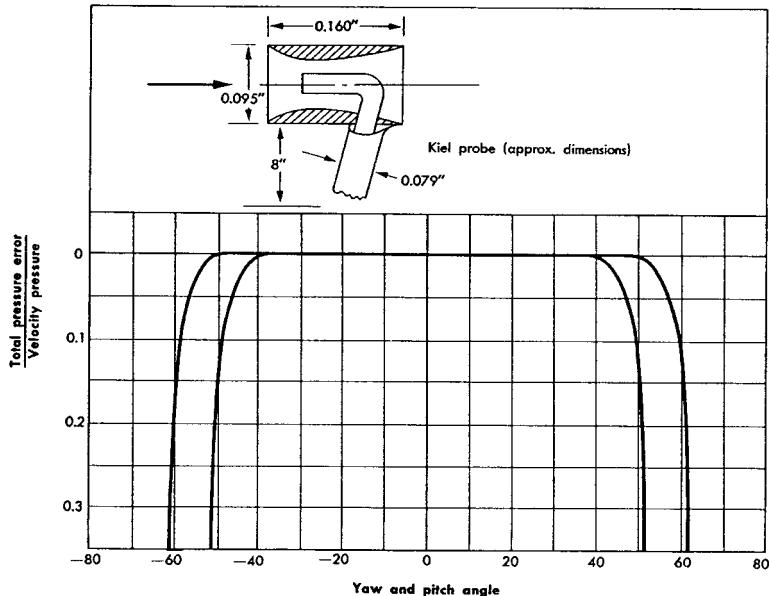


FIGURE 7.37

A pitot airspeed tube. (Source: With permission, Beckwith and Buck, 1961)

The most basic form of Doppler ultrasound system is the CW configuration, where the return signal is heterodyned with a sine wave at the transmitted frequency, then low-pass filtered and amplified. The frequency of this output signal is proportional to the average velocity of the ensonified blood. The output signal is audible to the operator and may be used to detect fetal heartbeats, or pulsating aneurisms in artery walls, in addition to blood velocity in vessels such as the common carotid arteries. In the latter case, obvious asymmetries in carotid blood velocity may indicate an asymmetrical obstruction, such as by atherosclerotic plaque in a carotid sinus. CW Doppler ultrasound systems are generally more qualitative in their medical applications than

**FIGURE 7.38**

A Kiel type air velocity probe. Note its relative insensitivity to off-axis air velocity. (Source: With permission, Beckwith and Buck, 1961)

are the more sophisticated pulsed Doppler systems. Pulsed Doppler systems can be range-gated to enable a blood velocity profile to be constructed for large vessels such as the descending aorta.

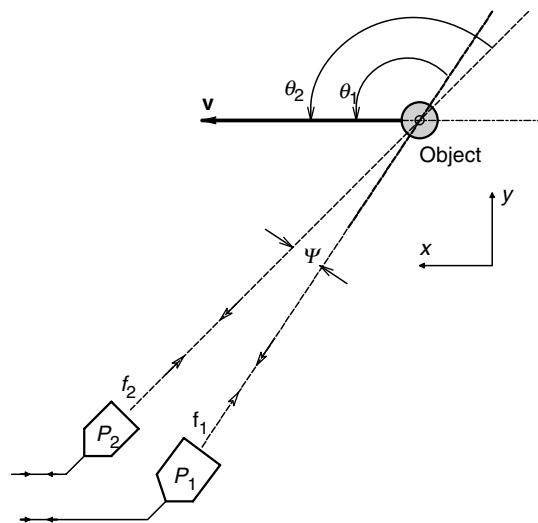
To be used quantitatively, all conventional Doppler blood velocity systems require a precise knowledge of θ , the angle between the blood velocity vector and a line connecting the transmitting/receiving transducers with the small volume of blood whose velocity is being sensed. Recall that the frequency of the Doppler shifted return signal is given by the relation:

$$f_r = f_t \left(1 + \frac{2v \cos(\theta)}{c} \right) \quad (7.139)$$

where c is the average velocity of sound in blood, v is the velocity of the small blood volume ensonified and f_t is the transmitted (carrier) frequency. The Doppler shift, taken as positive here because the velocity of the reflector, has a component toward the transducers, is detected electronically and is given by:

$$f_d = [f_t 2v \cos(\theta)]/c \quad (7.140)$$

Obviously, v is proportional to $f_r/\cos(\theta)$ and an error in θ will give an error in v . In 1985, Fox presented a closed form solution to the two dimensional Doppler situation which utilizes the outputs of two independent transmit-receive probes (Figure 7.39). Fox's solution yields the velocity magnitude, $|v| = \sqrt{v_x^2 + v_y^2}$ and the angle θ_1 between the velocity vector and a line from number 1 probe. Note that the probes lie in the xy plane with the velocity vector. The probes are separated by an angle ψ , and their beams converge on the moving, reflecting object at P . The Doppler frequency returned

**FIGURE 7.39**

Geometry relevant to Fox's angle independent Doppler system. TR1 = ultrasound transducer 1, transmits at frequency f_1 , transducer TR2 transmits at f_2 .

to each probe is given by:

$$f_{d1} = 2f_1|v| \cos(\theta_1)/c \quad (7.141a)$$

$$f_{d2} = 2f_2|v| \cos(\theta_2)/c \quad (7.141b)$$

Now the angle θ_2 can be written in terms of θ_1 :

$$\theta_2 = \theta_1 + \psi \quad (7.142)$$

After some arcane trigonometry and algebra, it can be shown that,

$$|v| = \sqrt{v_x^2 + v_y^2} = (cf_{d1}/2f_1)\sqrt{(1 - 2R \cos(\psi) + R^2)/\sin^2(\psi)} \quad (7.143)$$

and

$$\theta_1 = \tan^{-1}[(\cos(\psi) - R)/\sin(\psi)] \quad (7.144)$$

where R is the Doppler shift ratio corrected for carrier frequency:

$$R \equiv (f_{d2}/f_2)/(f_{d1}/f_1) \quad (7.145)$$

In order to calculate $|v|$ and θ_1 , one must measure f_{d1} and f_{d2} , knowing f_1 , f_2 and ψ . The accuracy of the method is limited by the accuracy which one can determine f_{d1} and f_{d2} . Due to the parabolic (laminar) flow profile in blood vessels, the finite size of the ensonified volume (typically 3 mm^3 for a 2.25 MHz carrier; Fox, 1978) and the random scattering nature of moving red blood cells, there is, typically, a bell-shaped distribution of f_{ds} , rather than a single, sharp peak. The mode of the distribution is generally taken as the desired f_d . The Fox two-probe method of determining the velocity vector in two dimensions is best implemented with a computer system which algorithmically processes the Fourier transformed Doppler return signals to determine their modes to estimate f_{d1} and f_{d2} and then calculates $|v|$ and θ_1 using the relations above and the known parameters, Ψ , f_1 and f_2 .

Fox and Gardiner (1988) extended the two dimensional, closed form solution for v to three dimensions. Their equations are too long to include here, but they have the same general form as the far simpler two dimensional case described above.

Their experimental results showed that the calculated $|v|$ remained within 5.6% of the theoretical value for Doppler angles up to 50° . Also, their angle estimate agreed with the theoretical values with a correlation coefficient, $r = 0.99937$.

The two and three dimensional Doppler flow velocimetry technique developed by Fox and colleagues is, of course, not restricted to the ultrasonic measurement of blood velocity. The Fox technique can be extended to the other Doppler modalities (lasers and microwaves), when a two or three dimensional estimate of object velocity is required.

The final system which we will discuss, which is used for the measurement of gas velocity and flow and which is used in modern pulmonary diagnostic instruments, is the pneumotachometer, or pneumotach. Basically, a pneumotach is a pneumatic analogue of a resistor. A resistor obeys Ohm's law, $I = V/R$. In the case of laminar gas flow through the pneumotach,

$$Q = (P_1 - P_2)/R_p \quad (7.146)$$

where Q is the volume flow of gas, $(P_1 - P_2)$ is the pressure drop across the pneumotach and R_p is the equivalent pneumatic resistance of the pneumotach. The pneumotach itself can have one of several forms. The body or duct of a pneumotach typically has a diameter of 1.5 in and it uses either a fine mesh screen or hundreds of parallel capillary tubes, about 1–2 in long, to obtain the desired acoustic resistance. Unfortunately, a capillary tube does not exhibit a pure, frequency independent resistance, but appears reactive at high frequencies. The acoustic impedance of a single capillary tube is given by [Olsen, 1943]:

$$Z_A(j\omega) = \frac{8L\eta}{\pi R^4} + j \frac{4L\omega\rho}{3\pi R^2} \text{ acoustic ohms} \quad (7.147)$$

where η is the viscosity of air at 20°C (1.86×10^{-4} poise), R is the radius of the tube in cm, L is its length in cm, ρ is the density of air (1.205×10^{-3} g/cm 3) and ω is the radian frequency of pressure variation across the pneumotach elements. If N identical capillary tubes are packed in parallel, the resultant acoustic impedance is:

$$Z_A(\text{NET}) = \frac{8L\eta}{N\pi R^4} + j \frac{4L\omega\rho}{3N\pi R^2} \text{ acoustic ohms} \quad (7.148)$$

The acoustic impedance of a capillary tube pneumotach appears resistive for frequencies from dc up to about one-tenth of the frequency, where $R_p = X_p$, or

$$f_{\max} \approx 0.3\eta/(\pi \rho R^2) \text{ Hz} \quad (7.149)$$

f_{\max} is about 50 Hz for $R = 0.01715$ cm and $L = 1$ cm. This means that as long as the frequency components in the pressure applied across the pneumotach impedance are below 50 Hz, the pneumotach will appear resistive and the pressure drop will be proportional to volume flow. A broader bandwidth pneumotach can be made from thin, parallel, rectangular slits. The acoustic impedance of one such slit is given by [Olsen, 1943]:

$$Z_A(j\omega) = \frac{12\eta L}{t^3 d} + j \frac{\omega 6\rho L}{5 td} \quad (7.150)$$

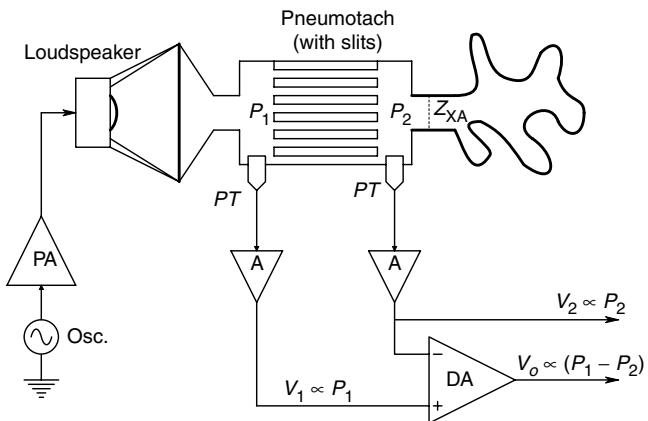
**FIGURE 7.40**

Diagram of pneumotach system used to measure acoustic impedance. Note: Osc = oscillator, PA = power amplifier, A = voltage amplifier, DA = differential amplifier, PT = pressure transducer.

where η is the viscosity of air, ρ is the density of air, L is the length of the slit in the direction of flow, t is the height of the slit perpendicular to the direction of flow and d is the width of the slit perpendicular to the flow direction. The critical frequency for this type of pneumotach is again taken as one-tenth the frequency, where $X_P = R_P$, or:

$$f_{\min} \approx \eta/(t^2 \rho) \text{ Hz} \quad (7.151)$$

For slits with $t = 0.01$ cm, the acoustic impedance of the pneumotach would appear real up to about 240 Hz.

We see that the pneumotach can be used to measure volume flow under laminar conditions by sensing the pressure drop across it with a differential pressure sensor. Average air velocity can easily be derived by multiplying the volume flow by the pipe's area. The pneumotach can also be used to measure the acoustic impedance of a series acoustic load at frequencies where the pneumotach itself appears resistive. Figure 7.40 illustrates such a system. A pressure source (a loudspeaker) is coupled to the pneumotach. Two independent pressure sensors are used to measure the pressure difference across the pneumotach ($P_1 - P_2$) and the pressure across the unknown acoustic impedance (P_2). The unknown acoustic impedance is then just:

$$Z_{XA}(j\omega) = \frac{P_2}{(P_1 - P_2)/R_P} (j\omega) \quad (7.152)$$

This method was used by Pimmel *et al* (1977) to investigate the impedance of the human respiratory system over a frequency range of 1–16 Hz, and by Peslin *et al* (1975) over a frequency range of 3–70 Hz.

7.3.3 Measurement of Linear Position

The technology of linear position measurement is a very broad area, ranging from position measurements of vehicles from within (navigation, inertial navigation and GPS) to position measurements of mechanical objects, radiation sources, estimation of vehicle position from without (as in the location and tracking of aircraft around an airport, or the tracking of submarines). It is not within the scope of this text to consider vehicle location problems by radar or sonar, nor will we consider the specialized topics of inertial navigation and LORAN. We will, however, consider satellite navigation using GPS. First, we will focus on the measurement of mechanical position.

Most mechanical systems operate under coordinate constraints. This means that they may only move in a path determined by their associated connections. For example, the position of the printing head of a dotmatrix printer at any instant is located somewhere along the length of its guide bar—this is a bounded one dimensional position problem which has been solved cleverly by not measuring the position of the print head as the printer operates, but by locating the zero or reference position at printer power-up by use of a simple photoelectric position indicator at the left hand side of the guide bar. Once the zero position is set, a stepping motor is used to position the print head along the guide bar as required. Since the stepping motor rotates in discrete angular increments with every applied pulse sequence, all the printer CPU has to do is use the net count of stepping motor command pulses stored in an up/down counter to know the location of the print head. The discrete nature of printing and symbols allows this simple indexing system to work effectively. However, there are many cases where there must be continuous feedback on position to optimize position control system performance.

In Section 6.5.1, we described the operation of the linear variable differential transformer (LVDT), a sensor widely used to give precise information in bounded linear displacement. LVDTs come in all sizes and cover the range of displacements from millimeters to ten centimeters. An LVDT is an ac, carrier operated sensor with an ac, double sideband, suppressed carrier output, which requires demodulation to a voltage proportional to core displacement by a phase sensitive rectifier (PSR) and low-pass filter (LPF). In some LVDT models, the carrier oscillator and PSR/LPF are built into the LVDT's case. In all cases, the response speed of the LVDT is limited by the carrier frequency and the time constants of the LPF.

One dimensional mechanical position can also be sensed photoelectrically. One scheme we have used makes use of a red LED and two photodiodes or phototransistors connected as shown in Figure 7.41. An opaque vane attached to an axially moving shaft blocks the amount of light impinging on the photodiodes. Near the center of its range,

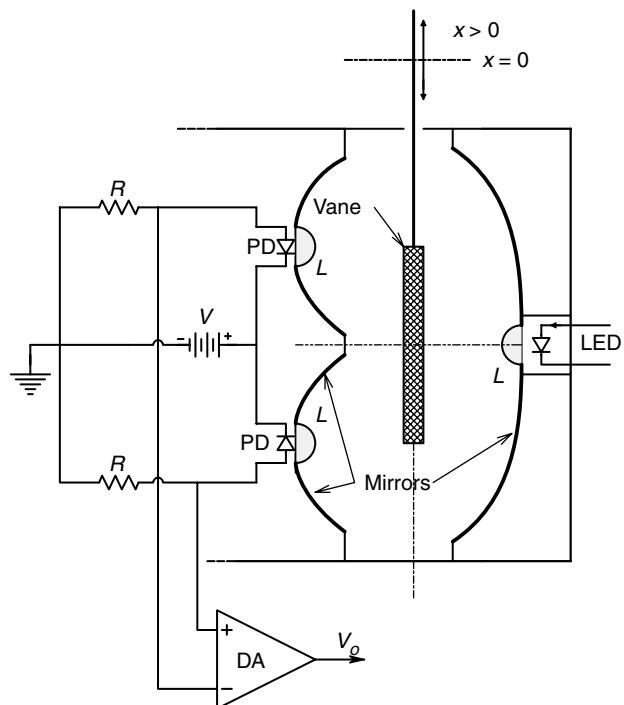


FIGURE 7.41

Section through a mechano-optical system using photodiodes used to measure linear displacements in the millimeter range. Note: PD = photodiode, L = lens, V = opaque vane, LED = light emitting diode.

the voltage output is linear with shaft position. This system was used by the author to provide position feedback to a servo mechanism, which was used to provide controlled stretches to small muscles in physiological studies. The United Detector Technology Co. describes single axis, position sensing photodiodes which can sense the linear position of a collimated spot of light. The UDT sensor can be operated differentially as a fourth quadrant (photovoltaic) solar cell—the ends of the linear sensor are connected to a dc differential amplifier, the center to ground. The spectral response of the UDT sensors peaks at about 850 nm. The sensor can also be used in the reverse biased photodiode mode. Rise times are in the range 0.5–7 μ s, depending on the model. The UDT linear position sensors have maximum displacement ranges of 0.21 or 1.18 in. A 1% output linearity in the central 75% of the length is claimed for the displacement of a 1 mm diameter light spot.

In some cases, linear position can be converted to rotational position by the use of a cable passed around a pulley to a weight. The weight keeps the cable taught and the pulley drives one of many angular position measuring sensors. Probably the simplest angular position measuring sensor for this application is a 10-turn potentiometer. The potentiometer can be used as a voltage divider, or be one arm of a Wheatstone bridge. This type of system has been used to measure the level of liquid in a tank or well.

Linear position of distant objects can also be measured by means of reflected laser light, ultrasound waves and microwaves. In the simplest implementation, as developed for the Polaroid™ camera rangefinder, a short ultrasound pulse is transmitted toward the object whose position is to be sensed and the time delay for the echo to return is used as a measure of the distance. The same principle applies to laser ranging, used in modern surveying and geophysics and with radar, whose uses should be obvious. All of the above pulse-echo systems rely on a precise knowledge of the propagation speed of the energy in the propagating medium. Unfortunately, the speed of sound is affected by air temperature, pressure and moisture content. The speed of sound in water is also affected by temperature, pressure, density and salinity. The speed of light in various media is only slightly affected by factors that affect the speed of sound. Also critical in any pulse-echo system design is the need for a precise, stable clock to measure the time intervals between pulse transmission and echo reception. Such clocks are often derived from thermostatted quartz crystal oscillators using frequency multiplication techniques.

Often, we do not wish to measure the absolute position of an object, but rather a small change in its position. In these cases, we may make use of interferometric techniques, where coherent light is commonly used. Fox and Puffer (1978) described the use of a holographic interferometric system to measure the three dimensional growth of plants in response to applied stimuli such as short flashes of light. Resolution of displacement was on the order of one tenth of a fringe, or 0.16 μ m, using a HeNe laser. Growth movement velocities under constant illumination of 0.4–0.05 μ m/were observed using the cactus, *Stapelia*.

Drake and Leiner (1984) described the use of a single mode, fiber optic, Fizeau interferometer to measure the displacement of the tympanic membrane of the cricket in response to a sinusoidal sound stimulus of 90 dB SPL over a range of 1–20 kHz. Their system could resolve peak displacements ranging 0.01–30 nm. The 0.01 nm limit was set by system noise and the 30 nm limit by linearity considerations in the interferometer. A HeNe laser was used. Drake and Leiner's system is illustrated schematically in Figure 7.42. Note that the distal end of the single mode optical fiber is brought to within a millimeter or so of the reflecting tympanic membrane. This distance is not exceptionally critical except if too large and the power of the return signal picked up by the optical fiber is too small for interferometer operation. The specimen is mounted on a thickness mode piezoelectric crystal which acts as displacement vernier to adjust the exact

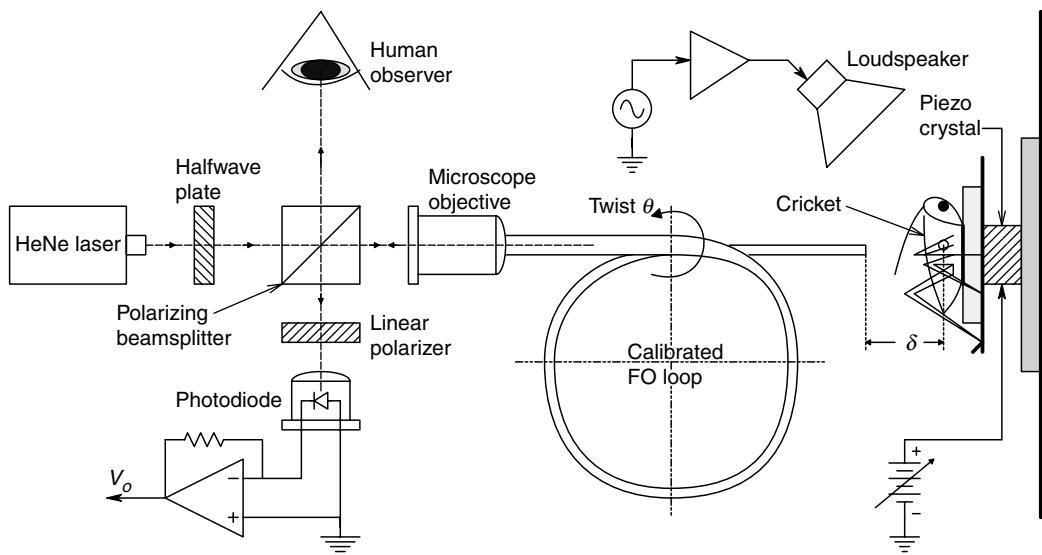
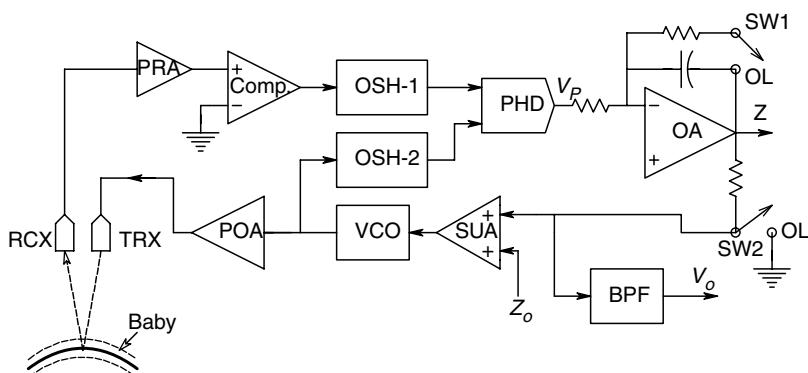
**FIGURE 7.42**

Diagram of the single optical fiber, Fizeau interferometer used by Drake and Leiner (1984) to measure displacements of 0.01–30 nm in a sound excited, insect tympanic membrane.

working distance between the end of the optical fiber and membrane for maximum interferometer sensitivity.

Northrop (1980), and Northrop and Decker (1977) described a 40 kHz ultrasonic system used as an incremental motion detector. The application of Northrop's system was a prototype, no-touch, infant apnea monitor and convulsion alarm for newborn babies at risk. The system's transmit and receive transducers were installed in the top of a conventional incubator system and the CW ultrasound beam was directed down to ensonify the surface of the baby's skin, a good reflecting surface. The return signal was amplified and processed electronically as shown in the block diagram of Figure 7.43. The apnea monitor system output is responsive to the small changes in the ultrasound path length. The frequency of the closed loop system's VCO is changed by feedback so

**FIGURE 7.43**

Block diagram of the 40 kHz, ultrasonic, infant apnea monitor system of Northrop (1980). Note: TRX = transmitting ultrasound transducer, RCX = receiving transducer, PRA = pre-amplifier, POA = power amplifier, COM = analog comparator, OSH = one-shot multivibrator, INT = integrator, BPF = bandpass filter, SUA = summing amplifier, VCO = voltage controlled oscillator, z_o = dc reference voltage.

that at any instant, there is a constant number of wavelengths in the air in the round trip airpath (TRX to BABY to RCX). That is, the controlled variable in this feedback system is the total phase lag between TRX and RCX. It may be shown that the analog output of the apnea monitor is proportional to incremental changes in the total airpath length. Such changes in airpath length would normally be from chest and abdominal movements caused by normal breathing of the sleeping infant. Changes in the normal amplitude and rate of these displacements are what are used to trigger the apnea alarm to summon medical help. The analog output is given by:

$$V_o = K_o \frac{2\pi(N + 1/2)c}{2X^2K_V} \Delta x \quad (7.153)$$

where K_o is a constant, K_V is the VCO constant in (r/s)/volt, c is the speed of sound in air, X is the average path length from the transducers to the baby, N is the integer number of ultrasound wavelengths in $2X$, the total airpath length and Δx is the (incremental) displacement of the baby's chest and abdomen. N is given by:

$$N = INT\left[\frac{2XK_V}{2\pi c}\right]V_C \quad (7.154)$$

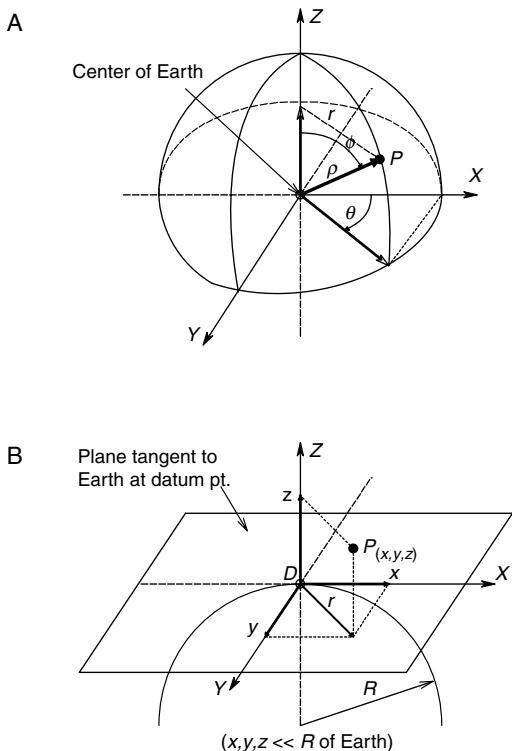
Here, V_C is a constant dc voltage. The ultrasonic apnea monitor was calibrated using an 18 in woofer loudspeaker whose dc displacement sensitivity was known. Limiting resolution of the Northrop system was found to be about $\pm 12 \mu\text{m}$, due to system noise and the wavelength at 40 kHz (0.86 cm). Output sensitivity for $X=21 \text{ cm}$, was 0.92 V/mm. Once the lock is established, the distance X may vary slowly over a wide range, limited only by the working range of the VCO and its ability to keep the steady state relation,

$$f_o = (N + 1/2)c/(2X) \quad (7.155)$$

is satisfied. f_o is the VCO output frequency. Large, rapid changes in X may cause the system to lose the lock. This event can be detected electronically [Northrop and Decker, 1977] and used to relock the system to a new N value. Note that the phase lock design principle of the Northrop system can be extended to microwaves. Such a microwave system might be useful in detecting motion of living persons buried in opaque but otherwise microwave transparent media, such as snow, or rubble from collapsed buildings.

7.3.3.1 Use of the Global Positioning System to Locate Objects

The purpose of the GPS system is to locate a GPS receiver at a position $P(\rho, \theta, \phi)$ in spherical coordinates or $P(x, y, z)$ in rectangular coordinates on or above the Earth's surface (Figure 7.44). GPS location has obvious application to the navigation of moving vehicles, including ships, wheeled and tracked vehicles, aircraft and missiles. GPS location is also used by hikers, explorers, surveyors, mapmakers, oceanographers and farmers. Most recently, its applications have been extended to tracking delivery trucks and keeping track of children at play. In these applications, a precision GPS receiver radios the receiver's coordinates to a base station. Shipboard GPS readouts are generally in Latitude Longitude—modern units can also give course vectors to entered waypoints

**FIGURE 7.44**

A. Spherical coordinates relative to locating an object P at or above the Earth's surface. B. The same object described in rectangular coordinates.

(e.g. buoys), receiver departure from the desired course and position on a moving map. GPS units used by hikers can plot to scale the path the hiker has taken. GPS receiver systems have evolved from the 73 kg Macrometer1000, developed at MIT in 1982 [Langley, 2000] to modern handheld and wrist mounted units in use today. Receiver antennas for the L1 and L2 GPS satellite signals at 1.57542 GHz and 1.22760 GHz, respectively, now are as small as a 2.5 cm square. Of course, larger antennas give stronger signals.

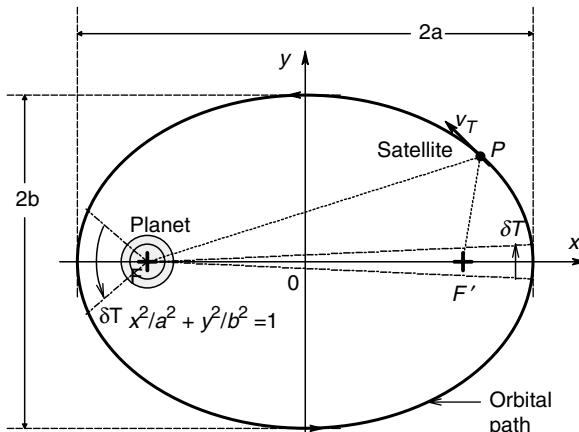
Described below are the four components of the basic GPS system.

1. The satellite vehicles (SVs)
2. The signals they send
3. The ground control system for satellite signal correction
4. GPS receiver design

We will also comment on the Russian GLONASS satellite navigation system and a system being developed by the European Union that may use both GPS and GLONASS technologies, as well as discuss enhancements of the GPS system that give it centimeter accuracy.

It is general knowledge that the Earth and the other eight planets are in orbits around the sun. The moon is in orbit around the Earth, as are many manmade satellites used for surveillance, weather, communications, climate research, astronomy and of course, navigation. The latter class includes the US GPS SVs, the Russian GLONASS SVs and certain SVs in equatorial geosynchronous orbits. The orbits of all orbiting celestial bodies, including SVs, obey Kepler's three laws.

Johannes Kepler was truly a great mind of his era. Kepler lived in the period 1571–1630. He published his First Law in 1605. In 1609, he published his Second Law in his text,

**FIGURE 7.45**

An elliptical orbit of a sattelite vehicle (SV).

Astronomia Nova. He published his Third Law in 1619, in his monograph, *Harmonices Mundi*. It was his third law which led Isaac Newton to his law of gravitation, not an apple. Kepler certainly was the founder of the science of celestial mechanics. In fact, he coined the word satellite. Kepler's three laws can be stated simply for satellites:

1. The orbits of satellites are ellipses (a circle is a special case of an ellipse where the two foci are at the center)
2. If two bodies revolve around each other under the influence of mutual gravitational attraction (i.e. the Earth and a SV), a line joining their mass centers sweeps out equal areas in the orbital plane in equal intervals of time
3. The square of the orbital period is proportional to the cube of the orbit's semimajor axis [Koch, 2002]

Figure 7.45 gives an illustration of the geometry of an elliptical orbit. Note that the elliptical orbit lies in a plane. The long axis of an ellipse is called its major axis. One-half its major axis is by definition, the semi-major axis, a . (In the limiting case where an ellipse becomes a circle, the semi-major axis is the circle's radius.) One-half of an ellipse's minor axis is its semi-minor axis, b . The eccentricity of an ellipse is defined as $\epsilon = \sqrt{(a^2 - b^2)/a}$. It can be shown that the sum of distances from any point P on the ellipse to the two foci is equal to $2a$. Note that focus F' of the ellipse is not centered on a mass, while F is. The GPS SVs are deliberately put in what is as close to a circular orbit as can be obtained. Due to mass inhomogeneities inside the Earth and its very slight degree of oblateness, the actual GPS SV orbits do have a slight eccentricity which is measured and corrected for in the operation of the GPS system.

As we all know, the Earth is not a perfect sphere, it is slightly fatter at the equator, having an mean equatorial radius of 6378 km, a polar mean radius of 6357 km, and a mean radius of 6371 km. Navigation on the Earth's surface has been traditionally done in terms of Latitude and Longitude, which requires accurate maps having latitude and longitude imprinted thereon. Latitude and Longitude are polar coordinates (ϕ and θ , respectively) in a spherical trigonometric system for locating a point P on the Earth's spherical surface. Specifically, latitude is a N/S measure of angular position on a series of circular planes passing through the Earth, perpendicular to its polar axis (the axis the Earth rotates on). The Equator is the 0° latitude locus farthest from the poles. Longitude is an E/W measure of angular position in relation to the Prime (Greenwich) Meridian. The circular planes defining Longitude angles all pass through (or contain)

the Earth's polar axis. Any point on the Greenwich meridian circle is, by definition, at 0° longitude.

The present configuration of the US Block II GPS satellite system uses 24 SVs in six orbital planes (four SVs in each plane). (Only four GPS SVs are required for a receiver to locate itself in (x,y,z,t) space.) There are 21 active SVs and three 'spares'. The orbital planes are spaced 60° apart at the equator, and are inclined at 55° with respect to the (horizontal) equatorial plane. Each of the six orbits has an orbital radius (from the center of the Earth) of 26,562 km. A given SV will appear over a given point on the Earth's surface under its orbit with a period of 11 h, 56 min. Each SV weighs *ca.* 787 kg. Each SV has on board four atomic clocks; two cesium vapor maser and two rubidium vapor maser clocks (Sections 1.4.2 and 8.9). Synchronized precision time keeping between all GPS SVs is essential to precisely triangulate a GPS receiver on Earth. The cesium beam, atomic maser clocks run with an accuracy of $\pm 2 \times 10^{-12}$, amounting to an error of several ns/day. Unfortunately, clocks in orbit are affected by both gravity and relativity.

Quoting Nelson, 1999:

"According to relativity theory, a moving clock appears to run slow with respect to a similar clock that is at rest. This effect is called 'time dilation'. In addition, a clock in a weaker gravitational potential appears to run fast in comparison with one that is in a stronger gravitational potential. This gravitational effect is known in general as the 'red shift' (only in this case it is actually a 'blue shift')."

GPS satellites revolve around the Earth with a (tangential) velocity of 3.874 km/s at an altitude of 20,184 km. Thus, on account of its velocity, a satellite clock appears to run slow by 7 microseconds per day when compared to a clock on the Earth's surface. But on account of the difference in gravitational potential, the satellite clock appears to run fast by 45 microseconds per day. The net effect is that the clock appears to run fast by 38 microseconds per day. This is an enormous rate difference for an atomic clock with a precision of a few nanoseconds. Thus, to compensate for this large secular (*sic*) rate, the (SV) clocks are given a rate offset prior to satellite launch of -4.465 parts in 10^{10} from their nominal frequency of 10.23 MHz, so that on average, they appear to run at the same rate as a clock on the ground. The actual frequency of the satellite clocks before launch is thus 10.2299999543 MHz.

Although the GPS satellite orbits are nominally circular, there is always some residual eccentricity. The eccentricity causes the orbit to be slightly elliptical and the velocity and altitude vary over one revolution. Thus, although the principal velocity and gravitational effects have been compensated by a (clock) rate offset, there remains a slight residual variation that is proportional to the eccentricity. For example, with an orbital eccentricity of 0.02 there is a relativistic variation in the apparent clock time having an amplitude of 46 nanoseconds. This correction must be calculated and taken into account in the GPS receiver.

The displacement of a receiver on the surface of the Earth due to the Earth's rotation in inertial space during the time of flight of the signal must also be taken into account. This is a third relativistic effect that is due to the universality of the speed of light. The maximum correction occurs when the receiver is on the equator and the satellite is on the horizon. The time of flight of a GPS signal to a receiver on the Earth (signal time delay) is then 86 milliseconds and the correction to the range measurement resulting from the receiver displacement (due to Earth's rotation) is 133 nanoseconds. An analogous correction must be applied by a receiver on a moving platform, such as an aircraft or another satellite . . ."

It should be evident to the reader that just the time keeping in the GPS system is complex. In practice, none of the 24 SVs has exactly the same orbital eccentricity.

Data broadcast from all GPS SVs have the following signal structure—two, pseudorandom noise (PRN) codes modulate the transmitted signal. Each SV has

its own unique PRN code. One is the coarse/acquisition (C/A) code the other is the precision (P) code. The C/A code is used by the receiver for satellite acquisition and identification, and for coarse position determination by lower accuracy receivers. The P-code is for higher precision position determination. It can be encrypted to create a secure signal called the Y-code, which is used by the military and other authorized users. The C/A code is a Gold code of register size 10, which has a binary sequence length of 1023 and a clocking rate of 1.023 MHz, thus it repeats itself every 1 ms. The P-code is of length 2.3547×1023 with a clock rate that is ten times the C/A code, or 10.23 MHz.

The GPS satellites transmit at two microwave L-band carrier frequencies, L1 and L2. These frequencies are derived from the SV's master clock by phase locked loop frequency multiplier techniques; the L1 carrier is at $f_{L1} = 154 \times 10.23 \text{ MHz} = 1.57542 \text{ GHz}$, and the L2 center frequency is at $f_{L2} = 120 \times 10.23 \text{ MHz} = 1.22760 \text{ GHz}$. The L1 carrier carries modulation by both the C/A and P-codes, and the Navigation Message (NM). The NM is a 50 bps signal which also modulates the C/A+P signals. It describes the SVs' orbits, clock corrections and other arcane system parameters. The L2 carrier carries only the P-code. The second P-code signal permits a dual frequency measurement of the ionospheric group delay [Nelson, 1999]. Figure 7.46 illustrates the modulation structure for the L1 and L2 signals. The P-code enables a simple GPS receiver to have a 2-sigma RMS horizontal position error of *ca.* 5 m.

Until 2 May, 2000, the US government deliberately degraded the C/A code broadcast from all GPS SVs in a process called Selective Availability (SA). When SA was active, part of the C/A code was 'dithered' to reduce GPS receiver position resolution so that 95% of position readings fell within a 90 m diameter circle. With SA off (the present condition), measured resolution is such that 95% of readings lie inside a 12.6 m diameter circle, an error area reduction of 98%. At the time the US DoD was implementing SA, only military users' GPS receivers could 'un-dither' the GPS signal. Recently, the differential GPS technique (DGPS) was developed to circumvent SA, rendering it obsolete. DGPS is described below.

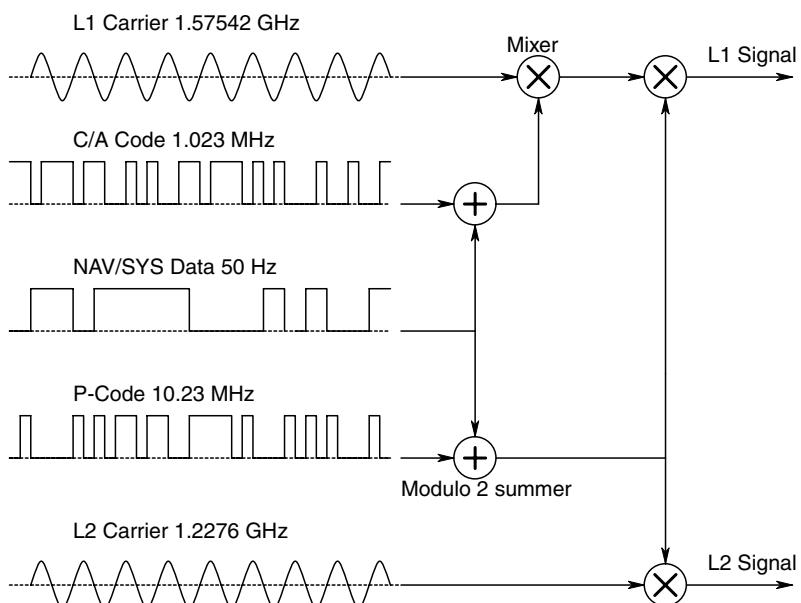
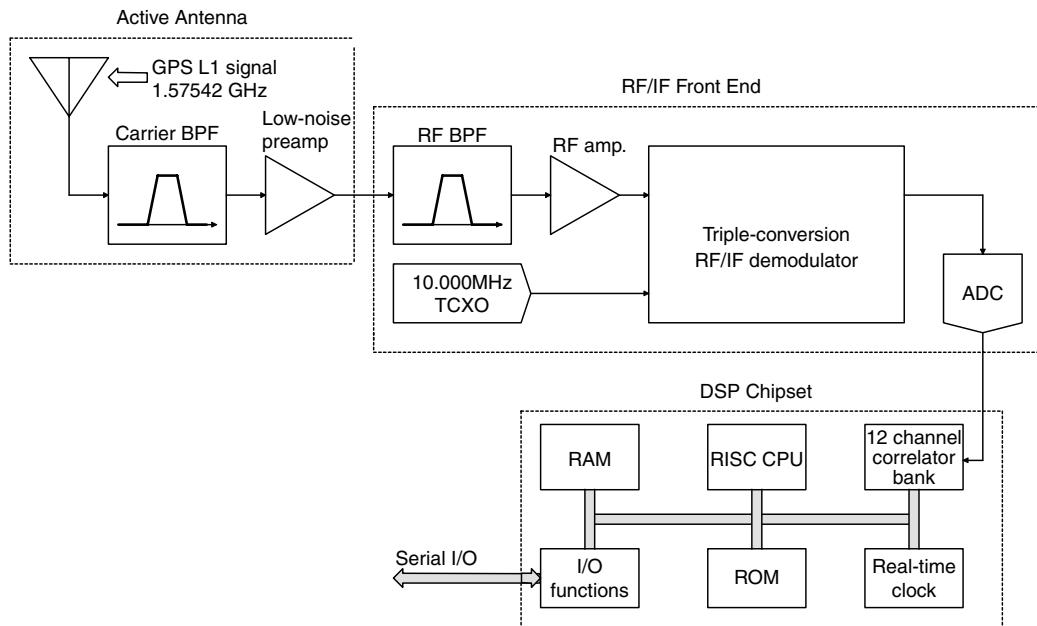


FIGURE 7.46

The modulation structure of the L1 and L2 GPS signals.

**FIGURE 7.47**

Block diagram describing a simple GPS receiver.

The control segment for the GPS system consists of a master control facility located at Schriever AFB in Colorado Springs, CO, and observer stations located in Hawaii, Ascension Island in the S. Atlantic, Diego Garcia Island in the Indian Ocean and Kwajalein Island in the Pacific. These monitor stations measure GPS SV signals and develop precise orbital models (ephemeris data) for each SV, which include clock corrections. This information is sent to the master control station in Colorado, which uploads the necessary ephemeris data and clock corrections to each SV. The SVs then send subsets of the orbital ephemeris data to GPS receivers in the slow navigation message. This continuous updating and correcting is a necessary process to maintain the accuracy of the GPS system.

Figure 7.47 illustrates the block diagram of a simple, civilian use, GPS receiver based on the design used in the Zarlink chipset used in the Orion GPS receiver. Note that it is of superheterodyne architecture—the output of a 10 MHz, temperature controlled quartz crystal local oscillator is processed by an on-chip phase locked loop (PLL) frequency multiplier and mixed to give intermediate frequencies of 175.42 MHz, 35.42 MHz and finally 1.405 MHz. The 2-bit, quantized output of the triple conversion receiver/detector is sent to a DSP package consisting of a parallel, 12-channel signal correlator used for SV identification by its unique PRN signature. The correlator's operation and its output are managed by a 32 bit, RISC microprocessor operating at 20 MHz. On-chip memory includes 512 kbytes 20 ns SRAM and 256 kbytes EPROM holding Orion GPS firmware. The firmware contains the GPS Almanac, which has coarse position (ephemeris) data for all possible GPS SVs. (All GPS receivers must have almanacs.)

An inexpensive, handheld GPS receiver only works with the L1 signal and processes one GPS SV's signal at a time, extracting the data it needs to calculate its position on an alternating basis. More expensive receiver designs have five, independent DSP channels for parallel computation of position. Very high end GPS receivers may have a dedicated channel for each SV in the constellation.

TABLE 7.1

Sources and magnitudes of GPS pseudo range error

Error	Value, m
Ionospheric (refraction) delays	4.0
Clock	2.1
Ephemeris (orbital position)	2.1
Multipath (reflections of UHF signal)	1.0
Tropospheric delays	0.7
Receiver	0.5
Cumulative	10.4

Differential GPS was developed by the US Coast Guard to facilitate precision navigation of military and civilian vessels. It was declared fully operational on 15 March, 1999. DGPS covers the US coastal waters, the Great Lakes, Puerto Rico and the US Virgin Islands, and portions of waters around Alaska and Hawaii. In a DGPS system, an accurately located ground base station receiver makes a position estimate at some time t_0 from five GPS SVs. It compares this position with its accurately known ground position and calculates error vectors. This error information is sent by low frequency radio link (300 kHz carrier) to the mobile DGPS receiver, which calculates a corrected position estimate from its position determined by satellites alone. According to the USCG, the DGPS positional error is 1 to 3 meters [USCG, 2003].

A GPS receiver normally calculates its position by measuring the time it takes the L1 signal from the SV to reach its position. From an *a priori* knowledge of where the SV is in its orbit at the time of transmission (ephemeris data), how long it takes to send the signal, and knowing the speed of the signal, it can compute what is called the pseudo range (PR) to the SV. PRs from four SVs are necessary to calculate the receiver's position, $P(x, y, z)$. (Arcane spherical trigonometry (not covered here) is used to find $P(x, y, z)$.) Each PR is corrected by data from the DGPS base station. The correction data is sent at an amazingly slow 100 or 200 baud, which generates a delay of 2–5 seconds before the corrected receiver position is available for display [DePriest, 2003]. (What were they thinking?). Pseudo range error can be attributed to various causes, as shown in Table 7.1.

The Achilles' heel of the DGPS system is that the farther the receiver is from the base station, the poorer the accuracy of the correction, hence poorer the position estimate by the DGPS receiver. For New England waters, USCG DGPS base stations are located at Sandy Hook, NJ; Horiches, LI, NY; Acushnet, MA; Portsmouth Harbor, NH; Brunswick, ME and Penobscot, ME.

While the 1 m accuracy of the DGPS system is certainly adequate for ship navigation, greater accuracy (<0.1 m) is required for land based operations such as surveying, forestry, robot farming, irrigation, etc. To meet this more stringent requirement, the GPS Wide Area Augmentation System (WAAS) was developed. WAAS is also called WAGPS or Space Based Augmentation System (SBAS). The European Community (EC) is developing their own version of the US WAAS called EGNOS and the Japanese have their version, called MTSAT. We will use the acronym WAAS for the US system. Similar to the USCG sponsored DGPS, which uses base stations largely on the coast, the US FAA sponsored development of the WAAS system uses *ca.* 25 precisely surveyed ground reference stations positioned across the US and several master stations. The WAAS uses its own three GEO SVs which are in high, geosynchronous equatorial orbits. The geosynchronous WAAS SVs useful to the US are called AOR-W, AOR-E and POR. An equatorial geosynchronous orbit has a radius of 35,768 km from the Earth's

center; the GEO SV's tangential velocity is 3075 m/s. SVs in geosynchronous equatorial orbits are stationary above a fixed point on the Earth's equator (i.e. their orbital angular velocity equals the Earth's rotational angular velocity, 4.167×10^{-3} degrees/s). AOR-W is at Lat. 0° (equator) and Long. 54° W. AOR-E is at Lat. 0° , Long. 15.5° W. POR is at Lat. 0° , Long. 178° E, almost mid-Pacific.

WAAS works in the following manner. The WAAS ground stations (and EGNOS ground stations) harvest GPS position information across the country and compare it to their accurate ground positions and each generates an error vector with ionospheric delay, clock errors and ephemeris error components. Each station transmits this information to a master ground station which analyzes and compiles a master error matrix, which is a function of Lat.-Long. coordinates on a rectangular grid over the Earth's surface. The grid is subdivided into eight, North-South strips approximately 45° of Long. wide. Some 1607 points are in the grid. The computed, interpolated/extrapolated GPS error estimates at the grid points are then uplinked to the geosynchronous satellites. An updated clock error matrix is uplinked every minute; ionosphere and ephemeris data are uplinked every two minutes. A WAGPS receiver first determines its position from the four or five GPS SVs it 'sees', then from the downloaded correction data from the GEO SV, it determines where it is (approximately) on the grid, determines the closest four points and interpolates its position correction from the error matrix. The WAAS system can work with three, two or one correction point if others are not available [DePriest, 2003]. Typical accuracy of the WAAS as currently implemented is 1 m (x, y), 1.5–2 m (z) (altitude).

Recall that WAAS was originally developed for aircraft navigation. Due to its wider area availability, several companies have developed WAGPS receivers that not only use the corrections downloaded from a GEO satellite, but phase information between L1 and L2, and sophisticated post processing algorithms. For example, the OmniSTAR 8000HP™ WAGPS system has an advertised horizontal accuracy of 10 cm 95%, and height accuracy of 20 cm 95%. The 95% means that there is a probability of 0.95 that a position reading will lie within a 10 cm radius from the true location for (x, y), and within 20 cm of the true height (z). OmniSTAR lists applications of this precise WAGPS system, including mooring, crane monitoring, seismic surveying, oil drilling, robot vehicle guidance, inshore dredging (e.g. for power cable trenches), trenching for pipelines, power cables, fiber optic cables, farming, rough surveying, etc.

GPS systems are essentially noisy. That is, if a series of position determinations are made using a fixed receiver, each reading is likely to be some distance from the receiver's true position. GPS error, ε , is defined as the vector distance from the true receiver position to the indicated position. There are several ways of statistically describing sets of errors [Wilson, 2003]. One can start by making a histogram of error magnitudes in the horizontal plane for a large number of measurements for the same receiver location. It is found that most such histograms can be fit by the continuous, Weibull distribution with shape factor $\beta=2$, or a continuous Rayleigh distribution. By integration of the Weibull distribution, it can be shown that:

$$\text{Probability}(\varepsilon \leq r) = 1 - \exp[-r^2/\text{MSE}^2] \quad (7.156)$$

where MSE is the mean squared error calculated by:

$$\text{MSE} = (1/N) \sum_{j=1}^N \varepsilon_j^2 \quad (7.157)$$

The root mean squared error (RMSE) is simply the square root of the MSE. From equation 7.156, we can illustrate the three principal error statistics used by GPS manufacturers and evaluators. CEP or circular error probable assumes that $\Pr(\varepsilon \leq r) = 0.5$. The r must be 0.83 for this to happen. This means that for large N , about half the positions determined will lie inside a circle of radius $r = 0.83 \times \text{RMSE}$. If $r = 1 \text{ RMSE}$, 63% of a large number of measurements will lie inside a circle of this radius. Finally the Rayleigh probability model tells us that if $r = 1.73 \times \text{RMSE}$, the probability is that 95% of the measurement points will lie inside a circle with this radius drawn around the known location of the receiver. One problem in evaluating GPS receiver performance in that manufacturers often just give a statement such as 'error = 1 m' without specifying how the error is evaluated.

In summary, we have seen that the evolution of accurate electronic location of a point on the Earth's surface, the ability to navigate on sea and in the air, has evolved from man's ability to launch satellite vehicles into orbits around the earth. This evolution has also depended on the development of integrated circuits and microcomputers, and effective microwave (L-band) radio equipment. Single point measurement position accuracy has increased from early GPS receivers using L1 signals from GPS SVs with selective availability (SA) enabled, which gave a whopping 100 m horizontal error and 156 m vertical error.

Government concern for enemies and terrorists using GPS to guide munitions kept SA active until May 2000, when it was finally turned off. By the late 1990s, several schemes had been developed to enhance basic GPS accuracy, with SA on, rendering SA obsolete. Basic GPS accuracy with SA off is *ca.* 15 m. The differential GPS system developed by the Coast Guard for general ship navigation initially gave 3–5 m accuracy, but now is of the order of 0.75 m using L1 and L2 satellite signals.

WAAS (WAGPS) was developed in the 1990s, specifically for aircraft navigation [FAA, 2003]. It, too, uses surveyed ground stations to determine errors, but instead of directly transmitting the error to the roving receiver, it uploads the processed error data from many surveyed base stations to a geosynchronous SV from which it is downloaded to the receiver. WAAS receivers with post processing can realize 10 cm accuracy. (Note that signal averaging can improve resolution.) We can expect to see mm accuracy in the future, accurate enough for AA or A2 land surveying. Will the transit and theodolite become obsolete for land surveying? Note that GPS accuracy is not without dollar cost. The most accurate GPS receivers must receive and use correction signals, as well as use parallel processing on the received L1 and L2 signals. This electronic and DSP complexity is expensive. Current prices for top-of-the-line WAAS systems lie in the range of 2–5k USD, while handheld units for hikers or boaters cost 100–200 USD.

7.3.3.2 The Use of Optical Interferometry to Measure Δx

There are many applications where it is necessary to accurately measure very small linear mechanical displacements and distances between objects on the order of nanometers. Such applications include the measurement of deflection of diaphragms and cantilevers in force, acceleration and pressure sensors, and thickness measurements in the microfabrication of LSI circuit chips and nanomachines. One means of measuring small distances without contact is to use optical interferometry. An interferometer compares the phase between a reference light beam and a measurement beam by using the phenomena of constructive and destructive interference which occurs when two, coherent, light wave beams are summed on a surface, or a photodetector. Many kinds of interferometers have been developed for various applications in metrology. These include the Fabry-Perot, Fizeau, Mach-Zehnder, Michelson, Nomarsky and Sagnac designs.

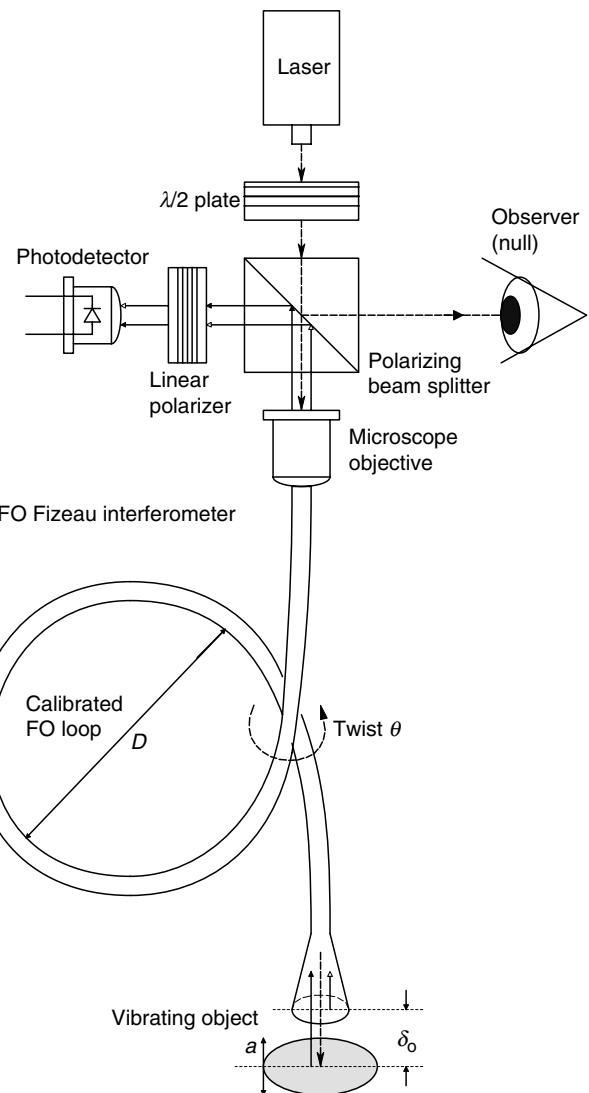
**FIGURE 7.48**

Diagram of a Fizeau interferometer used to measure small $a(t)$.

A number of other interferometer designs exist [Hecht, 1987, Chapter 9; Sirohi and Kothiyal, 1991, Chapter 4]. Figures 7.48, 7.49, 7.50 and 7.51 show simplified diagrams of these interferometers.

There have been several areas of non-invasive medical diagnosis in which information is gained by the no-touch, optical interferometric measurement of extremely small (e.g. of the order of nanometers) mechanical displacements of external surfaces. One such displacement is the extremely small vibration of the eardrum in response to incident sound pressure [Drake and Leiner, 1984]. Such measurement can be useful in the study of the mechanics of hearing, including the tympanal reflex in response to sudden, loud sounds. Another physical displacement is the movement of a tooth *in situ* in response to a lateral force. Such measurement of the mechanical compliance of teeth in gums can be useful in orthodontic research and in the study of dental health. Still another application of optical interferometry lies in the measurement of very small displacements

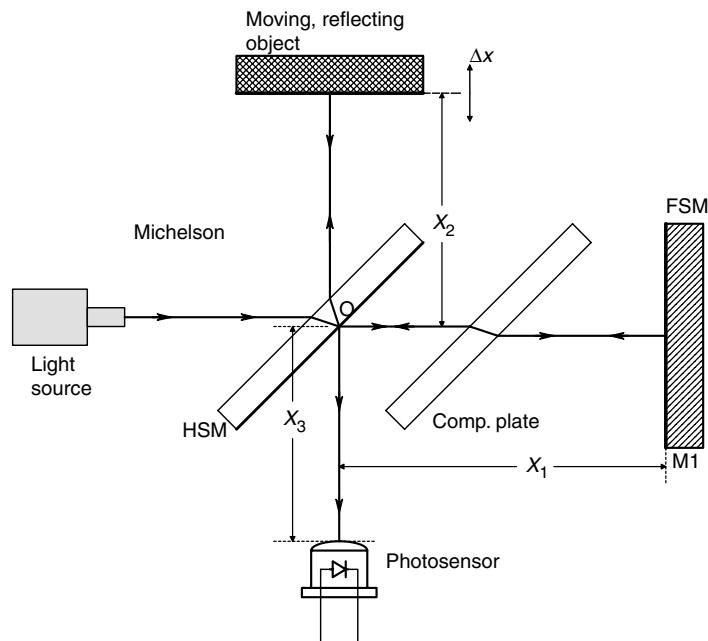


FIGURE 7.49
A classical Michelson interferometer.

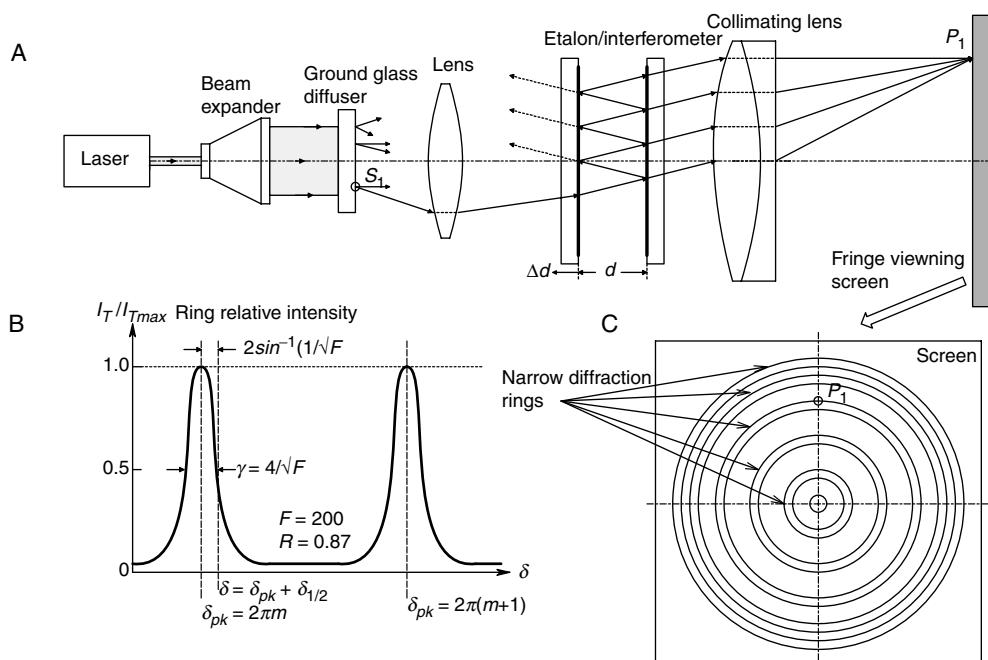
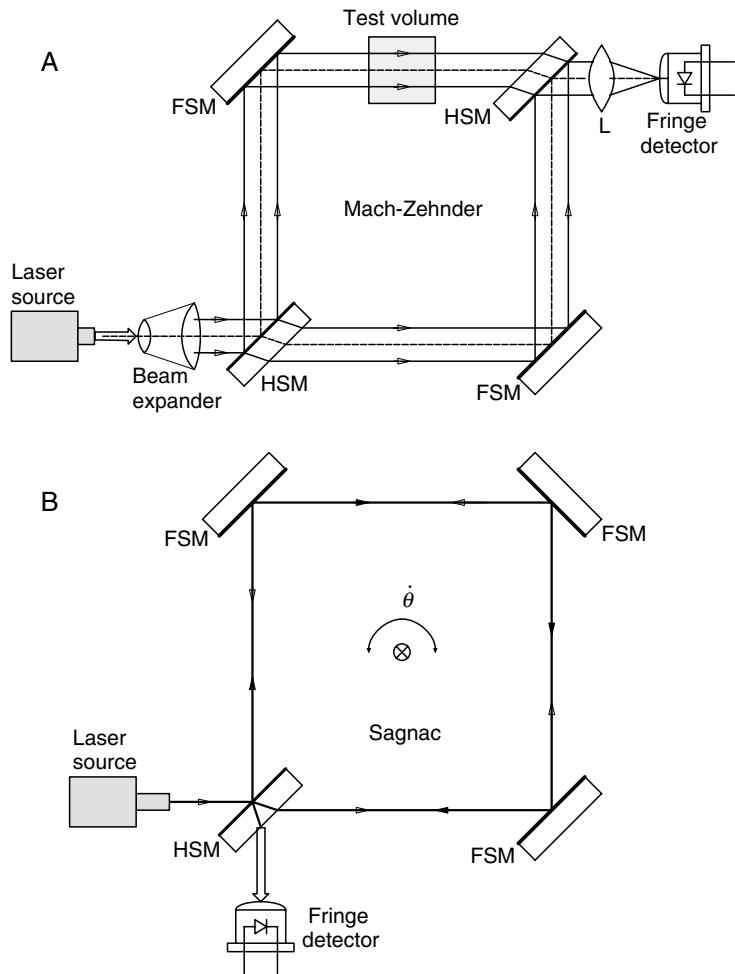


FIGURE 7.50
A. A Fabry-Perot interferometer. B. Relative ring intensity in the interferogram. Note sharpness of peaks. C. View facing interferogram.

**FIGURE 7.51**

A. The Mach-Zehnder interferometer. B. The Sagnac interferometer.

of the skin caused by blood flow in capillary beds. Microplethysmography of the skin surface can have application in studies of the control of peripheral circulation and in assessing microcirculation following vascular surgery. Quantitative measurement (amplitude and frequencies) of skin displacement caused by fine tremors in underlying muscles can have application in studies of neurological disorders, such as Parkinson's disease.

There are many types of optical interferometers which can be used to make the measurements described above. Most use a single coherent light source (i.e. a laser) and split the coherent beam into two optical paths, R and M , then combine the beams in such a manner that there is a linear summation of the output beam E vectors (E_M and E_R) at a common output plane or photosensor. At the output plane, there is alternate constructive and destructive interference. There are bright rings or lines where there is constructive addition of E vectors, dark rings or lines appear where there is destructive (subtractive) interference. In order to measure a physical quantity, such as a small surface displacement, the displacement must cause a phase change between E_M and E_R by changing the M -beam's light wave travel time. The wave travel time (or phase) can be

altered by changing the length of the M -path, or passing the M -ray through a medium having an index of refraction, $n = c/v$, different from the R -path.

As we will see, a convenient way of changing path length is to stick a mirror on the surface under measurement and to use the phase change in the reflected light in the M path when the surface moves. Other physical quantities can be measured by interferometry when they cause a phase change between the M and R beams. For example, when fiber optics are used for the R and M paths, differential temperature changes and mechanical strain or bending of the M fiber can produce phase changes. For an elementary discussion of the use of interferometry in length or displacement measurement, see Chapter 4 and others in the text by Sirohi and Kothiyal (1991). A detailed mathematical analysis of several major types of interferometer can be found in Chapter 9 in Hecht's *Optics* (1987).

Direct measurement of an object's displacement with an interferometer is limited to motions of less than $\pm \lambda/4$. Greater motion will result in a periodic output of the interferometer's photodetector and some means must be used to count interference fringes during the motion of the object to keep track of the total object displacement.

Before describing some specific types of interferometers, let us examine a general, heuristic treatment of optical interference. For interference to occur, the electromagnetic light wave must be temporally and spatially coherent. That is, the frequency and phase of the E vectors of the interfering waves must remain constant in time and space for stable interference to occur. A stable, monochromatic source such as a laser is generally used for interferometry. The optical intensity changes associated with interference occur as the result of the superposition of the E vectors at a point on the detector. The intensity at the detector is proportional to the net E vector squared. We will assume that the R and M rays are linearly polarized in the same direction (e.g. along the x axis). The E vector amplitudes vary in time and space. At a point on the output plane of the generalized interferometer, the x directed E vector amplitudes can be written:

$$e_r = e_r(t) = E_{Ro} \cos[(2\pi\nu)t] \quad (7.158)$$

$$e_m = e_m(t) = E_{Mo} \cos[(2\pi\nu)t + \phi_m + k\Delta x], \quad k = 2\pi/\lambda = 2\pi\nu/cr/m \quad (7.159)$$

$\nu = c/\lambda$ is the electromagnetic lightwave Hz frequency; $\nu = 4.74 \times 10^{14}$ Hz for a 633 nm HeNe laser.

Now the net intensity at the detector is proportional the E vector sum squared:

$$I_d \propto (e_r + e_m)^2 \text{ W/m}^2 \quad (7.160)$$

e_m and e_r are the time varying E vectors of EM plane waves at a point P at the detector surface. The fixed phase of e_r is taken as zero. The phase of the measurement wave, e_m , has a fixed phase lag component, ϕ_m , due to the extra propagation delay and a variable phase component, $k\Delta x$ radians, due to a small, relative displacement of the reflecting surface, Δx , that we are trying to measure. Note that the frequency of both waves is $\nu = c/\lambda$ Hz. Substituting equations 7.158 and 7.159 into equation 7.160, we find:

$$\begin{aligned} I_d \propto & (E_{Ro}^2/2)\{1 + \cos[2\pi(2\nu)t]\} + (E_{Mo}^2/2)\{1 + \cos[2\pi(2\nu)t + 2\phi_m + 2k\Delta x]\} \\ & + E_{Ro}E_{Mo}\{\cos[2\pi(2\nu)t + \phi_m + k\Delta x] + \cos[\phi_m + k\Delta x]\} \text{ W/cm}^2 \end{aligned} \quad (7.161)$$

Of course the photodetector does not respond to terms at lightwave frequencies ν and 2ν Hz. Thus the photo detector output is given by the time average of $(e_r + e_m)^2$ times a constant:

$$V_o = K_d \bar{I}_d = K_d \{(E_{Ro}^2/2) + (E_{Mo}^2/2) + E_{Ro}E_{Mo} \cos[\phi_m + k\Delta x]\} \text{ V} \quad (7.162)$$

We now use the trigonometric identity, $\cos(A + B) \equiv [\cos(A)\cos(B) - \sin(A)\sin(B)]$ where $A = \phi_m$ and $B = k\Delta x$ radians. If Δx is slowly varying (i.e. is a dc quantity), we can force $\phi_m = q\pi/2$, ($q = 1, 5, 9, \dots$) so that $\cos(\phi_m) = 0$ and $\sin(\phi_m) = +1$. Thus, the detector output is given by:

$$V_o = K_d \{(E_{Ro}^2/2) + (E_{Mo}^2/2) + E_{Ro}E_{Mo} \sin[k\Delta x]\} \text{ V} \quad (7.163)$$

From equation 7.163, we see that the photodetector output contains dc terms and a sinusoidal term which has sinusoidal peaks which add to give V_o maxima for $q\lambda/4$, where $q = \dots, -15, -11, -7, -3, 1, 5, 9, 13, \dots$, and minima for V_o at $q = \dots, -13, -9, -5, -1, 3, 7, 11, 15, \dots$. If the net displacement of the object exceeds $\pm\lambda/2$, it is necessary to count maxima and minima in V_o to find the total displacement of the object and to determine whether the displacement is towards or away from the interferometer. By placing a second (quadrature) photodetector next to the first, and delaying the phase of the output light ray to it by $\pi/2$ radians, we can write the second PD output voltage as:

$$V'_o = K_d \{(E_{Ro}^2/2) + (E_{Mo}^2/2) + E_{Ro}E_{Mo} \cos[k\Delta x]\} \text{ V} \quad (7.164)$$

The quadrature detector output, V'_o , has its maxima for $p\lambda/4$, $|p| = 0, 4, 8, 12, \dots$. V'_o minima are at $|p| = 3, 7, 11, \dots$. Electronic threshold logic can detect the maxima and minima of V_o and V'_o , and determine from their phase relation in time whether the object is approaching or receding from the interferometer.

If x is moving sinusoidally at an audio frequency so that $\Delta x = a_o \sin(\omega_m t)$ the dc terms are filtered out by high-pass filtering and again, ϕ_m is set equal to $q\pi/2$, $q = 1, 5, 9, 13, \dots$, there results:

$$v_o(t) = K_d E_{Ro} E_{Mo} \sin[k a_o \sin(\omega_m t)] \quad (7.165)$$

To keep the system in its linear output range, the argument of the $\sin[*]$ term must not exceed $\pm\lambda/4$. That is, a_o must be $<\lambda/4$. If $a_o > \lambda/4$, periodicity in the output vs a_o is observed, and unless one counts successive output maxima and minima, there will be ambiguity in the true value of a_o . Equation 7.165 can be expanded by the Bessel-Jacoby identity [Stark, Tuteur and Anderson, 1988]:

$$v_o(t) = K_d E_{Ro} E_{Mo} \sum_{n=-\infty}^{\infty} J_n(\xi) \sin(n\omega_m t) \quad (7.166)$$

where clearly, $\xi \equiv ka_o$ and $J_n(\xi)$ is a Bessel function of the first kind. It can be shown that $J_n(\xi)$ can be approximated by the series [Stark, Tuteur and Anderson, 1988]:

$$J_n(\xi) = \sum_{k=0}^{\infty} \frac{(\xi/2)^{(n+2k)} (-1)^k}{(n+k)! k!} \quad (7.167)$$

If $\xi \ll 1$, equation 7.167 can be approximated by:

$$J_n(\xi) \cong (\xi/2)^n / n! \quad n > 0, \text{ integer} \quad (7.168)$$

and in general [Dwight, 1969, 807.4],

$$J_{-n}(\xi) \equiv (-1)^{|n|} J_n(\xi) \quad (7.169)$$

Thus, the first few terms of equation 7.167 can be written, assuming $\xi \ll 1$:

$$v_o(t) \cong 0 + K_d E_{Ro} E_{Mo} \left\{ \xi \sin(\omega_m t) + 0 + (\xi^3/24) \sin(3\omega_m t) + 0 + (\xi^5/1920) \sin(5\omega_m t) + 0 + \dots \right\} \quad (7.170)$$

Since even, high order harmonic terms are zero and $\xi \ll 1$, the fundamental frequency term dominates the series of equation 7.167. a_o , the QUM, can be found by phase sensitive rectification of $v_o(t)$ using a ω_m frequency reference signal, followed by low-pass filtering (i.e. by using a lock-in amplifier). If the low-pass filter has gain K_F then its dc output, \overline{V}_o , will be:

$$\overline{V}_o = K_F K_d E_{Ro} E_{Mo} k a_o \text{ V dc} \quad (7.171)$$

We next examine two examples of optical interferometers that have been used in biomedical applications. The first is the fiber optic, Fizeau interferometer developed by Drake and Leiner (1984).

7.3.3.2.1 Measurement of Tympanal Membrane Displacement by Fiber Optic Fizeau Interferometer

This prototype instrument, shown in Figure 7.48, was used to sense the minute displacement of the tympanal membrane of a common cricket subject to external audible sound. (Figure 7.42) Drake and Leiner used a 15 mW HeNe (633 nm) laser as their source. The half-wave plate was rotated to minimize the observed intensity, indicating that all of the linearly polarized light from the laser was going through the polarizing beamsplitter into the 10×0.255 NA microscope objective used to direct all of the incident linearly polarized light (LPL) into the proximal end of the polarization preserving, glass optical fiber (PPOF). The PPOF was of two-step design—it had a $5 \mu\text{m}$ core diameter and a $125 \mu\text{m}$ cladding diameter. (Note that the expanded distal end of the OF is drawn for clarity to illustrate the fate of rays at the fiber's end; it is not really there.) By experimentally adjusting the PPOF loop diameter, D , and twisting it some small angle, θ , around the axis from the objective to the object, the PPOF loop could be made to act as a quarter-wave plate. Thus the light returning to the proximal end of the PPOF was rotated 90° with respect to the entering light. Drake and Leiner stated:

"All of the interfering light is then reflected toward the detector by the polarizing beam-splitter, while light reflected from the back side of the beam-splitter, from the elements of the microscope objective, and from the proximal end of the fiber, all pass straight back through the polarizing beam splitter and does not contribute unwanted radiation at the detector."

Since the Fizeau interferometer gives an output signal proportional to relative object displacement, Drake and Leiner mounted their object (a cricket tympanum) on

a piezoelectric crystal that could be displaced toward or away from the PPOF's distal end by $>\pm\lambda/2$, by applying a 10 Hz, triangular, voltage waveform and observing the magnitude of the ac signal output at the acoustic stimulus frequency, f_m . The ac output was maximum when the distance, a , was some multiple of $\lambda/4$. This particular dc voltage was applied to the crystal for subsequent tests of tympanal membrane displacement response, Δx . Calibration of the Fizeau interferometer was accomplished by applying a 1 kHz triangular wave around the dc bias voltage to the piezocrystal. The triangle wave amplitude was adjusted until the output peaks just began to fold over as seen on an oscilloscope display. The target was then known to be undergoing a $\pm\lambda/4$ displacement. Thus, an output calibration factor was determined in V/nm displacement. ($\lambda/4 = 158.2 \text{ nm}$)

Drake and Leiner displayed their output signal on an oscilloscope and on a spectrum analyzer. The typical ratio of fundamental frequency to second harmonic amplitude was 0.16. They found that their instrument could resolve $0.01 \text{ nm} < \Delta x_o < 30 \text{ nm}$ linearly. This is an amazing sensitivity considering that a lock-in amplifier was not used. They commented that the working distance, a , from the distal end of the fiber to the moving object can be increased to over a few mm by adding a Selfoc® lens to the tip of the PPOF cable.

Drake and Leiner mention that future applications of their Fizeau interferometer can include *in vivo* studies of human eardrum displacement, measurement of the basilar membrane deflection in the cochlea (presumably done *in vitro*), measurement of nerve axon displacement during an action potential and dental strain measurements.

Kaiser *et al* (2001), reported the RMS resolution of their improved, phase shifting, Fizeau interferometer to be *ca.* $\lambda/10^3$ or 0.6 nm. Their Fizeau interferometer altered the lightwave frequency of their LAD source to obtain a fine null.

7.3.3.2.2 Measurement of Skin Vibration by Optical Interferometry

The Michelson interferometer has been used to measure skin surface micro vibrations [Hong and Fox, 1993; Hong, 1994]. A true Michelson interferometer is illustrated in Figure 7.49 [Hecht, 1987]. Note that it uses a compensating plate in optical path X_1 so that each beam travels the same distance through glass. The compensating plate is identical to the half-silvered mirror beam splitter, except that it does not have silvering. Use of the compensation plate corrects for refractive index dispersion with λ and lets the Michelson interferometer be used with broadband (semi-coherent) light sources. If a monochromatic, laser source is used, the compensation plate is not necessary. A Michelson interferometer without the compensation plate is often called a Twyman-Green interferometer [Sirohi and Kothiyal, 1991]. As we showed above with the Fizeau interferometer, the optical intensity at the photodetector is proportional to the square of the sum of the two E vectors impinging on the detector. Let us take point O on the half-silvered mirror as the phase origin. Neglecting the half silvered mirror thickness, we can write for the Twyman-Green/Michelson interferometer:

$$e_r(t) = E_{Ro} \sin[2\pi\nu t + k(2X_1 + X_3)] \quad (7.172)$$

$$e_m(t) = \eta E_{Ro} \sin[2\pi\nu t + k(2X_2 + X_3)] \quad (7.173)$$

where the lightwave frequency $\nu = c/\lambda \text{ Hz}$, ($4.74 \times 10^{14} \text{ Hz}$ for a HeNe laser), $k \equiv 2\pi/\lambda$, the distances are defined in Figure 7.49, and $\eta < 1$ represents the fraction of incident

light reflected back to the detector. We know that the intensity at the detector is proportional to $[e_r(t) + e_m(t)]^2$:

$$\begin{aligned} I_d \propto & (E_{Ro}^2/2)\{1 - \cos[2\pi(2\nu)t + k(2X_1 + X_3)]\} + \eta^2(E_{Ro}^2/2)\{1 - \cos[2\pi(2\nu)t + k(2X_2 + X_3)]\} \\ & + 2(E_{Ro}\eta E_{Ro})\sin[2\pi\nu t + k(2X_1 + X_3)]\sin[2\pi\nu t + k(2X_2 + X_3)] \end{aligned} \quad (7.174)$$

Since the detector does not respond to the light frequency terms, equation 7.174 reduces to:

$$I_d \propto (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \cos[2k(X_1 - X_2)] \quad (7.175)$$

Using the $\cos(A - B) \equiv [\cos(A)\cos(B) + \sin(A)\sin(B)]$ identity, we can write:

$$I_d \propto (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \{\cos[2kX_1]\cos[2kX_2] + \sin[2kX_1]\sin[2kX_2]\} \quad (7.176)$$

Now, we adjust X_1 so that $\cos[2kX_1] \rightarrow 0$ and $\sin[2kX_1] = 1$. We can now consider a small sinusoidal skin vibration at frequency f_s and peak amplitude, $\Delta x_o < \lambda/4$:

$$X_2(t) = X_{2o} + \Delta x_o \sin(2\pi f_s t) \quad (7.177)$$

Thus:

$$I_d \propto (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \sin[2kX_{2o} + 2k\Delta x_o \sin(2\pi f_s t)] \quad (7.178)$$

Using the $\sin(A + B) \equiv [\sin(A)\cos(B) + \cos(A)\sin(B)]$ trigonometric identity, we can finally write:

$$\begin{aligned} I_d \propto & (E_{Ro}^2/2) + \eta^2(E_{Ro}^2/2) + \eta E_{Ro}^2 \{\sin[2kX_{2o}] \cos[(2k\Delta x_o) \sin(2\pi f_s t)] \\ & + \cos(2kX_{2o}) \sin[(2k\Delta x_o) \sin(2\pi f_s t)]\} \end{aligned} \quad (7.179)$$

Again, the $\cos[\xi \sin(2\pi f_s t)]$ and $\sin[\xi \sin(2\pi f_s t)]$ terms can be represented as Bessel functions [Stark, Tuteur and Anderson, 1988], *sic*:

$$\cos[\xi \sin(2\pi f_s t)] = \sum_{n=-\infty}^{\infty} J_n(\xi) \cos(n2\pi f_s t) \quad (7.180)$$

$$\sin[\xi \sin(2\pi f_s t)] = \sum_{n=-\infty}^{\infty} J_n(\xi) \sin(n2\pi f_s t) \quad (7.181)$$

where $\xi \equiv (2k\Delta x_o)$. Thus the electrical output of the detector contains dc terms, a fundamental frequency term (at f_s) and harmonics (at nf_s , $n \geq 2$). The dc energy can be eliminated by a high-pass filter and the fundamental frequency terms can be selected by a band-pass filter, or by use of a lock-in amplifier. Thus the output of the Michelson interferometer is nonlinear, even with a reflecting object's small motion.

If the object motion is fast enough, there will also be a significant Doppler frequency shift associated with the lightwave frequency of $e_m(t)$. This shift can be expressed as

a change in frequency of the e_m in the \mathbf{E} vector:

$$e_m(t) = \eta E_{Ro} \sin[2\pi v_m t + k(2X_2 + X_3)] \quad (7.182)$$

and

$$e_r(t) = E_{Ro} \sin[2\pi v t + k(2X_1 + X_3)] \quad (7.183)$$

where, $v_m \equiv v + \Delta v$. The Doppler shift is $\Delta v \equiv v(2\dot{X}_2/c)$ Hz [Northrop, 2002]. Now, ignoring light frequency terms, the intensity is of the form:

$$I_d \propto (E_{Ro}^2/2)(1 + \eta^2) + 2\eta E_{Ro}^2 \cos(A) \cos(B) \quad (7.184)$$

where $A \equiv [2\pi v t + k(2X_1 + X_3)]$ and $B = [2\pi v(1 + 2\dot{X}_2/c)t + k(2X_2 + X_3)]$. Using the identity, $\cos(A) \cos(B) \equiv (\frac{1}{2})[\cos(A - B) + \cos(A + B)]$, we have:

$$\begin{aligned} I_d \propto & (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \{\cos[(2k\dot{X}_2)t + (2k)(X_2 - X_1)] \\ & + \cos[2\pi(2v)t + (2k\dot{X}_2)t + (2k)(X_1 + X_2 + X_3)]\} \end{aligned} \quad (7.185)$$

The second term in the $\{\cdot\}$ of equation 7.185 is dropped because it is twice lightwave frequency. The first term is at the Doppler shift frequency, $\Delta v = v(2\dot{X}_2/c)$ Hz, which is generally in the audio range and resolvable by the photosensor. We can now use the identity, $\cos(C + D) \equiv [\cos(C)\cos(D) - \sin(C)\sin(D)]$, to find:

$$I_d \propto (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \{\cos[(2\pi v)(2\dot{X}_2/c)t] \cos[(2k)(X_2 - X_1)] \quad (7.186)$$

If the skin motion is sinusoidal so that

$$X_2(t) = X_{2o} + a_o \sin(\omega_s t) \quad (7.187)$$

Then

$$\dot{X}_2 = \omega_m a_o \cos(\omega_s t) \quad (7.188)$$

We set $X_1 = X_{2o}$ and after some algebra, we get the messy result:

$$\begin{aligned} I_d \propto & (E_{Ro}^2/2)(1 + \eta^2) + \eta E_{Ro}^2 \{\cos[4\pi\delta_2/\lambda](\omega_s \cos(\omega_s t))t] \cos[(2ka_o) \sin(\omega_s t)] \\ & - \sin[(4\pi\delta_2/\lambda)(\omega_s \cos(\omega_s t))t] \sin[(2ka_o) \sin(\omega_s t)]\} \end{aligned} \quad (7.189)$$

The $\cos[\xi \sin(\omega_s t)]$ and $\sin[\xi \sin(\omega_s t)]$ terms in equation 7.189 give rise to the motion frequency Bessel function series of equations 7.180 and 7.181 above. It is not known what sort of temporal modulation of the Bessel series the Doppler $\cos[\xi \omega_s t \cos(\omega_s t)]$ and $\sin[\xi \omega_s t \cos(\omega_s t)]$ terms will provide. For example, let us calculate the maximum Doppler shift for a peak sinusoidal skin deflection of 120 nm at 1 Hz, using a HeNe laser with

$\lambda = 633 \text{ nm}$. From equation 7.188 above, $\Delta\nu_{\max} = (k\omega_s a_o) \text{ Hz}$. So:

$$\Delta\nu_{\max} = [2 \times 2\pi(1) \times 1.2 \times 10^{-7} / (633 \times 10^{-9})] = 2.38 \text{ Hz} \quad (7.190)$$

Thus the peak Doppler shift is about three times the 1 Hz displacement frequency.

Hong and Fox (1993) reported on an ‘optical stethoscope’ that used a 670 nm diode laser as a coherent light source with a Michelson [actually, a Twyman-Green] interferometer to sense skin vibrations. Their interferometer was housed in a metal box about $5 \times 3.5 \times 3.5$ in. Its accoupled photodiode output was observed in real time on a digital oscilloscope with FFT spectrum analysis capability. A more refined version of this instrument was described by Hong (1994), in which he used a HeNe laser source and a fiber optic cable to impinge the light on the skin. In most of the phantoms and human subjects studied by Hong, object displacements were periodic at the heart rate, but displacements were in hundreds of μm rather than hundreds of nm! Thus the Twyman-Green interferometer was well out of its $\pm\lambda/4$ linear range and the apparatus evidently served as a mixer to extract the lightwave Doppler shift frequency in the measurement beam. The Doppler frequency shifts extracted followed the shape of the magnitude of the time derivative of the pressure waveform in the phantom studies, as evidenced by the output of a frequency-to-voltage converter (VFC). (Apparently, the ‘skin’ surface displacement of the phantom was proportional to the vessel pressure.) In human studies, Hong cites the peak displacement of the skin over the common carotid artery to be $510 \mu\text{m}$ over 86 ms. This is a velocity of $5.93 \times 10^{-3} \text{ m/s}$ and the peak Doppler lightwave frequency is $\Delta\nu = 2 \times 5.93 \times 10^{-3} / 633 \times 10^{-9} = 18.74 \text{ kHz}$. Obviously, this output frequency will vary over the cardiac cycle. Since there is no way to sense the sign of $\Delta\nu$, the VFC V_o was determined by:

$$V_o \propto \Delta\nu \propto |\dot{L}_2| \propto |\dot{p}| \quad (7.191)$$

Thus, in summary, Hong’s interferometer was apparently used as a Doppler mixer, not a true, nm distance measuring system.

7.3.3.2.3 Two Frequency, Heterodyne Interferometry

Figure 7.52 illustrates one form of a two frequency interferometer. A special laser produces two optical frequencies a few MHz apart (a fractional separation of $10^7/10^{14} = 10^{-7}$). The EM waves at the two frequencies, ν_1 and ν_2 , are linearly polarized at a right angle (orthogonally polarized). Both waves are partially reflected at beam splitter BS1, and sent to polarizer P1 which has its axis at 45° to both waves. The output of P1 is LPL with its E vector directed along P1’s pass axis. Thus, LPL which is the sum of the beams at frequencies ν_1 and ν_2 strikes the detector P1’s photo cathode. P1’s output current, as we have shown above, is proportional to the average of the square of the sum of the E vectors of the laser’s two output EM waves. The detector P1 cannot respond to terms at the lightwave frequencies: ν_1 , ν_2 , $2\nu_1$, $2\nu_2$, $(\nu_2 + \nu_1)$, but only responds to dc intensity terms and radio frequency ac term at frequency $(\nu_2 - \nu_1)$, which is in the MHz range. Hence the lower counter, when gated, counts pulses at the fixed frequency, $(\nu_2 - \nu_1)$. The horizontal, two component beam from BS1 passes through a polarizing beam splitter, PBS2. The LPL beam of frequency ν_1 is directed to the movable retroreflector prism, RR1, where it is returned to PBS2 with a Doppler shifted frequency of

$$v_{1r} = \nu_1 + \Delta\nu_d = \nu_1(1 + 2\delta\dot{x}/c) \text{ Hz} \quad (7.192)$$

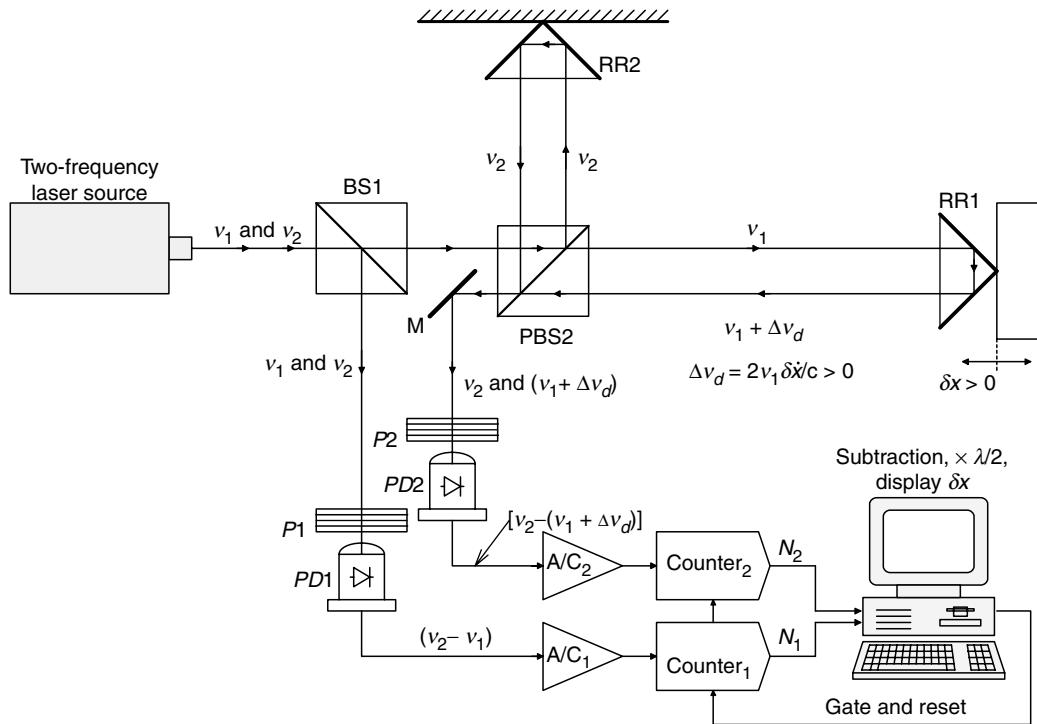


FIGURE 7.52
A heterodyne interferometer.

The orthogonally polarized beam of fixed frequency, v_2 , is sent to the fixed retroreflector, $RR2$, where it is returned to $PBS2$. Both of these orthogonally polarized beams are directed to a full reflecting mirror, M , and thence to polarizer $P2$, which also is at 45° with respect to the E vectors of the two impinging beams. The output light from $P2$ is LPL with its E vector along $P2$'s pass axis. Photodetector $PD2$ also responds to the average of the square of the sum of the two E vector components striking its photocathode. That is, dc light intensity components and a low frequency, ac term at frequency, $[v_2 - (v_1 + \Delta v_d)]$. Only the ac terms at the PD outputs are of interest; the dc is blocked.

The amplifier/comparators (A/C_1 , A/C_2) produce TTL output waves of frequencies $(v_2 - v_1)$ and $[(v_2 - v_1) - \Delta v_d]$, respectively. Assume the counters are reset and gated to count for ΔT seconds. The count $N_1 = \Delta T(v_2 - v_1)$ cycles. Similarly, $N_2 = \Delta T[(v_2 - v_1) - \Delta v_d]$. Assuming the velocity of $RR1$ is constant over ΔT , the computer subtracts N_2 from N_1 and finds $\Delta N = \Delta T\Delta v_d = \Delta T(2v_1\delta\dot{x}/c)$ cycles. Thus, the velocity, $\delta\dot{x}$, can be written as:

$$\delta\dot{x} = \frac{\Delta N c}{\Delta T 2v_1} = \frac{\Delta N}{\Delta T} (\lambda_1/2) \quad (7.193)$$

If ΔN is calculated periodically with period $T_c > \Delta T$, we can generate a series of velocity samples, which can be integrated to find X_o , the total $RR1$ displacement. Thus with the use of DSP, this interferometer can measure both target velocity and the distance traveled by a cooperative target.

Essential to the operation of this interferometer is the generation of two coherent rays of frequencies v_1 and v_2 such that $\Delta\nu = (v_2 - v_1)$ is of the order of tens of kHz to tens of

MHz. Phase locked, coherent frequency shifting of the LPL rays can be done in several ways [Sirohi and Kothiyal, 1991]. Quoting these authors:

"One technique makes use of rotating polarization components, namely quarter- and half-wave plates and (a) polarizer. For example, if circularly polarized light is transmitted through a rotating half-wave plate, the orthogonal components of the transmitted light show a frequency difference equal to twice the rotation rate of the plate. The principal difference in this method is that for reasonable rotation rates, it is not possible to obtain a frequency shift ($\Delta\nu$) much larger than a few kHz.

A moving diffraction grating can also be used as a frequency shifter, since it shifts the N th diffracted order by an amount Nvf , where v is the velocity component of the grating perpendicular to the grating lines (m/sec) and f is the spatial frequency of the grating (lines/meter). In other words, the frequency shift is equal to N times the number of grating lines that pass a given point per second. For a (moving) grating, the frequency shift is independent of the wavelength. To obtain a large frequency shift in the range of several kHz, a circular grating is useful.

An acoustooptic Bragg cell shifts the frequency of diffracted light in the same way as a moving diffraction grating. In this case the traveling acoustic wave serves as the moving grating. The frequency shift of the first diffracted order is equal to the frequency used to drive the Bragg cell independent of the wavelength of (the input) light. A Bragg cell gives a frequency shift in the range of several MHz.

'Some lasers can be designed to give two frequency outputs, for example, the Zeeman split HeNe laser.'

Indeed, the frequency shift in a Bragg cell can be written as:

$$\Delta\nu = \frac{2v_s}{\lambda_s} = f_s \text{ Hz} \quad (7.194)$$

where v_s is the velocity of sound in the Bragg medium and λ_s is the wavelength of the sound waves in the Bragg medium. Electro-optic modulation and diffraction of coherent light by Bragg cells are described in great detail in Chapter 12 of Yarov (1991).

7.3.3.2.4 The Fabry-Perot Interferometer

Figure 7.50 illustrates one form of the Fabry-Perot interferometer. This system has several unique features. It uses a diffuse, distributed parameter light source. Such a source can be made by expanding a laser beam with a telescope to about a 2 cm diameter disk directed on the back of a ground glass sheet. The ground glass then acts as a secondary source of numerous rays at many angles. In the figure, we trace one ray from point S_1 on the source, through a convex lens, thence to the etalon. The light ray reflects back and forth many times on the inside of the half-silvered, parallel plates of the etalon. A small fraction of the internally reflected rays are refracted and pass out of the etalon's plates. The emergent rays from the right surface of the etalon are collimated to a focus at point P_1 . Using a complex optical superposition, it is possible to derive an expression for the intensity of the concentric circular interference rings of the Fabry-Perot interferometer on its viewing screen [Hecht, Section 9.6.1, 1987; Vaughan and Vaughan, 1989; Betzler, 2002]. However, we will not repeat this derivation here.

Two things are unique about the Fabry-Perot refraction rings.

1. Their separation decreases with their radii
2. The rings are very sharp, unlike the diffraction rings of a Michelson interferometer, in which the intensity varies roughly sinusoidally with ring radius.

The etalon refraction cavity supports many internal reflections from each ray directed to it from the lens. The parallel internal faces of the etalon's glass plates are vapor coated with a thin, partially reflecting film of metal to produce intense internal reflections. When a light ray inside the etalon encounters the metal film, a large fraction of its intensity, R , is reflected internally, a very small fraction, A , is absorbed, and a small fraction, T , is transmitted outside the etalon cavity. Since $A \ll T$, we can write.

$$1 = A + R + T \cong A + T \quad (7.195)$$

Betzler 2002 defines a quality factor parameter, Q_R :

$$Q_R = \frac{4R}{(1-R)^2} \quad (7.196)$$

He also shows, using superposition of reflected waves, that the ratio of output to input intensity is:

$$\frac{I_t}{I_i} = \frac{1}{1 + Q_R \sin^2(\delta/2)} = \frac{1}{1 + \frac{4R}{(1-R)^2} \sin^2(\delta/2)} \quad (7.197)$$

Hecht shows that each narrow, bright interference ring has half its peak intensity for a phase difference, $\delta = \delta_{pk} \pm \delta_{1/2}$. Thus the half width of a peak is $\gamma = 2\delta_{1/2} \cong 4/\sqrt{F}$ radians. The larger the R is, the sharper the peaks will be, and the lower the absolute output intensity (more light is trapped inside the etalon). Hecht defines a parameter describing the sharpness of the interference rings, the finesse, F . The finesse is defined as the ratio of the peak separation to the half width. That is:

$$F = (\pi/2)\sqrt{F} \quad (7.198)$$

The finesse can typically range about 20–200 in a basic F-P interferometer.

The usual mode of operation of a F-P interferometer is to observe fringe movements as the etalon plate separation is varied by Δd by moving, say, the left hand plate. By detecting fringe movements optically, it is possible to drive a piezo motion generator connected to the right hand plate to renull the fringes. Since the piezo displacement to rennull the fringe pattern is $\Delta d' = V_c K_p = \Delta d$, we measure V_c at null to find Δd .

F-P interferometers have been used to measure a diversity of measurands, including atmospheric CO₂, magnetic fields, small displacements (nm), large displacements (mm), high pressure transients and underwater sound.

Interferometers are generally suited for precise direct measurement of surface displacements when the range of deflection is $< \lambda/4$. When the surface moves at a high frequency, the interferometer output contains many harmonics due to the $\cos[\xi \sin(\omega_s t)]$ terms, and lightwave Doppler shift terms. Most interferometers are better suited to measuring small, slow displacements, such as tooth drift in orthodontic research, or plant growth. Heterodyne interferometers using two light wave frequencies are suitable for measuring object velocities and displacements.

When a single wavelength interferometer is used to measure a distance which is a large multiple of $\lambda/4$, an additional, quadrature detector is used. The sign of Δx is found by examining the phase of the fringes in the direct and quadrature detector outputs as the object moves in time. The total number of fringes gives the distance to the nearest

λ and interpolation of the residual distance can be done on $\Delta V_o = (V_o - V'_o)$ from equations 7.163 and 7.164.

Fabry-Perot interferometers offer the advantage generating sharp diffraction rings, hence it is easier to detect object motion caused fringe shift.

Interferometer precision can range from 1 : 1,000 to over 1 : 10^5 , depending on the design and signal processing used. Accuracy is on the order of $\lambda/2000$. Sources of noise in interferometry include mechanical vibrations, changes in air refractive index due to temperature, barometric pressure and humidity changes, electronic noise in photo sensors, and drift in source laser amplitude, frequency and polarization.

7.3.3.3 Phase Lock Velocity and Range Sensing with Ultrasound and EM Radiation

In this section, we will examine a class of closed loop measurement system that gives simultaneous analog output voltages proportional to target range and velocity along the range vector. Recall that conventional radar works by measuring the time it takes a transmitted pulse to return to the receiver after being reflected from the target. Doppler radar measures the frequency shift of the reflected pulse and computes the target velocity from this information. Sonar works in essentially the same manner. The LAVERA phase lock method adjusts the transmitted frequency under closed loop control in order to keep the phase lag constant between received and transmitted waves.

Several years ago, the author developed a means of measuring target range and velocity without explicit consideration of the Doppler effect [Northrop, 1997]. This system was developed for use with an amplitude modulated laser beam. Many conventional laser based velocimeters are based on a time-of-flight rangefinder principle in which laser pulses are reflected from a reflective (cooperative) target back to the source position and the range is found from the simple relation:

$$L = c \Delta t / 2 \text{ m} \quad (7.199)$$

If the target is moving towards, or away, from the laser/photosensor assembly, the relative velocity can be found by approximating $v = \dot{L}$, by numerically differentiating the sequence of return times for sequential output pulses, $Dt = \sum \Delta t_k$. In 1992, Laser Atlanta made a vehicular velocimeter which used this principle.

The laser velocimeter and rangefinder (LAVERA) system we will describe in this section uses quite a different principle to provide simultaneous velocity and range output. The LAVERA system was devised by the author [Northrop, 1997] and reduced to practice by graduate student Todd Nelson (1999) for his MS research in Biomedical Engineering at the University of Connecticut. The motivation for developing the system was to make a walking aid for the blind that could be incorporated into an active cane.

The system uses sinusoidally amplitude modulated, CW light from an NIR laser diode (LAD). The beam from the LAD is reflected from the target object, travels back to the receiver photosensor where its intensity is converted to an ac signal whose phase lags that of the transmitted light. The phase lag is easily seen to be:

$$\Delta\phi = (2\pi)(\Delta t/T) = (2\pi)(2L/c)/T \text{ radians} \quad (7.200)$$

where as discussed above, Δt is the time it takes a photon to make a round trip to the target, T is the period of the modulation and c is the speed of light. The system we developed is a closed loop system that tries to keep $\Delta\phi$ constant regardless of the target range, L . This means that as L decreases, T must decrease, and conversely, as L increases, T must increase to maintain a constant phase difference. Figure 7.53 illustrates

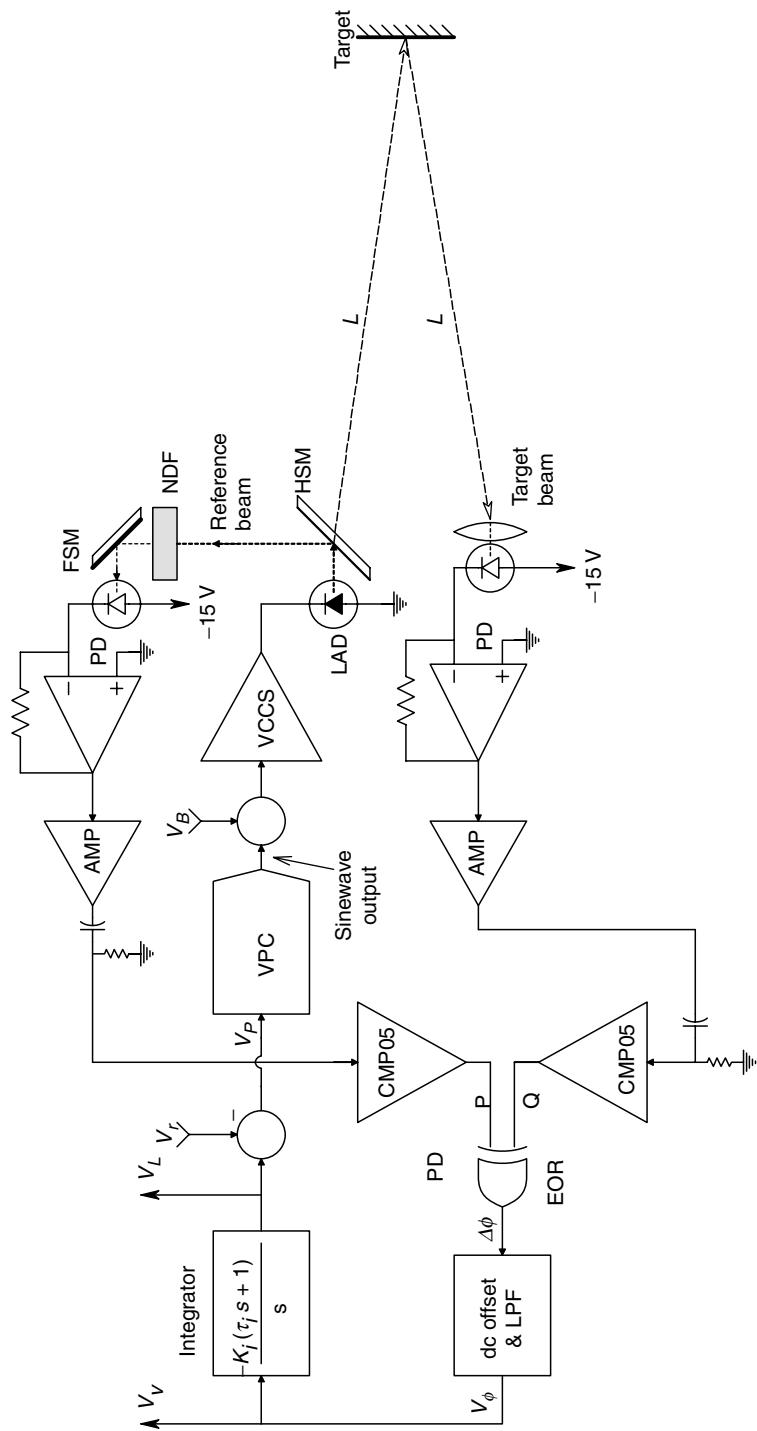


FIGURE 7.53
Block diagram of the laser velocity and rangefinder (LAVERA) system devised by the author.

the functional architecture of the LAVERA system. Note that the dc bias voltage, V_B , is required to set a mean intensity of the diode laser beam (light intensity cannot go negative) which is sinusoidally amplitude modulated. The transmitted intensity is thus:

$$I_t = I_o + I_m \sin(2\pi t/T), \quad I_o > I_m \quad (7.201)$$

R-C high-pass filters remove the I_o component of the transmitted and received light signals before they are input to the comparators.

We start the analysis by considering the voltage-to-period converter (VPC). The VPC is a VCO in which the output period is proportional to V_P . Thus:

$$T = 1/f = K_{VPC} V_P + b \text{ s} \quad (7.202)$$

By way of contrast, the output from a conventional voltage-to-frequency oscillator is described by:

$$f = K_{VCO} V_C + d \text{ Hz} \quad (7.203)$$

Next, consider the phase detector. This subsystem measures the phase difference between the transmitted, modulated signal intensity and the delayed, reflected, modulated signal intensity. Although an analog multiplier can be used directly to measure this phase difference, in this system, the analog sine waves proportional to light intensity are passed through comparators to convert them to TTL waves, whose phase difference is sensed by an exclusive or (EOR) gate phase detector (PD). The EOR PD is a quadrature detector (i.e. its output duty cycle is 50% when one input lags (or leads) the other by odd multiples of $\pi/2$). Figure 7.54 illustrates the EOR PD's circuit and transfer function. The op-amp inverts the TTL output of the EOR gate and subtracts a dc level so that V_z swings $\pm 1.7 \text{ V}$.

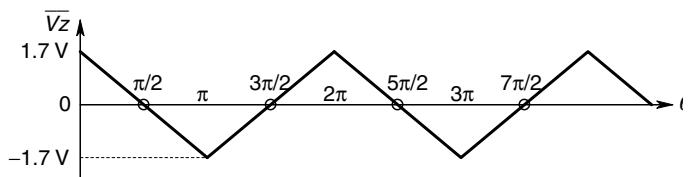
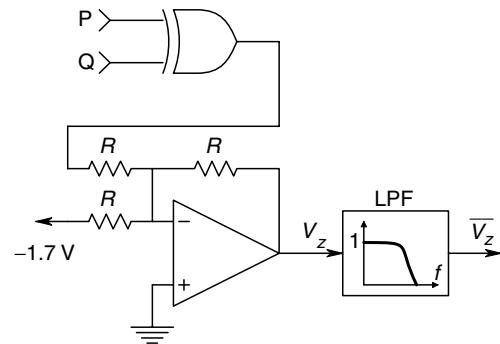


FIGURE 7.54

An exclusive OR gate used as a quadrature phase detector. Note that zero dc output occurs for odd multiples of $\pi/2$ radians phaseshift between the P and Q digital inputs.

For a heuristic view of how the system works, we can write the average PD output as:

$$V_\phi = \overline{V_z} = -(3.4/\pi)\phi + 1.7 \text{ V} \quad (7.204)$$

In the steady state with fixed L , $V_v=0$ in a type 1 feedback loop. For $V_v=0$, $\phi_{ss} = \pi/2$. But we have already established that the phase lag for a returning, modulated light beam is:

$$\phi = (2\pi)(2L/c)(1/T) \text{ radians} \quad (7.205)$$

If equation 7.205 for ϕ is solved for T which is substituted into the VPC equation, we have:

$$T_{ss} = (2\pi)(2L/c)(1/\phi_{ss}) = K_{VPC} V_{PSS} + b \quad (7.206)$$

↓

$$8L/c = K_{VPC}[V_{LSS} - V_r] + b \quad (7.207)$$

If we let $V_r \equiv b/K_{VPC}$, the final expression for V_{LSS} is obtained:

$$V_{LSS} = 8L/(cK_{VPC}) \quad (7.208)$$

Thus, the steady state V_L is indeed proportional to L . Going back to the VPC equation, we find that for a stationary target, the steady state VPC output frequency is given by:

$$f_{ss} = 1/T_{ss} = c/(8L) \text{ Hz} \quad (7.209)$$

At $L=1\text{m}$, $f_{ss}=37.5\text{ MHz}$; when $L=30\text{ m}$, $f_{ss}=1.25\text{ MHz}$, etc.

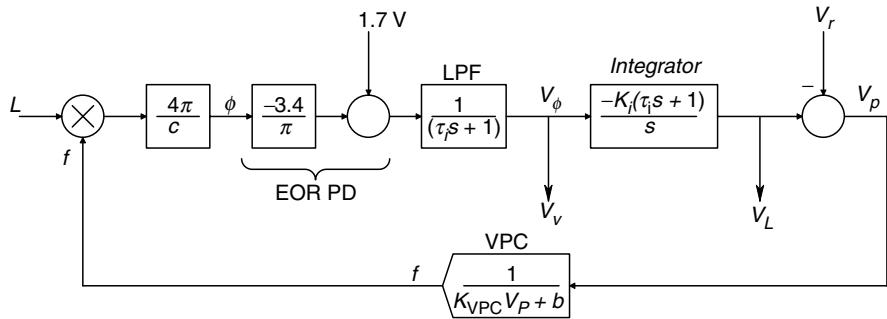
Now, suppose the target is moving away from the transmitter/receiver at a constant velocity, $v = \dot{L}$. V_v is no longer zero, but can be found from the integrator relation:

$$V_v = -\dot{V}_L/K_i = \frac{-8\dot{L}}{K_i K_{VPC}} \quad (7.210)$$

Notice the sign of V_v is negative for a receding target, also as the target recedes, the VPC output frequency drops. Clearly, practical considerations set a maximum and a minimum for L . The limit for near L is the upper modulation frequency possible in the CW system. The limit for far L is the diminishing intensity of the return signal, and noise.

To appreciate the closed loop dynamics of the LAVERA system, it is possible to make a systems block diagram as shown in Figure 7.55. Note that this system is nonlinear in its closed loop dynamics. It can be shown [Nelson, 1999] that the closed loop system's undamped natural frequency, ω_n , and damping factor are range dependent, even though the steady state range and velocity sensitivities are independent of L . The undamped natural frequency for the LAVERA system of Figure 7.55 can be shown to be:

$$\omega_n = \sqrt{\frac{4(3.4)c K_i K_{VPC}}{64\tau_l L_o}} \text{ r/s} \quad (7.211)$$

**FIGURE 7.55**

Systems block diagram of the LAVERA system, relevant to tracking a moving target.

The closed loop LAVERA system's damping factor can be shown to be:

$$\xi = \frac{1}{2} \left[\frac{64 L_o}{4(3.4)c K_i K_{VPC}} + \tau_i \right] \sqrt{\frac{4(4.3)c K_i K_{VPC}}{64 \tau_i L_o}} \quad (7.212)$$

The LAVERA system architecture was also extended to ultrasound and microwave distance measurement systems [Northrop, 1997; Nelson, 1999]. At the heart of all LAVERA systems is the use of a voltage-to-period converter (VPC) instead of a voltage-to-frequency converter (VFC). The VPC is a voltage controlled oscillator in which the period of the output waveform is directly proportional to the input (control) voltage, that is,

$$T = 1/f = b + K_p V_c \text{ s} \quad (7.213)$$

(In a VFC, $f = 1/T = a + K_v V_c$ Hz). Thus the output frequency of a VPC is given by:

$$\omega_0 = 2\pi/[b + K_p V_c] \text{ r/s} \quad (7.214)$$

For small changes in V_c , we can write:

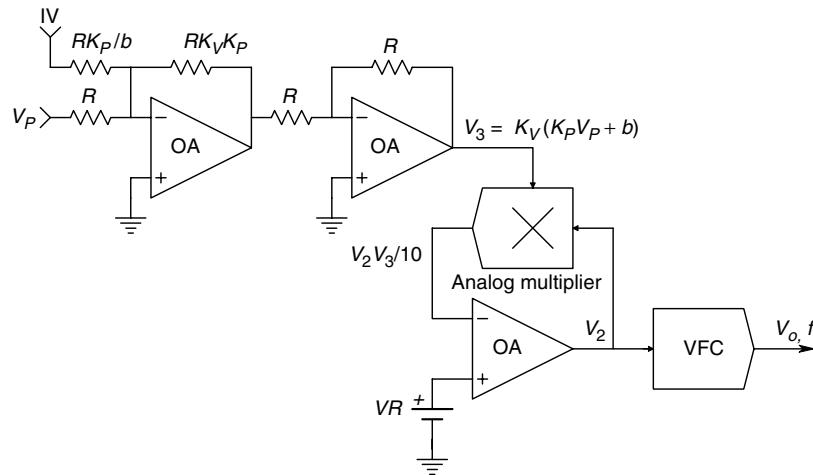
$$\Delta\omega = \left(\frac{\partial\omega_0}{\partial V_c} \right) \Delta V_c = \frac{-2\pi}{(b + K_p V_c)^2} \Delta V_c \quad (7.215)$$

Note that there is no distance dependence in the velocity and range sensitivity expressions; they remain constant over the system's operating range.

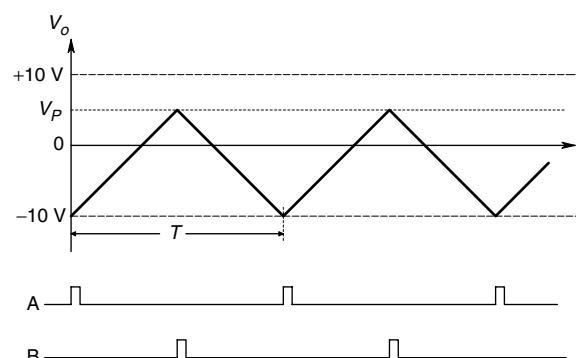
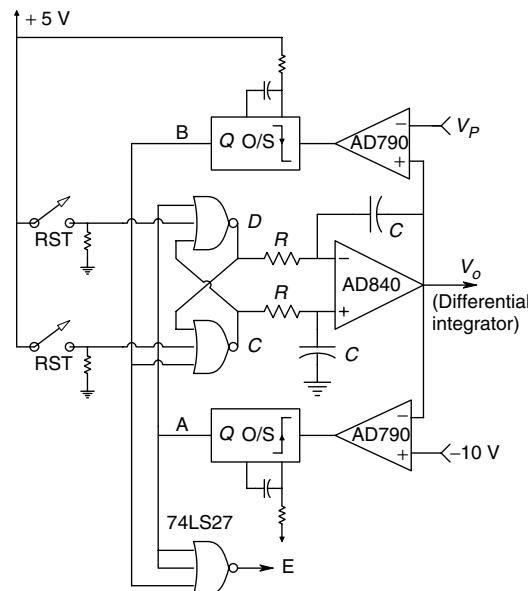
Examples of an analog and a digital VPC are shown in Figures 7.56 and 7.57, respectively. Note that the AVPC uses a standard VFC but takes the reciprocal of the input analog control voltage. Note that $f = K_F V_2$ for the VCO. In the nonlinear op-amp circuit, $V_R = V_2 V_3 / 10$. Also, from Figure 7.56, $V_3 = K_V (K_P V_1 + b)$, so finally:

$$f = \frac{K_F 10 V_R / K_V}{K_P V_p + b} \text{ Hz} \quad (7.216)$$

As in the case of VFCs, there are practical limits to the range of the output frequency in VPCs.

**FIGURE 7.56**

An analog voltage-to-period converter (VPC) oscillator. The voltage-to-frequency converter oscillator (VFC) can have an analog sine wave output.

**FIGURE 7.57**

A digital VFC system.

The DVPC of Figure 7.57 is really a hybrid circuit in which the outputs of the R-S flip-flop are integrated by the op-amp. Assume C is HI and D LO— V_o ramps positive until it reaches the input signal, V_p . The upper AD790 comparator then goes LO, and the upper one-shot generates a narrow positive output pulse that causes the NOR-gate R-S FF to change states. Now C is LO and D HI—the integrator output now ramps negative until it reaches -10 V , the threshold for the lower comparator to go HI. When the lower comparator goes HI, it triggers the lower one-shot to give a narrow positive output pulse that again causes the FF to change state, and the process repeats itself. One can see intuitively that the more positive that V_p is, the longer will be the E output pulse period.

Let us now derive the transfer function of this DVPC, $T=f(V_p)$. First, we need the transfer function of the differential integrator. Using superposition, and assuming an ideal op-amp:

$$V_o(s) = V_D(s) \left[\frac{-1}{sRC} \right] + V_C(s) \left(\frac{1}{[(s+1/RC)RC]} \right) \left[1 + \frac{1/sC}{R} \right] = [V_C(s) - V_D(s)] \left[\frac{1}{sRC} \right] \quad (7.217)$$

Note that the R-S FF's outputs $V_D(s)$ and $V_C(s)$ can be treated as steps of voltage occurring at $t=0$ and thus, can be written as $V_C(s)=3.4/\text{s}$ and $V_D(s)=0.25/\text{s}$ or $V_C(s)=0.25/\text{s}$ and $V_D(s)=3.4/\text{s}$. Thus, with appropriate attention to initial conditions, and assuming that $V_p(t)$ remains constant over the VPC period, it is easy to show that the integrator sawtooth output has slope of $m=\pm RC/3.15$. So, starting at $V_o(t)=-10\text{ V}$, it is found that it takes one-half sawtooth period for $V_o(t)$ to reach V_p . $T/2$ is easily found to be:

$$T/2 = (V_p + 10)RC/3.15\text{ s} \quad (7.218)$$

Note that it also takes the same $T/2$ for $V_o(t)$ to ramp down to -10 V again.

Since the bottom 3-input NOR gate gives an output pulse every time the RSFF changes state, the pulse output of the VPC, E , has period $(V_p + 10)RC/3.15\text{ s}$. In terms of frequency, the lowest output frequency of E will be:

$$f_{E\min} = 3.15/(20RC) \text{ pps} \quad (7.219)$$

If the minimum V_p allowed is -9.8 V , the maximum DVPC output frequency will be:

$$f_{E\max} = 3.15/(0.2RC) \text{ pps} \quad (7.220)$$

For example, let $C=1\text{ nF}$, $R=1\text{ k}$, then $f_{E\max}=15.75\text{ Mpps}$, and $f_{E\min}=157.5\text{ kpps}$ occurs when $V_p=+10\text{ V}$.

The LAVERA system is unique in that it provides simultaneous range and velocity output voltages for a moving target. If a second integrator is added in the loop, the closed loop system becomes Type II, giving zero steady state error in V_V when the target is moving at constant velocity (not accelerating). The velocity output is independent of range. This independence disappears if a VFC is used in the system instead of a VPC. The general LAVERA systems architecture can be extended to pulsed or CW ultrasound, microwaves and pulsed lightwaves. Sinusoidally amplitude modulated, CW lightwaves were used in the system described above because lower bandwidth amplifiers than required can be used for pulsed systems.

7.4 Measurement of Force and Torque

Measurement of force is important in many branches of science, engineering and biomedicine. For example, the magnetic attractive or repulsive force between current carrying conductors is used in the definition of ampere as the unit of electric current. The measurement of forces in a uniform gravitational field is how we weigh things. The thrust of a jet engine is measured as a force. In biomechanics, force plates are used to study the forces involved in various gaits (walking, etc.). Physiologists and physical therapists are concerned with the contractile force produced by muscles under various physiological and pharmacological conditions. Aerospace engineers and naval architects are concerned with the lift and drag forces on airframes, wings, hulls, etc. A locomotive or tractor trailer exerts force on a load to accelerate it to speed and then keep it rolling. Traction force times velocity is the power expended in moving the load (horsepower). The range of forces which are commonly measured is from micrograms to hundreds of thousands of kilograms. Obviously, no single sensor can cover this dynamic range, so there are many types and sizes of force sensors (load cells) available to the instrumentation engineer, depending on the application.

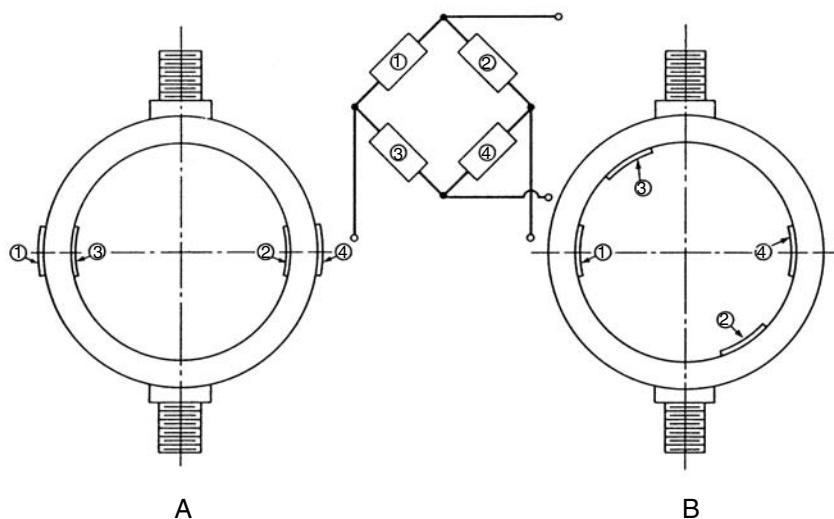
Torque measurements are generally made on rotating shafts connecting a power source (electric motor, Diesel engine, etc.) to a device doing work (pump, rolling mill, conveyor belt, etc.). Torque measurements are used to calculate rotary output power ($\text{torque} \times \text{angular velocity}$). In general, direct torque measurements are more difficult to make than are force measurements, because the shaft is rotating and slip rings must be used to couple the output signal to the stationary, outside world. There are means, as we will see below, of measuring torque, optically or magnetically, which do not require direct contact with the rotating shaft.

7.4.1 Load Cells for Force Measurement

Paradoxically, most force measurements involve the small displacement of a stressed, linearly elastic member. Displacement or strain is proportional to the stress produced by the force load. The transduction of the small displacement may be done optically by interferometry, or electrically by a sensitive LVDT, a bonded or unbonded wire strain gauge bridge, a semiconductor strain gauge bridge, a circuit which senses small changes in capacitance, or by a piezoelectric crystal (crystal force transducers only respond to time varying forces). Any force applied to a piston acting on a fluid will produce a rise in pressure, which, if there are no leaks, will remain constant and can be measured with a pressure transducer. The most basic type of load cell is one which uses four, bonded strain gauges in a Wheatstone bridge configuration, illustrated in Figure 7.58. The four-arm bridge is inherently temperature compensated and the circular load cell configuration works for compressive as well as tension loads.

Interface, Inc., of Scottsdale, AZ offers a large line of bonded strain gauge bridge load cells with ranges 22.2–8.89E5 Newtons full scale in tension or compression. The Interface load cells are quite precise and linear. Non-linearity is $\pm 0.03\%$ of rated output, hysteresis is $\pm 0.02\%$ of rated output, and non-repeatability is $\pm 0.01\%$ of rated output.

Schaevitz Engineering, Pennsauken, NJ markets LVDT based load cells with ranges from 9.80E-2 Newtons full scale, to as much as 2.22E6 Newtons full scale. Frequency response of an LVDT based load cell is set by the carrier frequency used with the LVDT, which can be as high as 10 kHz. The Schaevitz load cells have linearities of 0.2% of full range, resolutions of 0.1% of full range and repeatability of 0.1% of full range. Deflection at full scale force may be 0.005 in if the Schaevitz 004-XS-B sub-miniature LVDT is used in the cell.

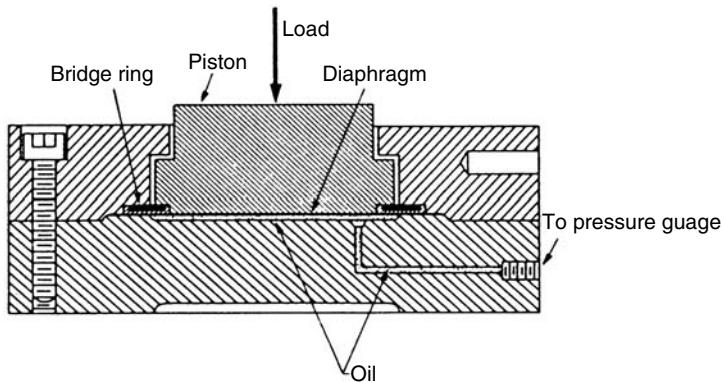
**FIGURE 7.58**

Two, ring type, load cells used to measure tensile or compressive forces. A. The strain gauges respond only to the bending strains; the axial strain components cancel out in the bridge. B. The Wheatstone bridge has greater sensitivity because gauges 1 and 4 respond to both axial and bending strain components. (Source: With permission, Beckwith and Buck, 1961)

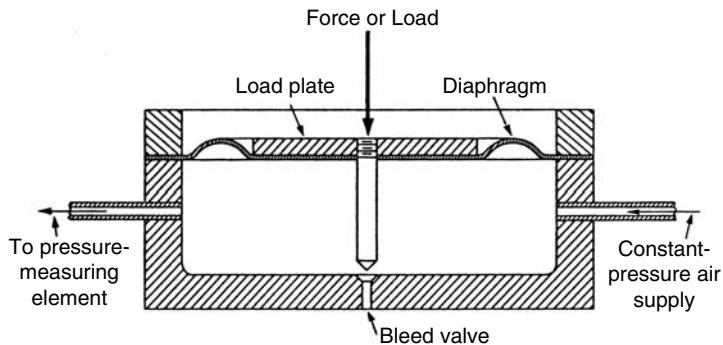
Still another type of load cell is the capacitive cell marketed by Kavlico Corp., Chatsworth, CA. In this cell, the applied force causes the deflection (5×10^{-4} in at full scale force) of a ceramic diaphragm, one surface of which is metallized, which in turn causes a proportional capacitance change which is sensed electronically. Kavlico load cells have full scale ranges of 0.98 N to 2.67E2 N. System accuracy is $\pm 5\%$ of full scale, and repeatability and hysteresis are typically $\pm 0.36\%$.

Grass Instruments, Quincy, MA make a physiological force sensor which is basically a cantilever to which are bonded four strain gauges that make a temperature compensated Wheatstone bridge. The Grass force sensors have sensitivities which can be decreased by the addition of insertable spring sets. The Grass FT-03C has a working range of ± 50 g with no insertable springs, but can be given a full scale range of 2 kg with the stiffest spring set. Deflection at full rated load is 1 mm. The mechanical natural resonant frequency is 85 Hz for the 50 g range and 500 Hz for the 2 kg range. The strain gauge bridge output is 1.5 mV/mm/V applied to the bridge. Obviously, differential amplification is needed.

Accurate, linear load cells for extreme compressive forces can be made using hydraulic technology. Figure 7.59 illustrates such a hydraulic cell. Beckwith and Buck (1961) stated that capacities as high as 5×10^6 lbs with accuracies of $\pm 0.1\%$ of full scale have been attained with this design. Pneumatic load cells offer the advantage over hydraulic fluid designs over a lower force range in terms of temperature insensitivity. Air is compressible and does not freeze or boil over a very wide temperature range. Pneumatic load cells are generally made using a compliant diaphragm, rather than a piston. In addition, they use a feedback design in which the internal air pressure is adjusted by a needle valve attached to the moving diaphragm such that, if the external, downward load force exceeds the upward pneumatic force on the diaphragm, the diaphragm drops downward and decreases the outflow of air by decreasing the pneumatic bleed conductance. This causes the internal air pressure to rise, raising the diaphragm and increasing the outflow conductance. An equilibrium is reached, in which the internal pressure times

**FIGURE 7.59**

Section through a hydraulic load cell. This system responds only to compressive forces. (Source: With permission, Beckwith and Buck, 1961)

**FIGURE 7.60**

Section through a pneumatic load cell. (Source: With permission, Beckwith and Buck, 1961)

the effective diaphragm area equals the applied load. Also, the volume inflow must equal the volume outflow at equilibrium (Figure 7.60). Assume the needle valve has a pneumatic conductance which is a function of diaphragm displacement given by:

$$G_o(x) = G_{oMAX} \quad x < 0 \quad (7.221a)$$

$$G_o(x) = G_{oMAX}(1 - x/x_{MAX}) \quad 0 \leq x \leq x_{MAX} \quad (7.221b)$$

$$G_o(x) = 0 \quad x \geq x_{MAX} \quad (7.221c)$$

Here, negative x is the upward displacement of the membrane from $x=0$. It is not difficult to show, that in the steady state,

$$P_i = F/A = P_o G_i/(G_i + G_o) \quad (7.222)$$

where G_i is the fixed, pneumatic input conductance, A is the effective area of the diaphragm, F is the external load force and P_o is the pressure in the air reservoir supply. From equation 7.222, we see that the practical bounds on the measurement of F are set by $F_{MAX}=P_oA$ at the upper end and $F_{MIN}=P_oAG_i/(G_i - G_{oMAX})$ at the low end.

One problem with simple pneumatic load cells is that they may become unstable under certain conditions. If the force on the pneumatic load cell is a mass, the air in the chamber behaves like a nonlinear spring, and oscillations and instability in the operation of this sensor can occur unless pneumatic damping is used in its design. Beckwith and Buck (1961) report that 40 ton capacity pneumatic load cells are commercially available.

Closed loop, electromagnetic force cells are used for weighing light objects in the range of 0.1 mg–100 g. In this design, an upward force is generated electromagnetically to equal the external, downward force, as shown in Figure 7.61A. Equilibrium position can be sensed either electro-optically or with a high resolution LVDT. The electromagnetic force is generated by a coil and permanent magnet assembly similar to that used in a loudspeaker, except that there is no spring required. The coil current is made proportional to the integral of the downward displacement. A systems block diagram describing this system is shown in Figure 7.61B. There is zero steady state displacement error and it can be shown that:

$$\frac{V_c}{Mg} = \frac{1}{G_M K_F} \text{ V/N} \quad (7.223)$$

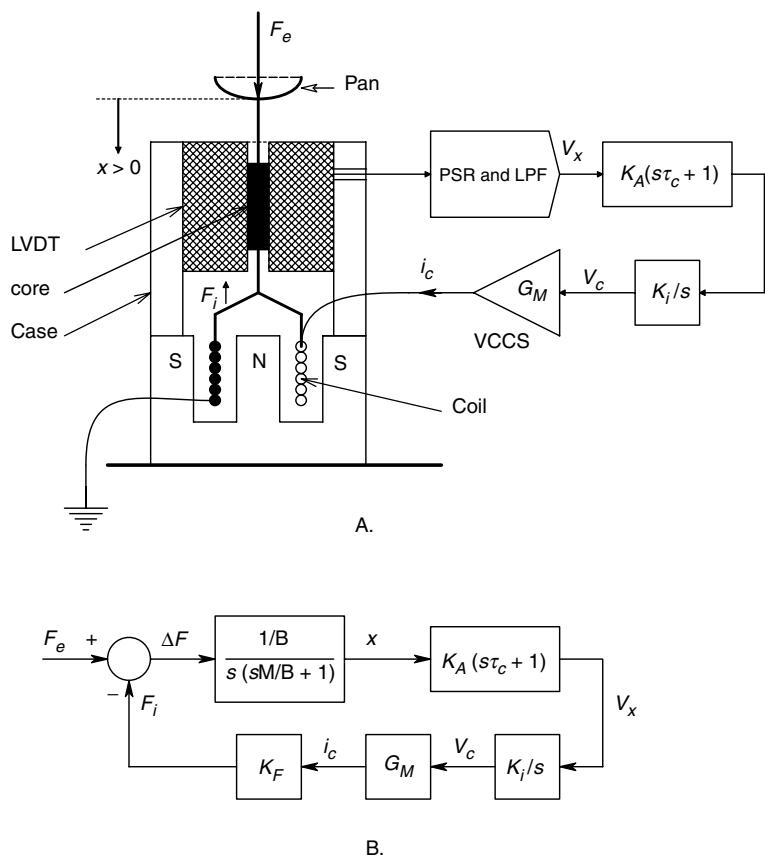


FIGURE 7.61

A. Section through an electrodynamic, feedback load cell. Position is sensed by an LVDT. The LVDT output is conditioned by a phase sensitive rectifier and lowpass filter (PSR and LPF), then by a proportional plus derivative compensator and integrator. The integrator's output, V_C , is then conditioned by a transconductance amplifier (VCCS) whose output, i_C , is the input to the force coil. B. Systems block diagram of the electrodynamic, feedback load cell.

Other obvious and far less technical means for measuring forces less than several hundred kg include spring scales, which make use of Hooke's law and a calibrated scale, where displacement of the moving end of the spring is proportional to the applied force. Steelyards, laboratory beam balances, platform balances, etc., all make use of calibrated weights and lever arms pivoted on knife edge bearings. A human operator adjusts the position of one or more known weights on the long arm of the balance until the arm is level and stationary, signifying that the unknown force times its lever arm equals the sum of the known forces times their lever arms.

7.4.2 Torque Measurements

Torque measurements on rotating shafts are made with more difficulty than are those on stationary shafts. The reason for this difficulty is that the wires from sensor elements, such as a strain gauge bridge cemented to the shaft, must be brought out to the non-rotating world through slip rings. Noise introduced by the brushes making contact with the rotating slip rings can reduce the resolution of a rotating torque measuring system below that of an identical system used on a stationary shaft without slip rings. Figure 7.62 illustrates the configuration of bonded strain gauges on a shaft in order to measure torsional strain caused by a torque couple applied to the ends of the shaft. Such a configuration may be shown to be insensitive to temperature changes, axial load and bending of the shaft.

There are other means of measuring torque on a rotating shaft that do not require electrical contact with the shaft. In one approach, alternating black and white markings are painted around the circumference of each end of the torque carrying shaft under measurement. A photosensor consisting of an IR LED and a phototransistor is placed above both rotating, black/white bands. The output of each photo sensor is a square wave with frequency proportional to the shaft's angular velocity. However, the phase between

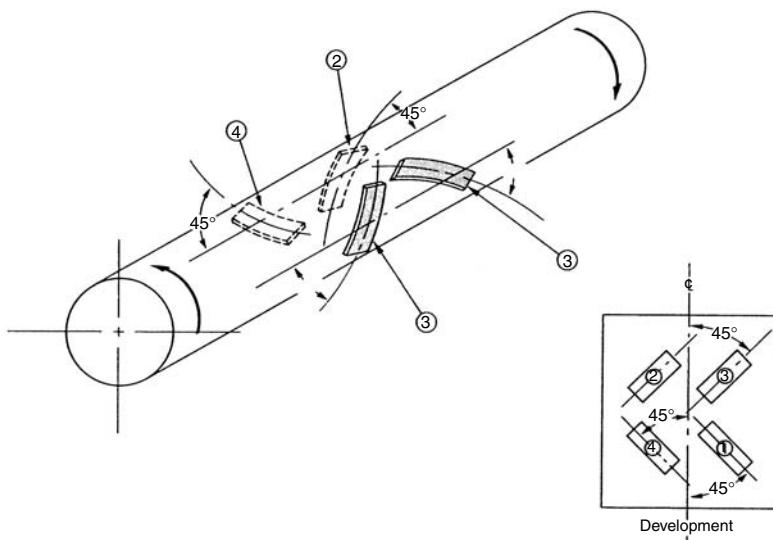


FIGURE 7.62

Four bonded strain gauges cemented to a rotating torque shaft. This configuration is insensitive to temperature, axial load, or bending. Slip rings must be used to make electrical connections to the gauges (not shown). Slip rings are noisy. (Source: With permission, Beckwith and Buck, 1961)

the square waves is determined by the displacement angle produced by the torque between the black/white bands. This angle is given by:

$$\phi = KT d \quad (7.224)$$

where d is the distance between the bands, T is the torque on the shaft and K is a constant determined by the material and dimensions of the shaft. The electrical phase angle between the two square wave trains, θ , can be measured and is related to ϕ by:

$$\theta = N\phi \quad (7.225)$$

where N is the integer number of black/white cycles inscribed around the circumference of the torque shaft at each end. The frequency of each square wave is just:

$$f = N \times RPM/60 \quad (7.226)$$

RPM is the shaft's revolutions per minute. One way of measuring the phase angle, θ , is to use a phase detector IC, such as the Motorola MC4044. The MC4044 has a working range of $\pm 360^\circ$ and which can generally be used to resolve angles as small as $\pm 0.1^\circ$. Obviously, to measure low values of torque, a phase detector is needed, which can resolve to one-hundredths of a degree or better and work over a wide range of frequencies.

A frequency independent, millidegree resolution phase detector design was devised by Du (1993), based on a design by Nemat (1990). The Du system is illustrated schematically in Figure 7.63. The periodic reference signal (A) is converted to a TTL waveform by a fast comparator, the output of which is the input to a phase lock loop in which the VCO output is made to follow the input frequency and is multiplied it by a factor of 360,000. Thus, every input cycle produces 360,000 PLL clock cycles and the input phase is effectively divided into millidegrees (i.e. there is one PLL clock pulse for every 0.001° of input, regardless of the exact input frequency). The reference and complimentary, phase shifted square wave inputs act through a NAND gate ($G1$) to enable a 24-bit, high frequency binary counter, which counts the high frequency output of the PLL for N cycles of the input wave. For purposes of discussion, assume that the input signals are 100 Hz. Thus, the PLL's high frequency output will be 36 MHz. The number of high frequency pulses counted depends, of course, on the phase shift between the 100 Hz input square waves, as well as N .

If the phase shift is zero, no pulses are counted. On the other hand, if, for example, the phase shift is 3.6° , then 3600 PLL clock cycles will be counted for every one of the N gated input cycles. After N input waveform cycles, the cumulative count in the binary counter will be 9.216×10^5 , if $N=256$ and $\theta=3.6^\circ$. In order to obtain an average count for N input cycles, the cumulative count in the counter is loaded in parallel to a 24-bit shift register (three 74F299 ICs). Division by N is easily accomplished by shifting the register contents to the right by k clock cycles from the controlling computer. We make $N=2k$ in this system, with $k=0, 1, 2, \dots, 8$. The 16 least significant bits of the shifted binary number in the register is then loaded into the computer and displayed as a decimal angle in the format, 88.888° . For our 3.6° phase shift example, this would be 3600 or 3.600° . The PLL allows precise measurement of the phase difference between the input signals A and B, even though their frequency may not be fixed. In fact, it will vary as the shaft speed varies.

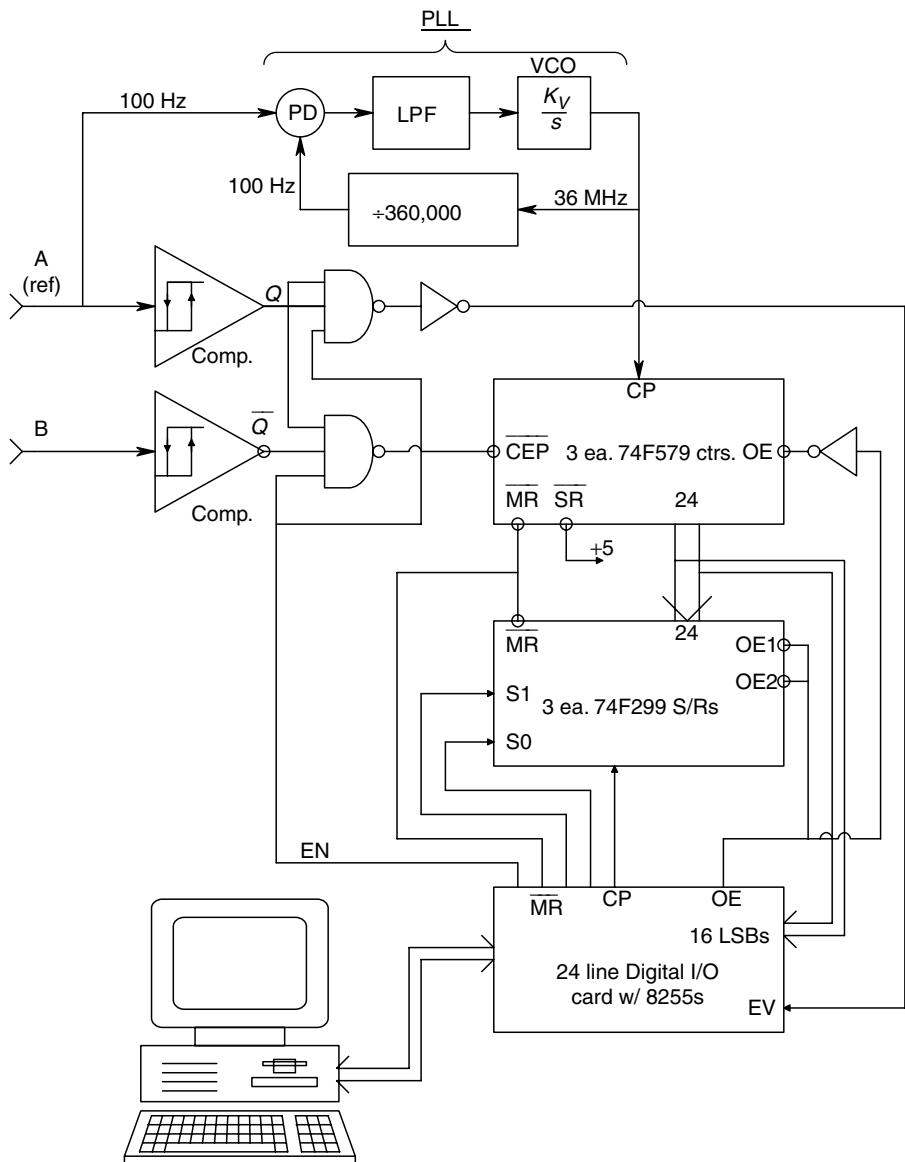


FIGURE 7.63
Millidegree phase meter designed by Du (1993).

A limitation of the Du system emerges when the phase angle, θ , is large. Now, the counter must count upwards of 360,000 pulses per cycle. The cumulative count for $N=256$ would be about 92.2 million before division by 256. Thus, if intended to work with large phase angles, the Northrop and Du system needs counters capable of holding raw counts of over 2^{26} . A practical upper bound on the input signal frequency is set by the maximum operating frequency of the PLL and the maximum counting rate of the counter. A Signetics NE564 PLL will operate up to 50 MHz. The 74F579 decade counters work up to 115 MHz.

An alternate to the use of photoelectronic measurement of the shaft's torque angle, ϕ , is the use of a variable reluctance pickup, such as shown in Figure 7.64. Here, a voltage

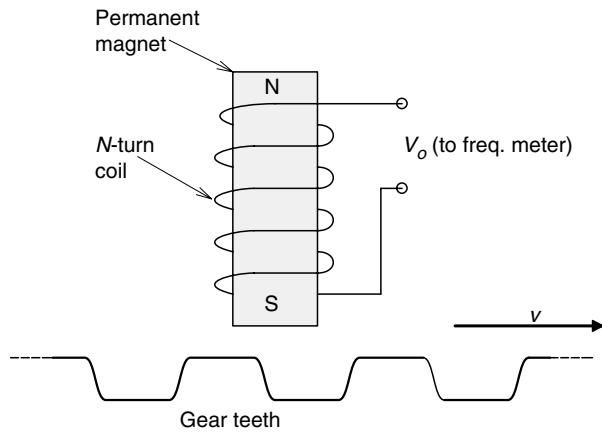
**FIGURE 7.64**

Diagram of a variable reluctance pickup. As a steel gear tooth passes under the magnet and coil, there is a transient increase in the flux linkages through the coil, hence an induced EMF pulse.

pulse is induced in the coil every time a steel gear tooth passes under the permanent magnet. Two gears, concentric with the torque shaft are used, as in the case of the photoelectronic pickup described above. Since the output pulses are biphasic, fast analog comparators, such as LT1016s, are used to convert them to TTL pulse trains for input to the phase detector.

Other schemes of no-touch torque measurement in rotating shafts have been described, which make use of the magnetoelastic (Villari) effect [Lion, 1959]. Here, the magnetic permeability of ferromagnetic materials is seen to change in a nonlinear manner when the material is subjected to mechanical stress, such as torsion. Lion reports that the Villari effect has been used to measure forces in load cells. As the force is increased, there is a nonlinear decrease in the inductance of a coil wound around the ferromagnetic core under compression. Villari sensors may be adaptable to the measurement of torque [NASA Tech Briefs, 1991]. Besides their obvious nonlinearity, the Villari effect has a strong temperature dependence which would need compensation in a field grade instrument. The inductance change resulting from the mechanical strain of the core must be detected with some form of inductance bridge or by sensing the change in the output frequency of an oscillator in which the inductance is a frequency determining component.

7.5 Pressure Measurements

Many interesting transducer systems have been developed to measure fluid pressure, as gauge pressure, absolute or differential pressure. Applications of pressure sensors are broad, including, but not limited to, aircraft altimeters, submarine depth meters, pressure meters for gas tanks, sensors for aircraft rate of climb and airspeed indicators, medical/physiological pressure sensors for blood, cerebrospinal fluid, intra-ocular pressure, intrathoracic pressure, kidney dialysis pressure, etc., hydraulic system pressure, oil pressure in bearings for turbines, generators, Diesel engines, etc., liquid level indicators for storage tanks, purge liquid pressure in oil well reconditioning, etc. Pressures involved in controlled explosions, such as combustion chamber pressures in internal combustion engines, or the gas pressure in the chamber or barrel of firearms, can be measured with special pressure sensors having high frequency response, and high temperature and pressure capabilities. Also of interest are the pressure sensors used in vacuum systems (i.e. sensors responsive to very low pressures).

In the following sections, we discuss the important, commercially available pressure sensors and the mechanisms by which they work. It will be seen that nearly all pressure sensors use some type of membrane or diaphragm which deflects linearly with pressure, relative to the zero gauge pressure condition. As we will see, measurement of this deflection has been done in a number of ingenious ways.

7.5.1 High Pressure Sensors

In this category of pressure sensor, we place all transducers acting on pressures around and above atmospheric pressure, including those that measure gauge pressure (pressures above atmospheric), absolute pressure, and pressure differences. At the low end of the measurement range, we find pressure sensors specialized for biomedical applications and applications in automobile fuel and emissions control systems. One design approach for low pressure sensors used by Schaevitz Engineering is to use compliant bellows attached to the core of an LVDT. The deflection of the bellows is proportional to the pressure difference across its walls, and models are available which use vacuum, vented gage, or differential pressure references. Such designs can be made quite sensitive (0–2 in H₂O in the Schaevitz Model P-3000 series). Due to the high compliance (low stiffness) of the bellows and the mass of the LVDT core, such sensitive pressure sensors have limited frequency response, and are generally suited for measuring pressures which vary slowly.

Another approach to pressure measurement is seen in the design of pressure sensors made by Conal Precision Instruments, Inc., Franklin Lakes, NJ. Conal pressure sensors use the pressure of a bellows against a quartz crystal used in a 40 kHz oscillator to reduce the oscillator's frequency. Unfortunately, the relation between oscillator period and input pressure is nonlinear and a calibration equation must be used to convert period to a function proportional to pressure. Conal claims resolution of 0.0001% of FS and hysteresis of 0.01% of FS. Full scale pressure ranges from 20–1000 psiA are available.

Many of the pressure sensors described below use the deflection of a thin diaphragm, or the strain produced in this diaphragm by pressure as the primary step in converting pressure to an electrical output. When pressure is applied to a flat diaphragm on the side opposite the gauges (Figure 7.65), the central gauge or gauges experiences tension, while the gauge cemented to the edge of the diaphragm sees compression. In terms of the deflection of the center of the diaphragm, deflection is linear with applied pressure for deflections up to about 30% of the diaphragm thickness [Beckwith and Buck, 1961]. The central deflection of a diaphragm is given approximately by:

$$\delta y = \frac{3\Delta P R^4(1 - \mu^2)}{16 Y t^3} \quad (7.227)$$

where ΔP is the pressure difference across the diaphragm, R is the diaphragm's radius, t is its thickness, Y is its Young's modulus and μ its Poisson's ratio. The radial stress in the diaphragm at its edge is:

$$\sigma_r = (3/4)(R/t)^2 \Delta P \text{ psi} \quad (7.228)$$

The tangential stress at the center of the diaphragm is given by:

$$\sigma_t = (3/8)(R/t)^2(1 + \mu)\Delta P \quad (7.229)$$

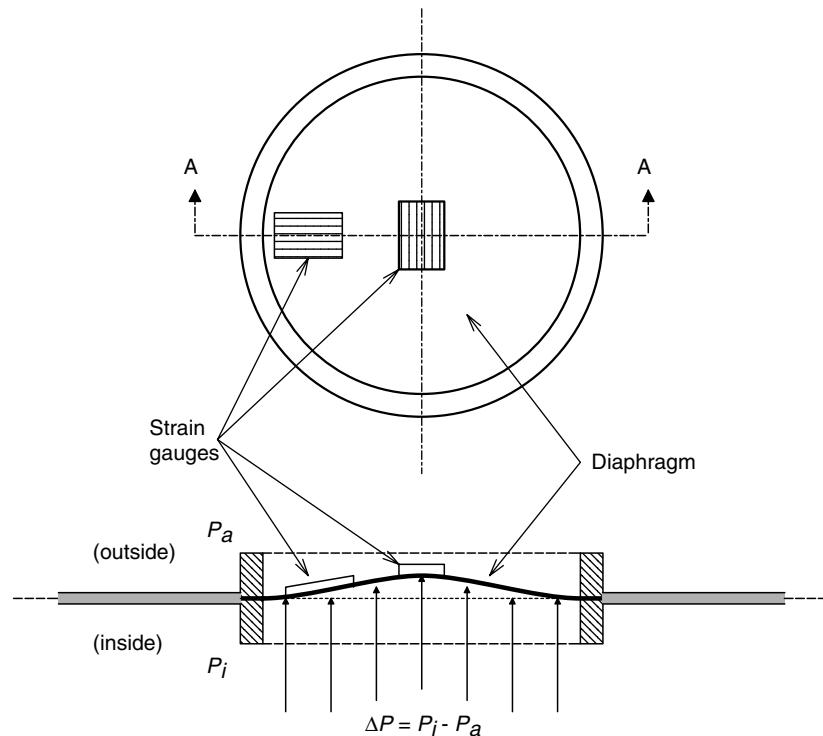
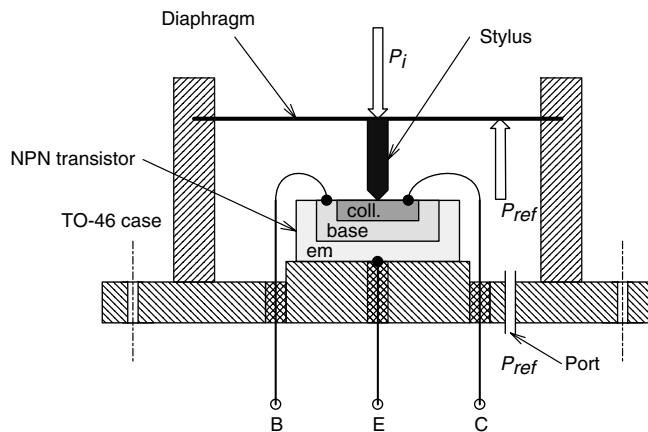
**FIGURE 7.65**

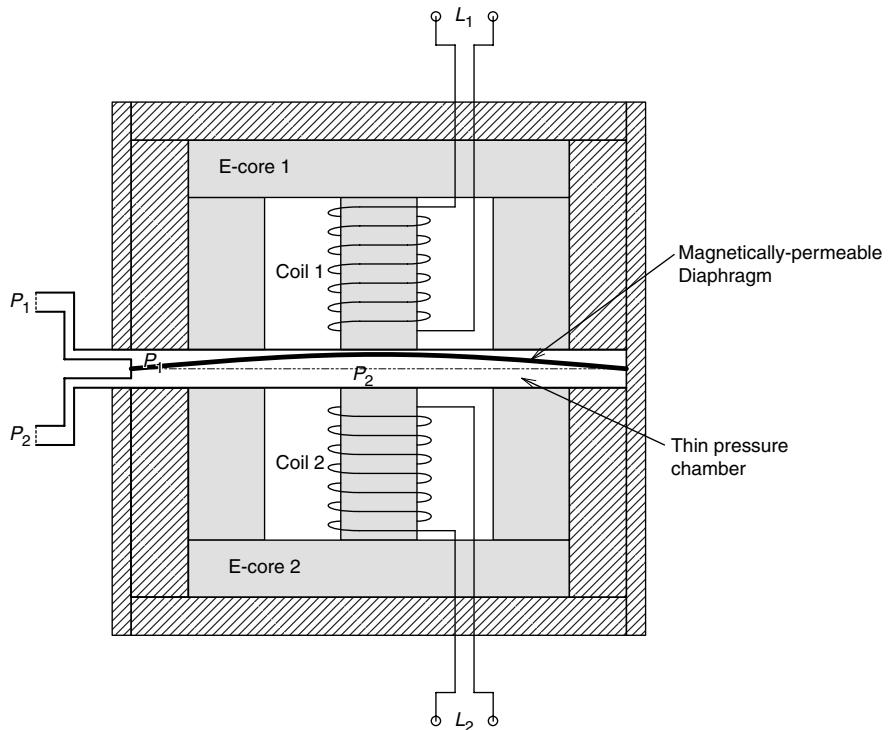
Diagram of a diaphragm pressure sensor in which the sensing elements are a pair of strain gauges bonded to the diaphragm. The strain gauges are put in one arm of a Wheatstone bridge, giving temperature compensation. (Source: With permission, Beckwith and Buck, 1961)

**FIGURE 7.66**

Section through a Pitran™ pressure sensor. (Source: Reprinted with permission, Stow Laboratories, Inc.)

The parameters are the same as for equation 7.227 above [Beckwith and Buck, 1961, Chapter 12].

Stow Laboratories, Inc., Hudson, MA, make miniature, Pitran™ pressure sensors. In the Stow design, illustrated in Figure 7.66, pressure acting on a diaphragm causes a pointed, insulated stylus to bear mechanically on the 'layer cake' of an *n-p-n* silicon transistor. This transferred pressure produces a large, reversible, effective decrease in the

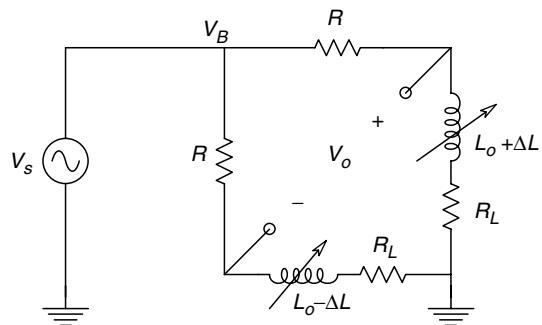
**FIGURE 7.67**

Section through a differential pressure sensor made by Validyne Engineering Corp. Sensitivity can be altered in the field by changing diaphragms.

transistor's β , lowering I_C and thus raising V_{CE} . Stow claims that Pitran have 0.5% linearity and hysteresis, and a 65 dB dynamic range. Maximum membrane deflection is 2 μ in and the Pitran's mechanical resonant frequency is about 150 kHz. Pitran models have FS pressure ranges of 0.1–20 psid.

A family of pressure sensors based on a differential, variable reluctance transducer principle is sold by Validyne Engineering Corp, Northridge, CA. A cross-sectional view of the Validyne pressure sensor is shown in Figure 7.67. The Validyne sensor design uses a diaphragm of magnetically permeable stainless steel clamped between two identical blocks. Embedded in each block are identical 'E' core and coil assemblies. With zero deflection of the diaphragm, both coil and core assemblies see an identical air gap and therefore, both coils have identical inductances. When a pressure difference exists across the diaphragm, its deflection causes an increase in the air gap or reluctance of one coil assembly and a decrease in the reluctance of the flux path of the second coil. Coil inductance, L , is inversely proportional to the reluctance of the magnetic flux path, so we see that a small bowing of the diaphragm due to ΔP will cause a corresponding $+\Delta L$ in one coil and a $-\Delta L$ in the other. This inductance change is symmetrical and can be sensed with a simple ac bridge, such as shown in Figure 7.68. It is left as an exercise to the reader to show that when the bridge sinusoidal excitation frequency is $\omega_0 = (R + R_L)/L_0$ r/s, the bridge output voltage is:

$$\frac{V_1 - V_2}{V_B} = (\Delta L/L_0)[R/(R + R_L)] \quad (7.230)$$

**FIGURE 7.68**

A simple ac inductance bridge which can be used to give an output proportional to ΔL , which in turn is proportional to ΔP in the sensor shown in Figure 7.67.

Hence, the bridge output is a DSBSC signal, and can be demodulated by a phase sensitive rectifier (PSR) and low-pass filter. The output of the PSR/LPF is seen to be linearly proportional to ΔP across the transducer diaphragm. For greater range, stiffer (thicker) diaphragms are used. Diaphragms are available to give full scale ranges from 0.0125–12,500 psi. Excitation voltage is typically 5 V RMS, 3–5 kHz. Linearity and hysteresis are typically $\pm 0.5\%$.

A pressure sensor design that is more commonly encountered than the types described above uses four strain gauges bonded to the protected side of a deflecting diaphragm. In such a full Wheatstone bridge design, two of the bridge arms increase resistance with diaphragm strain, and two decrease resistance. The close proximity of the strain gauges affords nearly complete temperature compensation. In other designs, only two gauges are active, increasing resistance with strain. The other two are inactive and in close proximity to the active gauges, so a temperature compensated output can be realized. There are a number of miniature pressure sensor designs that make use of integrated circuit fabrication techniques. In the design described by IC Sensors, Inc., Milpitas, CA, four, piezoresistive strain gauges are diffused into the surface of a single silicon crystal diaphragm to form a fully active Wheatstone bridge. The single crystal diaphragm has negligible hysteresis. Entran Devices, Fairfield, NJ makes an EPI series of ultra miniature pressure sensors using a fully active, four arm Wheatstone bridge of piezoresistive material bonded to a silicon diaphragm. In the EPI-050 series of sensors, the diameter of the body is 0.05 in and the body length is 0.25 in—full scale pressure ranges 2–300 psi, and the corresponding mechanical resonance frequencies range 100 kHz–1.7 MHz. Useful mechanical frequency response is given as 20% of the resonance frequency. Ultra miniature sensors such as the Entran EPI-050 series can be mounted in the tips of catheters and used invasively to measure pressures in the chambers of the heart, blood vessels, etc., in physiological studies.

Another approach to pressure sensor design has been devised by Motorola Semiconductors, Phoenix, AZ. Motorola's MPX, IC series of pressure sensors do not use a conventional, four-active-arm Wheatstone bridge. Instead, a single, *p*-type, diffused silicon strain resistor is deposited on an etched, single crystal silicon diaphragm, as shown in Figure 7.69. A second resistor is deposited to form an X shaped, 4-terminal resistor with two current taps (1 and 3) and two output voltage taps (2 and 4). When current is passed through terminals 1 and 3, and pressure is applied at right angles to the current flow, a transverse electric field is generated, producing an EMF between pins 2 and 4. Motorola suggests that this effect can be considered to be an electromechanical analog of the Hall effect, with pressure replacing the magnetic field. In the Motorola MPX3100 sensor series, a complete signal conditioning system with four op-amps and laser trimmed resistors has been built on the margin of the silicon wafer holding the diaphragm and the transverse voltage strain gauge element (Figure 7.70). The output of the MPX3100 is

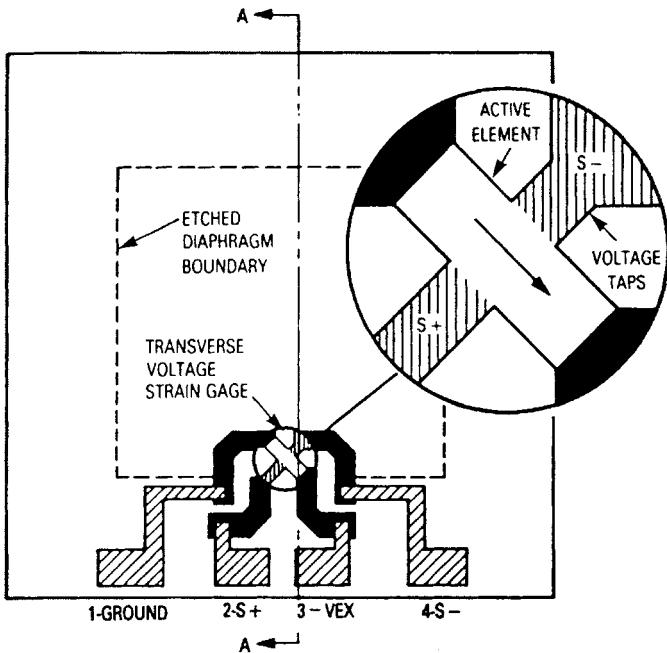
**FIGURE 7.69**

Diagram of the etched microcircuit diaphragm of a Motorola MPX series pressure sensor. The patented X shaped structure is a four-terminal resistor with two voltage and two current taps. Motorola claims that the sensor is the electromechanical analog of a Hall effect device. (Source: Courtesy of Motorola Semiconductors)

quite linear, $\pm 0.2\%$ FS. Temperature compensation of full scale span is $\pm 2\%$ FS over a $0\text{--}85^\circ\text{C}$ range. In the design of the MPX2000 series pressure sensors, five laser trimmed resistors and two thermistors are deposited on the margin of the silicon chip around the pressure diaphragm. These resistors and thermistors are used to achieve a $\pm 1\%$ FS temperature effect over a $0\text{--}85^\circ\text{C}$ span. A similar temperature range on one of the non-thermally compensated sensor models, such as the MPX10, would be a whopping $\pm 16.2\%$ FS.

Measurement of extremely high pressures in the range of $10,000\text{--}>250,000$ psi requires special transducer designs. The simplest is a cylindrical pressure cell, illustrated in Figure 7.71. Two active bonded strain gauges respond to the outer surface's hoop stress caused by pressure difference across the cylinder's wall. Two reference strain gauges are located on the top of the cylinder. It may be shown that the outer surface's hoop strain is given by:

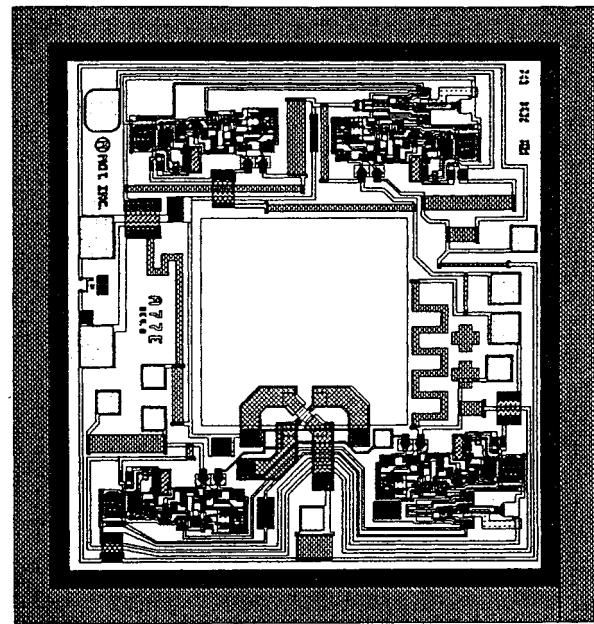
$$\varepsilon_H = \frac{P_i d^2 (2 - \mu)}{Y(D^2 - d^2)} \text{ inch/inch} \quad (7.231)$$

where Y is Young's modulus, D is the cylinder's outer diameter, d is its inner diameter, μ is Poisson's ratio and P_i is the internal pressure [Beckwith and Buck, 1961].

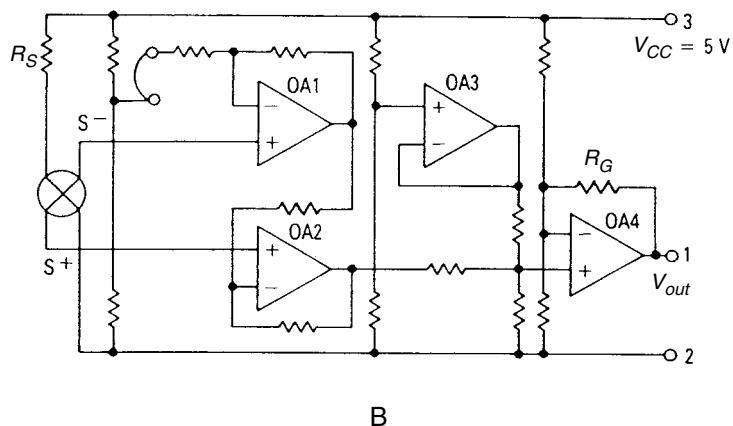
A different sensor mechanism which is used to measure extremely high pressures makes use of the fact that the electrical resistance of a conductor changes if it is compressed by an applied, external pressure. Such sensors are called bulk compression gauges (Figure 7.72). The resistance of the compressed conductor may be approximated by

$$R(P) = R_o(1 + bP) \quad (7.232)$$

where R_o is the resistance of the conductor at 1 atmosphere and standard temperature, and b is the pressure coefficient of resistance.



A

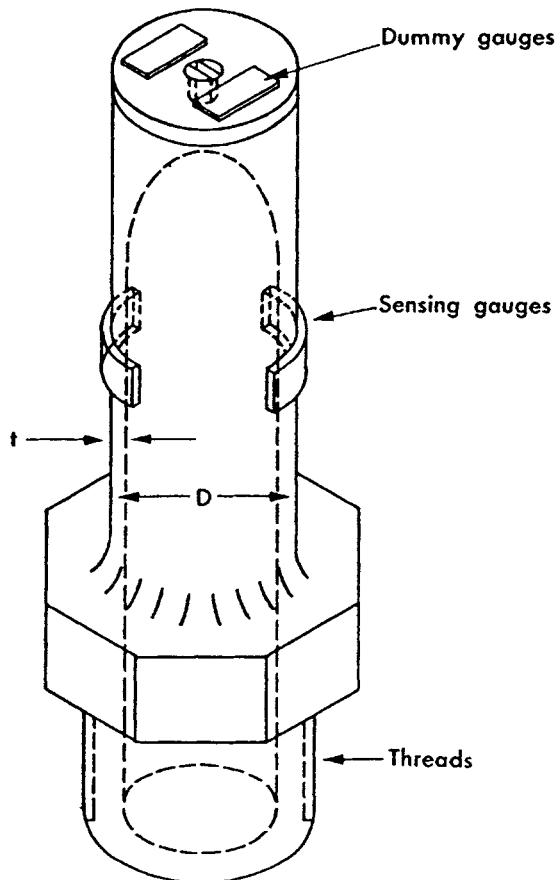


B

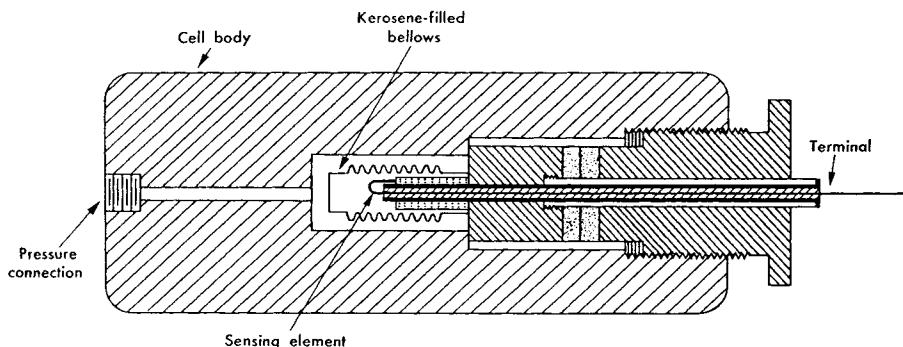
FIGURE 7.70

A. Top view of the integrated circuit of a Motorola MPX3100 pressure sensor. B. Schematic of the temperature compensated MPX3100 pressure sensor. (Source: Courtesy of Motorola Semiconductors)

Two metals commonly used for bulk compression gauges are an alloy of gold and 2.1% chromium and manganin. The gold alloy is preferred because of a much lower tempco than manganin ($\Delta R/R = 0.01\%$ for a $21.1\text{--}82.2^\circ\text{C}$ change in temperature for the gold alloy *vs.* 0.2% for manganin). The $(\Delta R/R)/P$ for the gold alloy is $0.673 \times 10^{-7} \Omega/\Omega/\text{psi}$, and is $1.692 \times 10^{-7} \Omega/\Omega/\text{psi}$ for manganin. Although the gold alloy has a lower pressure sensitivity, its lower tempco makes its use preferred in bulk compression gauges [Beckwith and Buck, Chapter 12, 1961]. A bulk compression pressure sensor, which works up to $1.5 \times 10^6 \text{ psi}$, was described by Hall (1958). Bulk compression pressure

**FIGURE 7.71**

Cylindrical high pressure sensor cell which uses four bonded strain gauges in a Wheatstone bridge; the two gauges on the end are inactive and are used for temperature compensation. (Source: With permission, Beckwith and Buck, 1961)

**FIGURE 7.72**

Section through a very high pressure, bulk modulus compression, pressure sensor. The sensing resistor is made one arm of a Wheatstone bridge. (Source: With permission, Beckwith and Buck, 1961)

sensors must be temperature compensated. Technical difficulties are encountered at very high pressures in bringing the insulated resistance element's leads out of the case without leaks or plastic flow problems.

The final type of pressure sensor that we will describe in this chapter is the quartz crystal, piezo electric gauge. As we have seen in Section 6.4.3, quartz transducers are only suitable for measuring transient and time varying pressure changes which contain

frequencies above their low frequency (high-pass) pole. Quartz transducers do not respond to static (constant) pressures. Since quartz pressure sensors work well at elevated temperatures and respond to transient changes in pressure with frequency responses which typically extend to hundreds of kHz, they are often used to measure the transient pressure changes associated with combustion in internal combustion engines, as well as the transient pressure associated with internal ballistics of firearms (peak chamber pressures in military and hunting rifles are typically 50–65,000 psi). Some quartz dynamic pressure sensors are water cooled to prevent the quartz crystal from reaching its Curie temperature and failure in certain high temperature applications such as measurement of cylinder pressure in an internal combustion engine. The output of a quartz piezoelectric pressure sensor is generally conditioned by a charge amplifier (Section 2.7.1). Kistler Instrument Corp., Amherst, NY, make a wide selection of quartz pressure sensors. For example, the Kistler model 6205 sensor is specifically designed for ballistics research and ammunition testing. It has a working range of 0–75,000 psi, a charge sensitivity of -0.09 pC/psi , a resonant (upper) frequency of 300 kHz, a rise time of $1.5 \mu\text{s}$, nonlinearity and hysteresis of $\leq 1\% \text{ FS}$, operating temperature range of -50 to 200°C . The Kistler model 6213 piezoelectric pressure sensor has a maximum range of 145,000 psi, and their special Z series of sensors operate over an extended temperature range of -196 to 400°C . PCB Piezotronics, Inc., Buffalo, NY also makes a full line of quartz pressure sensors, accelerometers and load cells.

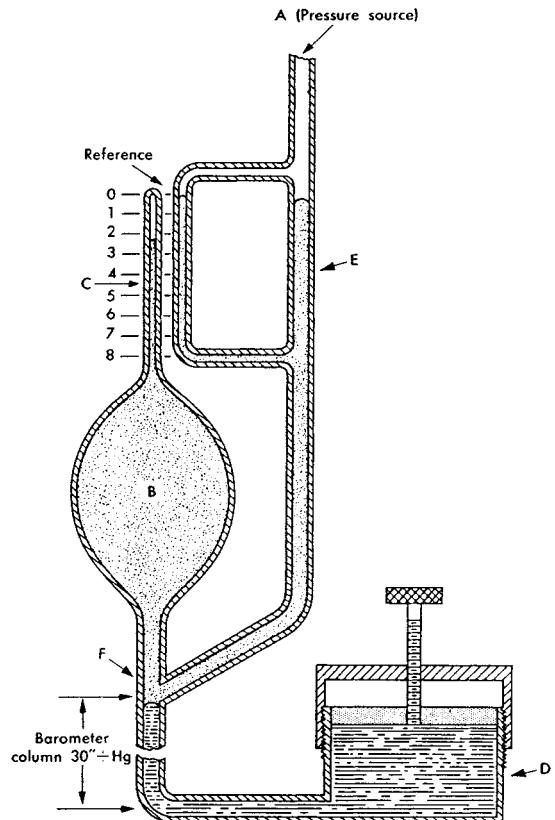
7.5.2 Low Pressure Sensors

Here, we define low pressure sensors as those sensors designed to measure pressures that are significantly lower than atmospheric. Their primary application is in vacuum system measurements. Two means of measuring low pressures do not have electrical outputs. The McLeod gauge, shown in Figure 7.73, is a mercury filled glass system which is operated manually and read out visually. Its useful range is typically 5–0.005 mm Hg pressure. The McLeod gauge makes use of Boyle's law, $P_1 = P_2 (V_2/V_1)$. P_1 is the pressure we desire to measure, P_2 is the increased pressure in the McLeod gauge obtained by compressing the gas from volume V_1 to a smaller volume, V_2 . The gas whose pressure is under measurement should not contain water vapor, or the compression process will cause condensation. Beckwith and Buck (1961), describe the operation of a McLeod gauge:

"Measurement is made as follows. The unknown pressure is connected to the gauge at point A, and the mercury level is adjusted to fill the volume represented by the darker shading. Under these conditions the unknown pressure fills the bulb B and capillary C. Mercury is then forced out of the reservoir D, up into the bulb and reference column E. When the mercury level reaches the cut-off point F, a known volume of gas (V_1) is trapped in the bulb and capillary.

The mercury level is then further raised until it reaches the zero reference point in E. Under these conditions the volume remaining in the capillary is read directly from the scale, and the difference in heights of the two columns is the measure of the trapped pressure (P_2). The initial pressure may then be calculated by use of Boyle's law.

Another directly indicating pressure gauge is the General Electric 'Molecular Vacuum Gauge'. The GE gauge is calibrated over a 0.002–20 mmHg range. In a cylindrical chamber which is at the pressure under measurement, a synchronous motor drives a fan. The gas in the chamber is given an angular momentum by the fan and its molecules strike the blades of an adjacent, restrained fan connected to a pointer and a helical spring. The higher the gas pressure in the chamber, the higher the density of molecules striking the pointer's fan and the

**FIGURE 7.73**

Section through a McLeod vacuum gauge. Mercury is the fluid used.

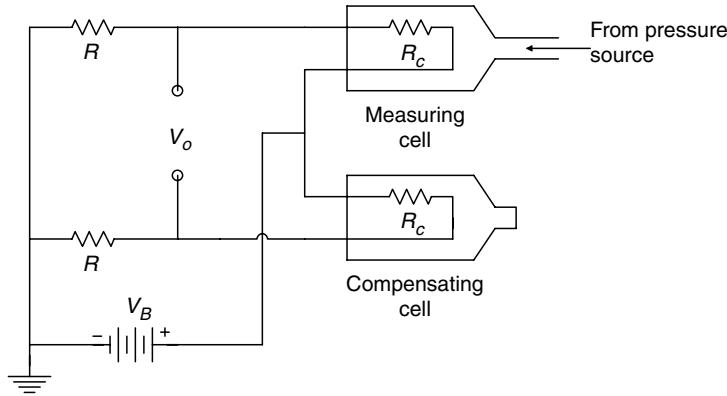
more torque produced on it, hence the greater the deflection of the pointer. The GE gauge's scale is nonlinear, with the highest resolution at the low end of the scale (0–0.1 mmHg covers the first 90° of meter deflection, while 0.1–20 mmHg covers the remaining 150° of scale)."

It is, of course, important to have low pressure sensors which can be read out electrically. The first and simplest electrical vacuum gauge is the Pirani gauge. The Pirani gauge consists of a filament of platinum, tungsten or nickel in a glass bulb connected to the vacuum system. Various gases have different heat conductivities, so the Pirani gauge must be calibrated for the gas or gases being used. A simple operating mode of the Pirani gauge is to place it in one arm of a Wheatstone bridge circuit, as shown in Figure 7.74. An identical gauge *in vacuo* is used as a temperature reference. Recall from Section 7.3.2.1 on hot wire anemometers that in a vacuum, there will be a rise in the Pirani gauge's filament temperature given by:

$$\Delta T = I_o^2 R_o \Omega_o / (1 - \alpha I_o^2 \Omega_o) \quad (7.233)$$

where I_o is the RMS current through the Pirani filament, R_o is the resistance of the filament at ambient (reference) temperature, α is the tempco of the wire, ΔT is the equilibrium increase in temperature of the wire above ambient temperature and Ω_o is the thermal resistance of the filament *in vacuo*, in °C/W. We also showed in Sec. 7.2.2.1 that the resistance of the hot filament *in vacuo* is given by:

$$R(\Delta T, 0) = R_o / (1 - \alpha I_o^2 \Omega_o) \quad (7.234)$$

**FIGURE 7.74**

Schematic diagram of a Pirani low pressure sensor.

Now, as the gas pressure is increased from zero, filament heat is no longer only dissipated by radiation and conduction through the leads, but also by heating gas molecules and giving them increased kinetic energy. Thus, as the pressure increases, Ω decreases from its vacuum value. At low pressures (below 10 mmHg) [Lion, 1959], the thermal resistance may be given by

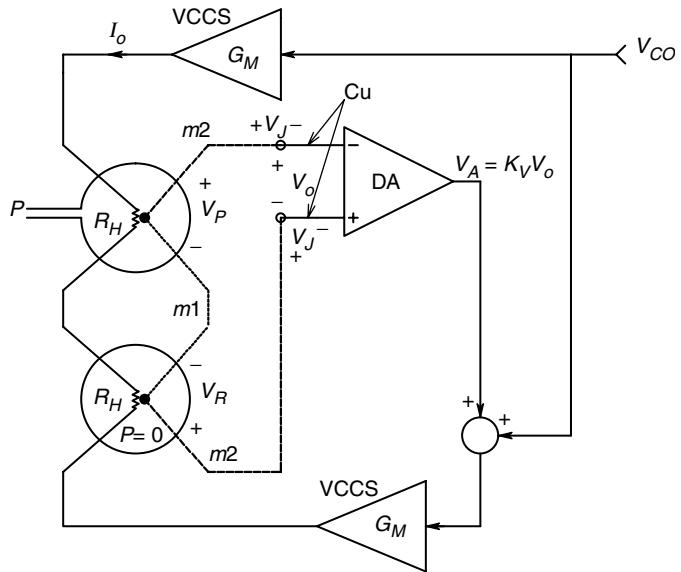
$$\Omega(P) = \Omega_0(1 - \sigma P) \quad (7.235)$$

where σ is dependent on system geometry, filament material and the heat conductivity of the gas. If equation 7.235 above is substituted into equation 7.234, then we see that the Pirani gauge filament resistance may be approximated by:

$$R(P) = R(\Delta T, 0) \left[1 - \alpha \frac{I_0^2 \Omega_0 \Delta P}{1 - \alpha I_0^2 \Omega_0} \right] \quad (7.236)$$

The linear working range of most Pirani gauges is from 10^{-5} –1 mmHg, although modifications have allowed measurements of pressures as low as 5×10^{-9} mmHg. Of interest is the closed loop, feedback operation of the Pirani gauge. In this mode of operation, the Wheatstone bridge is balanced at some reference pressure, P_r . As P increases from P_r , the thermal resistance decreases, and the Pirani filament cools and its temperature and resistance decrease. Decreasing resistance causes the bridge output to increase. This increase in V_o is applied to increase the voltage across the bridge, thereby increasing the power dissipation in the Pirani filament and increasing its temperature and resistance, restoring bridge balance. The other resistors in the bridge are assumed to have very low thermal resistances and tempcos.

Another low pressure (vacuum) sensor makes use of the heater/vacuum thermocouple device introduced in Section 6.4.1 as a means of measuring true RMS current. In the vacuum version of this sensor, the thermal resistance of the heater decreases with increasing gas pressure. This is because at increased gas pressure, there are more gas molecules present to absorb heat energy from the heater and convey it to the walls of the glass envelope, where it is lost by radiation and convection. Hence the heater temperature drops (at constant power input) as the gas pressure increases, and thus the EMF of the associated thermocouple also decreases with the pressure increase.

**FIGURE 7.75**

Schematic diagram of a feedback, vacuum thermocouple, low pressure sensor. The two thermocouples must have matched characteristics. Amplifiers G_M are voltage controlled current sources, DA is a differential amplifier (VCS). For the reference thermocouple we have a heater temperature rise given by:

Thermocouple vacuum sensors work in the range from 10^{-3} to 10 mmHg. Below pressures of 10^{-3} mmHg, heat loss through the heater wire supports exceeds that from the gas molecules, and the thermocouple pressure sensor loses linearity and sensitivity. Thermocouple pressure sensors are ideally suited for feedback operation. One such closed loop system is shown schematically in Figure 7.75. In this system, feedback is used to nearly match the dc EMF from the pressure responsive thermocouple to the dc EMF from the reference (vacuum) thermocouple. For the reference thermocouple we have a heater temperature rise given by:

$$\Delta T_r = I^2 R_W \Omega_o \quad (7.237)$$

The thermocouple EMF is proportional to ΔT_r :

$$V_r = K_T (I^2 R_H \Omega_o) \quad (7.238)$$

For the pressure responsive thermocouple we have:

$$V_p = K_T [I_o^2 R_W \Omega_o (1 - \sigma P)] \quad (7.239)$$

Here, we assume that at very low pressures, the thermal resistance of the heater decreases linearly with increasing pressure, P . The current in the pressure responsive thermocouple heater is constant.

$$I_o = G_M V_{co} \quad (7.240)$$

G_M is the transconductance of the two VCCSs, and V_{co} is a dc voltage. Thus the current in the reference thermocouple's heater is:

$$I = G_M (V_{co} - K_V V_o) \quad (7.241)$$

where K_V is the differential amplifier gain and V_o is the difference between the reference thermocouple's EMF and the pressure responsive thermocouple's EMF. That is:

$$V_o = V_r - V_p \quad (7.242)$$

If the relations above are substituted into equation 7.242, we obtain a quadratic equation in V_o :

$$0 = V_o^2 - V_o \frac{1 + 2V_{co}\mu}{\mu K_V} + \frac{P\sigma V_{co}^2}{K_V^2} \quad (7.243)$$

where

$$\mu = K_V K_T R_W \Omega_o G_M^2 \quad (7.244)$$

If we assume that $2V_{co} K_V K_T R_W \Omega_o G_M^2 \gg 1$, then one root of the quadratic equation can be shown to be:

$$V_o = P\sigma V_{co}/2 \quad (7.245)$$

The differential amplifier's output is:

$$V_A = P\sigma K_V V_{co}/2 \quad (7.246)$$

Thus, the proposed design for the feedback thermocouple, low pressure sensor system is directly proportional to the pressure. Note that the calibration depends on the value of the constant, σ , which will depend on the composition of the gas. Hydrogen would have a larger σ than would air, for example.

The ionization vacuum gauge is routinely used to measure very low pressures, in the range of 10^{-8} – 10^{-3} mmHg. Special designs have been used down to 10^{-10} mmHg [Lion, 1959, Section 1-57]. The ionization gauge is constructed similar to a triode vacuum tube, as shown in Figure 7.76. A glass envelope is kept at the pressure under measurement. The envelope contains a heated cathode which emits electrons.

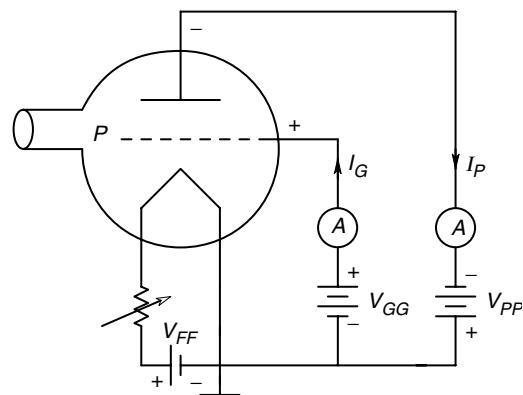


FIGURE 7.76

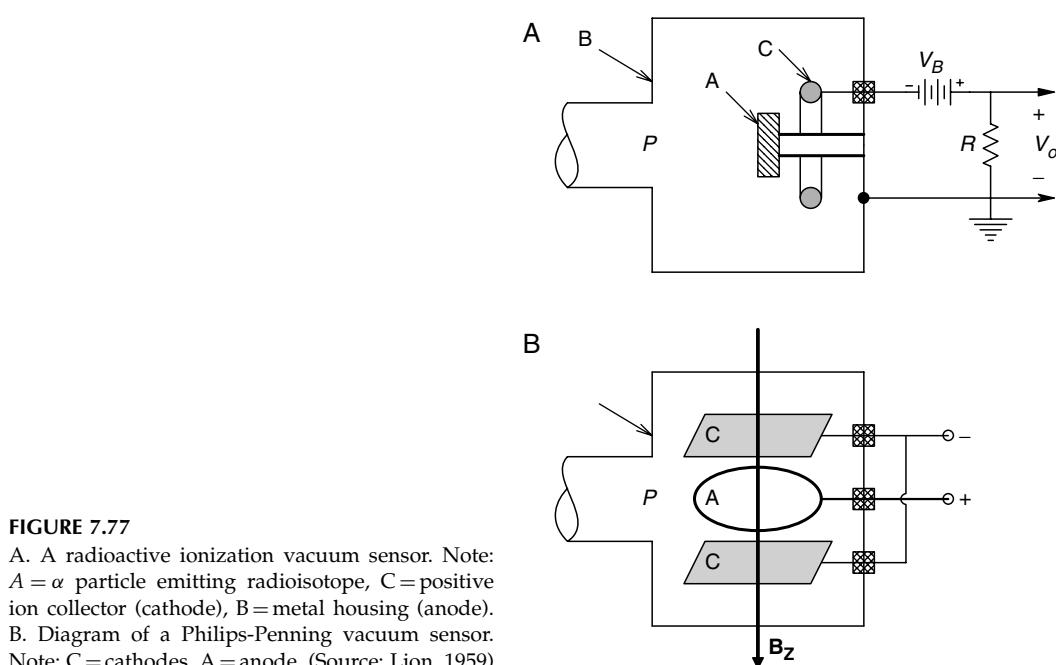
A vacuum triode ionization gauge used to measure very low pressures ($10^{-8} \leq P \leq 10^{-3}$ mmHg).

Surrounding the cathode is a grid and outside the grid is a plate. Unlike a vacuum tube triode, the ionization detector's grid is maintained at a high (100–250 V) positive potential with respect to the cathode. The plate is kept at a negative potential (−2 to −50 V). Electrons from the hot cathode are accelerated by the field from the grid, collide with gas molecules and ionize them. The positive ions in the space between the grid and the plate are collected by the negative plate; those formed in the space between the cathode and the grid are accelerated toward the cathode. The free electrons and negative ions are collected by the positive grid. The rate of ion production is proportional to the density of gas molecules in the tube (or to the gas pressure) and to the number of electrons available to ionize the gas. Lion (1959) shows that the gas pressure is proportional to the ratio of the positive ion current at the plate to the grid current. Thus:

$$P = K(i_{p+}/i_{g-}) \quad (7.247)$$

where K is about 10^{-2} mmHg (mA/ μ A). K depends on the gas composition, the geometry of the tube and the voltages used. The emission of photoelectrons from the plate, caused by light and soft X-rays from the grid sets the lower bound of ion gauge sensitivity. One mode of operation of the ion gauge is to keep the grid current constant (1–20 mA) and to measure the plate current. A log ratio I_C (*cf.* Section 2.6.7) can also be used to measure P over the wide range of this sensor.

Instead of creating an ion current by collisions with energetic electrons, a radioactive, α particle emitter can be used to make a radioactive ionization gauge, illustrated in Figure 7.77A. The number of positive ions formed is proportional to the gas pressure as long as the range of the α particles is longer than the dimensions of the chamber. The useful range of pressures is 10^3 – 10^{-3} mmHg. Below 10^{-3} mmHg, the mean free path of the α particles increases beyond the dimensions of the chamber so that the probability of an ionizing collision with a gas molecule is reduced and the ion current drops off



sharply. One advantage of the radioactive ionization gauge is that it is not damaged if it is turned on at atmospheric pressure, and it requires no degassing.

The final low pressure sensor we will describe in this section is the Philips-Penning Gauge, illustrated in Figure 7.77B. In this gauge, collisions between gas molecules and moving electrons create an ion plasma, which contributes to the gauge current. Electrons are emitted as the result of the high electric field and bombardment of the electrode surfaces by ions. The working range of the Philips-Penning Gauge is 4×10^{-7} – 10^{-3} mmHg [Lion, 1959]. The electrodes of this gauge are run at potentials of 1–3 kV and a transverse dc magnetic field of 300–8000 Oersteds is used to decrease the mean free path of electrons and thus extend the low pressure range. One version of the Philips-Penning gauge, the inverted magnetron, was reported to work at pressures as low as 10^{-12} mmHg [Lion, 1959].

7.6 Introduction to Substance Detection and Measurement Using Photons

7.6.1 Introduction

In this section, we will describe various methods of detecting the presence of and measuring the concentration or mass of certain chemicals (analytes) using electromagnetic (EM) radiation, ranging from the far infrared (FIR) to far ultraviolet (UVC). Figure 7.78 illustrates the EM spectrum. The (photon) wavelength range used for spectrometry spans 430 μm –430 pm, going from FIR to UVC. When EM waves in this wavelength range interact with matter of finite thickness, some fraction R of the incident energy is reflected, a fraction A is absorbed as the radiation passes through, and a fraction T is transmitted. It should be obvious that $R+A+T=1$. In using light to measure how much of a substance is present, we generally ignore R and examine A or T as a function of wavelength (i.e. $A(\lambda)$). The transmittance of a substance through which ‘photon’ EMR passes is defined by:

$$T(\lambda) \equiv P_{out}/P_{in} < 1 \quad (7.248)$$

where P_{in} is the radiant input power to the substance and P_{out} is the emerging radiant power. The transmittance is a function of the wavelength of the incident light in space. The absorbance, A , aka optical density, OD, is defined as:

$$A = OD = -\log_{10}[T(\lambda)] = \log_{10}[1/T(\lambda)] \quad (7.249)$$

The optical absorbance as a function of wavelength or wavenumber is a unique, physical-chemical ‘fingerprint’ of the analyte. The particular absorbance peaks are caused by the interaction of incident photons having energy $h(c/\lambda)$ with certain molecular bonds. Molecular subgroups such as –H, –OH, –CH₃, –NH₃, etc. are attached to other larger molecular units such as benzene rings, purine rings, sugars, fatty acids, etc. All chemical bonds are elastic, and allow the joined molecules to move in complex vibrational modes when they absorb low energy photons having their natural mechanical frequencies. Stretching, twisting (rotating), waving, rocking, etc. are some of the complex elastic motions possible. When the photons striking a particular molecular bond have the correct energy to excite a resonant mode of the bond, they are selectively absorbed, causing

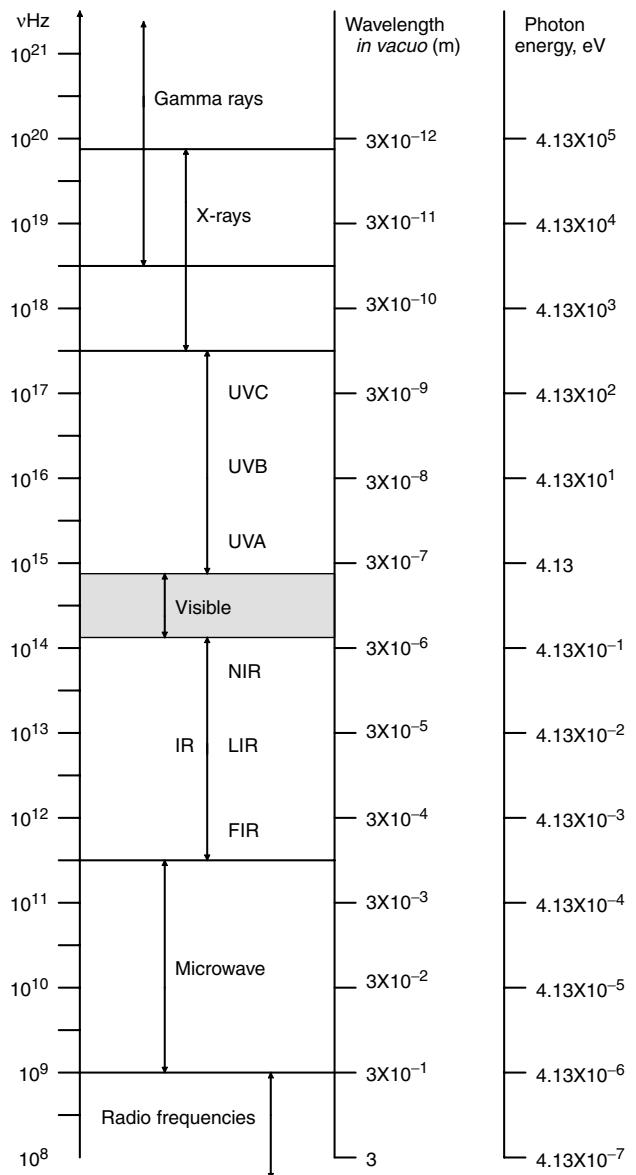


FIGURE 7.78
The electromagnetic spectrum.

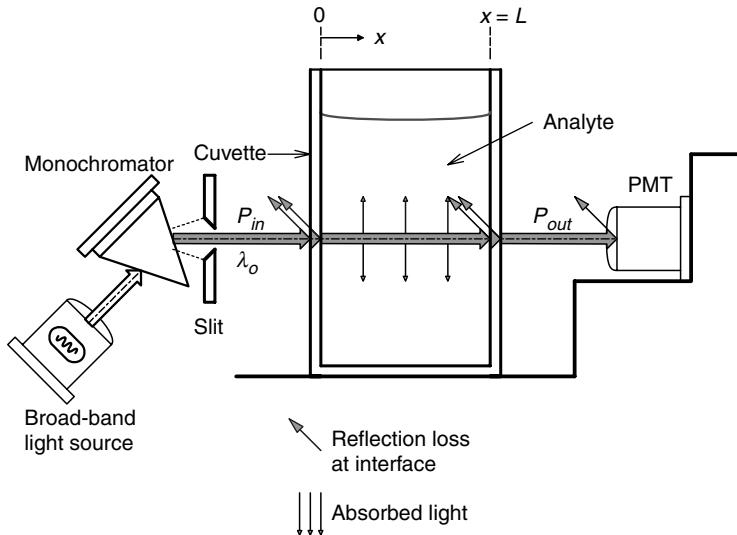
enhanced bond resonance. Thus, the EMR exiting the sample has absorbance peaks at the wavelengths (energies) of the photons which excite molecular resonances. Table 7.2 illustrates the types of photon/molecular interactions as a function of wavelength and wavenumber [Northrop, 2002].

Beer's Law is often used to model absorbance (Figure 7.79). First, note that a small fraction of the impinging radiant power, $R = R_{inC} + R_{inS} + R_{outS} + R_{outC} + R_{PMT}$, is reflected at the air/cuvette interfaces, the cuvette/analyte interfaces and the air/PMT interface, where there is a difference in refractive index. The actual spectral radiant power absorption is assumed to take place in the medium of interest (the analyte) inside the cuvette. Cuvettes can be made of glass, quartz, plastic, or exotic materials such as germanium metal, or IRTRAN™, depending on the range of wavelengths used. Beer's Law states that the rate of decrease of radiant power with distance as light passes through a homogeneous medium is proportional to the concentration C of the absorbing substance (for solutions) times the power.

TABLE 7.2

Wavelength, wavenumber, energy levels, and type of molecular energy transition when photons are absorbed by molecules

λ range	ξ range, cm^{-1}	Type of radiation	Energy transition	$h\nu/q$ range, eV
100–10 cm	0.01–0.1	VHF–μwave radio	Spin orientation	1.24E–5–1.24E–4
10–1 cm	0.1–1	μwave	Molecular rotations	1.24E–4–1.24E–3
1–0.1 cm	1–10	μwave–FIR	Molecular rotations	1.24E–3–1.24E–2
100–10 μm	10–100	FIR–MIR	Molecular vibrations	1.24E–2–1.24E–1
10–1 μm	100–10 ³	MIR–NIR	Molecular vibrations	1.24E–1–1.24
10 ³ –100 nm	10 ³ –10 ⁴	Visible–UV	Valence e [–] transitions	1.24–1.24E1
100–10 nm	10 ⁴ –10 ⁵	UVC (vacuum UV)	Valence e [–] transitions	1.24E1–1.24E2
10–1 nm	10 ⁵ –10 ⁶	UVC–X-rays	Inner shell e [–] transitions	1.24E2–1.24E3
1–0.1 nm	10 ⁶ –10 ⁷	X-rays	Inner shell e [–] transitions	1.24E3–1.24E4
<0.1 nm	> 10 ⁷	γ-rays	Nuclear transitions	>1.24E4

**FIGURE 7.79**

A simple, single beam optical system illustrating optical absorption, reflection, and Beers' law.

In mathematical terms:

$$-\frac{dP}{dx} = \alpha' PC \quad (7.250)$$

↓

$$-\int_{P_{in}}^{P(L)} \frac{dp}{p} = \int_{x=0}^L \alpha' C dx \quad (7.251)$$

↓

$$\log_{10}[P_{in}/P(L)] = \alpha CL, \quad \alpha = \alpha'/2.303 \quad (7.252)$$

Thus, it is clear that $OD = \alpha CL$ and $\%T = 10^{[-\alpha(\lambda)CL+2]}$. Note that α , the extinction coefficient, aka the absorptivity, of the solution, is a function of wavelength with many peaks and valleys that can be related to photon energy being absorbed by aspects of the analyte's molecular structure. The absorptivity is the analyte's spectrometric is called the 'finger print'.

Note that in Figure 7.79, the monochromator was represented as a prism. Modern spectrometers generally use one or two diffraction gratings and slits to obtain a narrow wavelength beam of light (high full power/half width or Q) for analysis. For example, the Czerny-Turner single grating monochromator uses a plano grating and two concave focusing mirrors (*cf.* Figure 7.80). A detailed mathematical treatment of grating monochromator design, dispersion, errors and aberrations can be found in Jobin Yvon (2002).

If continuous scanning of a range of the spectrum is not necessary, then a group of laser diodes emitting at a set of peak wavelengths, (λ_k), can be used to quantify known peak absorption wavelengths of the analyte(s).

7.6.2 Dispersive Spectrophotometry

Dispersive spectrometry (DS) methods constitute a large number of UV, visible light and IR spectroscopy systems. DS refers to the fact that the light or EM radiation passed through the sample (analyte) must pass through a monochromator either before or after the analyte, to describe the analyte's absorption/transmission characteristic spectrum. The simplest monochromator is a prism followed by a slit to select a narrow band of wavelengths centered around some λ_o . A beam of 'white' light is directed into the prism and a fan shaped beam of light, dispersed by wavelength, exits the prism. The desired λ_o is selected by rotating the prism. The slit defines the desired wavelength band. Modern monochromators use one or two diffraction gratings which consist of a glass plate into which are etched very fine, closely spaced, parallel lines. Refraction, reflection and interference occur when a beam of white light impinges on a grating, creating an output light beam with a peak λ_o , which is dependent on the incident and exit angles. A grating output generally has a higher Q (peak power wavelength λ_o /half power wavelength width) than a prism, and by using two gratings in series, one can obtain an even more spectrally pure output beam, albeit at the expense of intensity.

Figure 7.80 illustrates a single beam spectrophotometer. White light from a tungsten or halogen source is directed to a grating. The exit slit further increases the Q of the exit

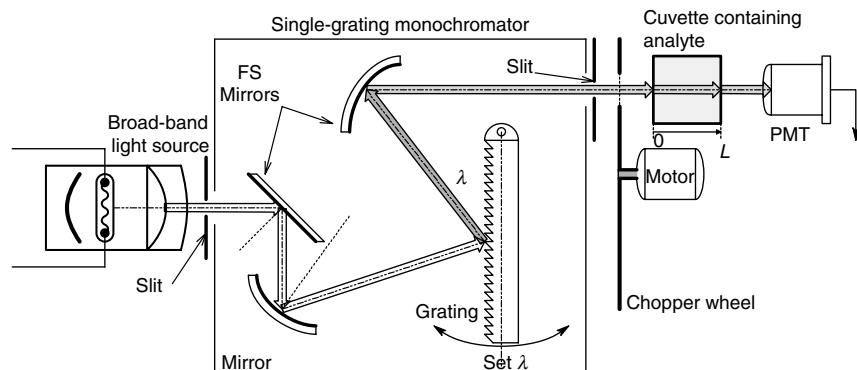


FIGURE 7.80

A single beam spectrophotometer.

beam, which is passed through a chopper wheel which periodically interrupts the beam. The chopping rate is typically in the hundreds of pps. The chopped beam then passes though a solution of the analyte in a special chamber known as a cuvette. The light path through the analyte in the cuvette is L cm, where L can be 0.1, 1, 2, 10 cm, etc., depending on the absorptivity of the solution. The exiting light power from the cuvette is measured by a calibrated photomultiplier tube (PMT). The PMT dc dark current is rejected by the phase sensitive rectifier (PSR), which responds only to the ac (chopped) signal from the PMT. The intensity of the light source must be stabilized by using a regulated dc power supply. The test beam's λ_o is selected by varying the grating angle. Unfortunately, as λ_o is varied, so does the intensity of the input beam to the cuvette. To compensate for this effect, a run is made with a cuvette full of distilled water or a reference solution, and the $P_{outRef}(\lambda)$ recorded and used to compensate for the run with the analyte. A phase sensitive demodulator is used to detect $P_{outAn}(\lambda)$, and a computer calculates $A(\xi)$, $A(\lambda)$, $T(\xi)$ or $T(\lambda)$.

Figure 7.81 illustrates one version of a dual beam spectrophotometer. This design is more complex optically, because it simultaneously measures PMT dark current, reference beam transmission and analyte transmission. A special, 3-state chopper wheel is used that has holes and mirrors in it such that the PMT alternately sees dark, the analyte beam and the reference beam. Sample and hold circuits allow appropriate corrections and comparisons to be made from the 3-state-in-time PMT waveform.

The spectrogram for an analyte is usually plotted as $\%T$ vs wavenumber, ξ , in cm^{-1} . $\xi(\text{cm}^{-1}) \equiv 10^4/\lambda (\mu\text{m})$. ξ is proportional to the EM wave's frequency, ν (Hz) = $c(\text{m/s})/\lambda(\text{m})$. By using a wavenumber abscissa, $\%T$ peaks crowded together at longer wavelengths are separated and given more detail.

Figure 7.82 illustrates the use of an attenuated total reflection (ATR) plate in spectrophotometry where we wish to measure the wavelength dependent absorption

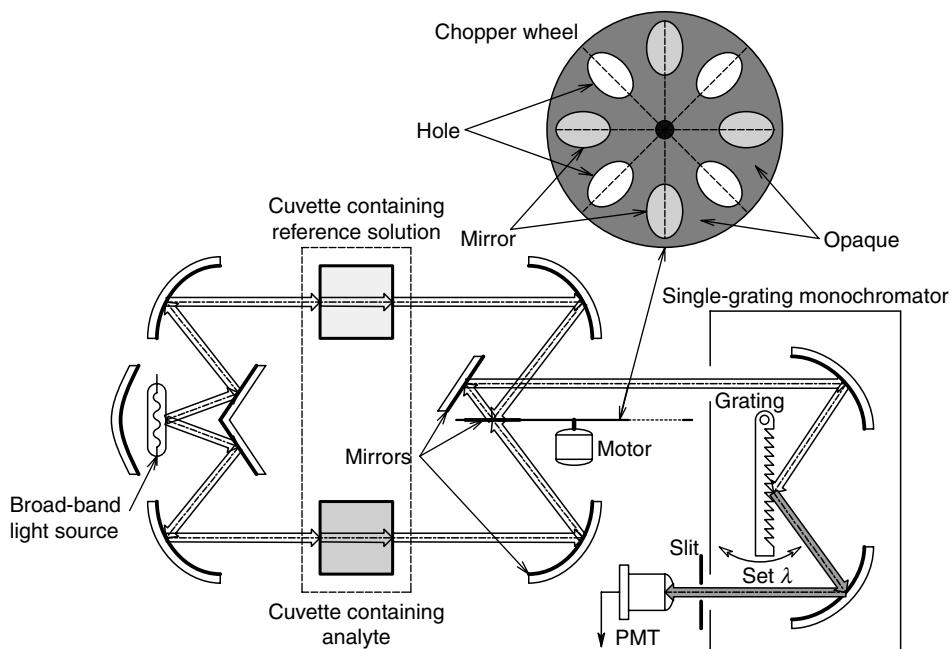
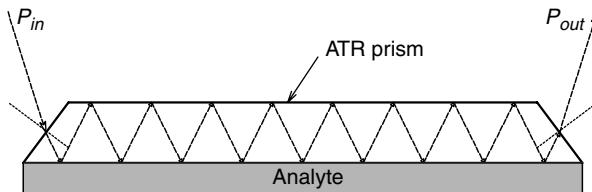


FIGURE 7.81

A dual beam spectrophotometer. A grating monochromator is used.

**FIGURE 7.82**

An attenuated total reflection (ATR) plate or prism. The ATR plate allows multiple beam interactions with the surface of the analyte.

of backscattered light from thin films of analyte, including superficial tissue (dermis, capillaries, etc.). This approach can be used to measure analytes in liquid blood samples and may find application in measuring capillary blood constituents such as glucose, cholesterol, alcohol, heroin (diacetyl morphine) etc. The input light enters the ATR prism and is directed through its bottom surface into the absorbing sample. The light that is back-scattered from the sample re-enters the ATR prism and is totally reflected from its top surface, being directed again into the tissue, where it is again back-scattered, and so on. The repeated re-entry and back-scattering of the beam from the tissue increases the sensitivity of the spectroscopic process by effectively increasing L in Beer's law. The beam exiting the ATR plate thus has had a multitude of opportunities to interact with analyte molecules in intimate contact with the lower surface and contains the spectral absorption signature of the analyte. Increases in sensitivity of a factor of 20 can sometimes be obtained.

Raman spectroscopy is a special type of dispersive spectroscopy. It is based on Raman effect, named after the Indian physicist, C. V. Raman, who was awarded the 1930 Nobel prize for his discovery. In his seminal paper, published in the journal *Nature* in 1928, Raman observed that when light interacted with matter, some small fraction of its energy was scattered, and the frequency of the scattered light was shifted in a complex manner from that of the incident beam (Stokes and anti-Stokes scattering). Most of the scattered light energy is returned at the input wavelength, where this light undergoes elastic, Rayleigh scattering.

Since its discovery, the Raman effect has been developed into a sensitive, analytical chemical tool with results that rival conventional spectroscopic techniques (UV, VIS, IR) [Kaiser, 1998; Berger *et al*, 1999]. The advent of monochromatic laser sources has expedited the rise of Raman spectroscopy as an analytical tool. In the past few years, Raman spectroscopy has been applied to various biomedical applications, such as measuring various important analytes in blood and serum [Shafer-Peltier *et al*, 2002; Lambert *et al*, 2002; Koo *et al*, 1999; Frank, 1998]. Also, much attention has recently been given to developing non-invasive, biomedical Raman systems [Hanlon *et al*, 2000].

There are two kinds of Raman spectroscopy systems currently in use —Conventional Raman, and Stimulated Raman spectroscopy (SRS). Figure 7.83 illustrates a typical, conventional, *in vitro* Raman system. A high power diode laser emitting in the NIR, typically at 830 or 850 nm, is used as the source. Its beam is chopped or otherwise amplitude modulated to enable synchronous detection of the scattered light. NIR source light is generally used to minimize any natural, background fluorescence from a complex, biological sample. Note that the shorter visible wavelengths and UV light excite fluorescence. Most of the input light is elastically scattered from the sample (Rayleigh light) with no frequency shift by analyte molecules. Inelastic (Raman) scattering in all directions occurs at about 10^{-6} – 10^{-8} times the intensity of the Rayleigh scattered light. Energy from the monochromatic input beam is absorbed in the process of exciting vibrations of the various interatomic bonds of the analyte. The Raman scattered (output) light thus has a complex spectrum depending on the chemical structure of the analyte(s). In general, the Raman red shifts in the output spectrum have frequencies given by

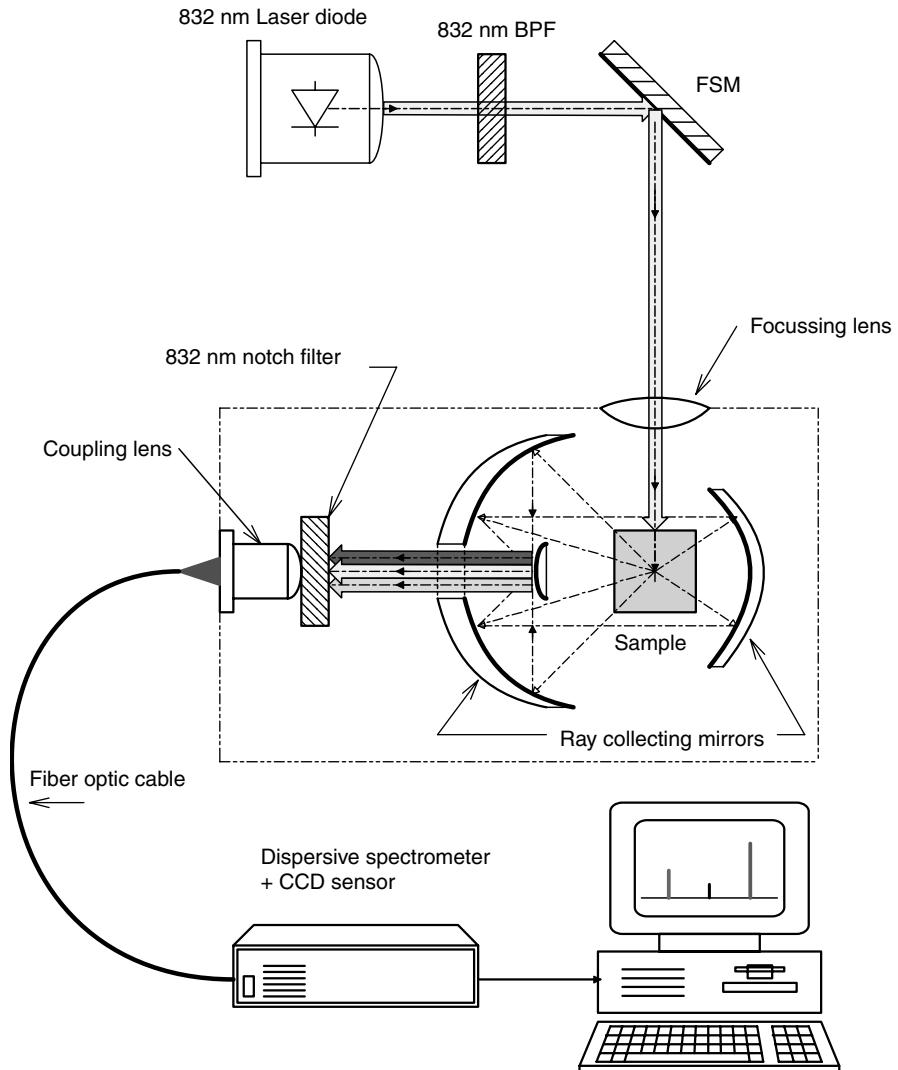
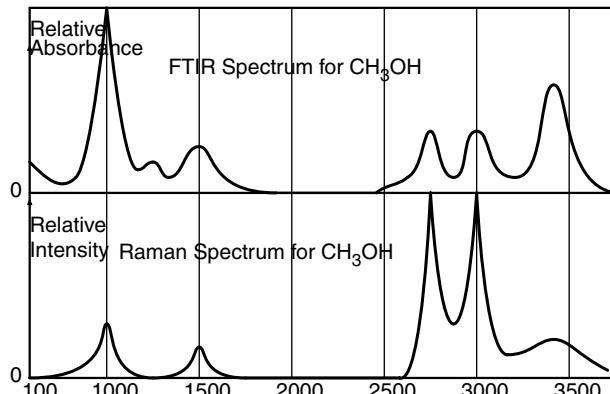


FIGURE 7.83
A conventional Raman spectroscope.

$(v_{in} - v_{Bk}) = \Delta v_{Sk}$ Hz. v_{in} is the input (excitation) lightwave frequency, v_{Bk} is the 'natural frequency' of the k th class of atomic bond. Δv_{Sk} is also called the Stokes shift of the k th bond class. The energy absorbed in the photon excitation of a particular, interatomic bond resonance is proportional to the Δv_{Sk} of its spectral peak. That is, $E_{abs} = h\Delta v_{Sk}$ J. A weaker emission is seen to occur for the anti-Stokes, or blue shift, Raman emission: $(v_{in} + v_{Bk}) = \Delta v_{aSk}$ Hz.

To effectively gather more of the weak, Raman scattered light, often a parabolic mirror is used near the sample. An optical notch filter is also used to attenuate the Rayleigh scattered light of frequency, v_{in} . A Hi-Q, holographic, optical notch filter attenuates by ca. 10^{-6} at the center of its stop band, and passes ca. 90% of the input light at other wavelengths. The weak Raman light is passed through a dispersive monochromator that either projects the desired Raman spectrum on a linear, CCD sensor array where the k th pixel intensity is proportional to Δv_{Sk} , or light at a given Δv_{Sk} is directed onto a single

**FIGURE 7.84**

A comparison of the absorbance spectra of methanol by FTIR and by Raman spectroscopy.

photomultiplier sensor or some photodiode. For *in vitro* Raman spectroscopy, relatively large input intensities are used, typically in the hundreds of mW. Such high input power cannot be used for *in vivo* Raman measurements because of potential heat damage to tissues, including the retina.

Raman spectra are generally displayed in units of wavenumber, $\xi \text{ cm}^{-1}$, which is proportional to the Hz frequency of the light. Note that 1 cm^{-1} equals $3 \times 10^{10} \text{ Hz}$, or 30 GHz. The frequency of 830 nm input light is $v_{\text{in}} = 3.61 \times 10^{14} \text{ Hz}$, and its wavenumber is $\xi_{\text{in}} = 1.203 \times 10^4$.

Figure 7.84 illustrates both Raman and FTIR absorbance spectra from methanol. Notice that in the two spectrograms, some congruent peaks are sharper and stronger than in the other spectrogram.

Stimulated Raman spectroscopy (SRS) is a technique that avoids the non-directional scattering of the frequency-shifted light emitted in conventional, single source Raman spectroscopy. In SRS, two lasers are used to excite the analyte—a pump laser and a tunable probe laser. The intersection of their beams defines an analyte volume. When the frequency (or wavenumber, $\Delta\xi$) difference between the two lasers equals the Δv_{Sk} of a Raman active mode, there is an increase in the irradiance of the transmitted probe beam and a corresponding decrease of the irradiance of the transmitted pump beam. In one SRS system designed to sense D-glucose in biological samples, the Δv between the pump and probe lasers is made to be the Δv_{Sk} of a major glucose Raman peak (i.e. $\xi = 518 \text{ cm}^{-1}$) [Tarr and Steffes, 1993; 1998]. (Recall that $v\lambda = c$ and $\xi \text{ cm}^{-1} = 3.333 \times 10^{-11} v$, v in Hz.) The actual frequencies of the pump and probe lasers are not as important as the Δv between them. The pump frequency should be in the NIR, if possible, to avoid exciting intrinsic fluorescence of the specimen. The advantage of the SRS technique is that no expensive monochromator is used. As shown in Figure 7.85, the probe laser's output is both wavelength tunable and chopped, either mechanically, or electronically by switching the laser diode current on and off. The beams from the probe and the pump lasers are combined, either using fiber optic mixing, or mirrors and a half-silvered mirror and then directed into the sample. Transmitted light from the pump laser is greatly attenuated by an optical notch filter, letting through the modulated probe laser beam, which is directed to a suitable photosensor. The sensor's output contains a modulated signal component consisting of the probe beam increased as a function of the Raman-Stokes energy absorbed by the analyte, as well as modulated components from sample background fluorescence, plus Raman emissions from the water solvent and from glass (cuvette, lenses, fiber optics). These latter components are artifacts with little Δv dependence and are easily subtracted from the desired signal following synchronous (lock-in) demodulation of the sensor output. Note that in SRS, the probe

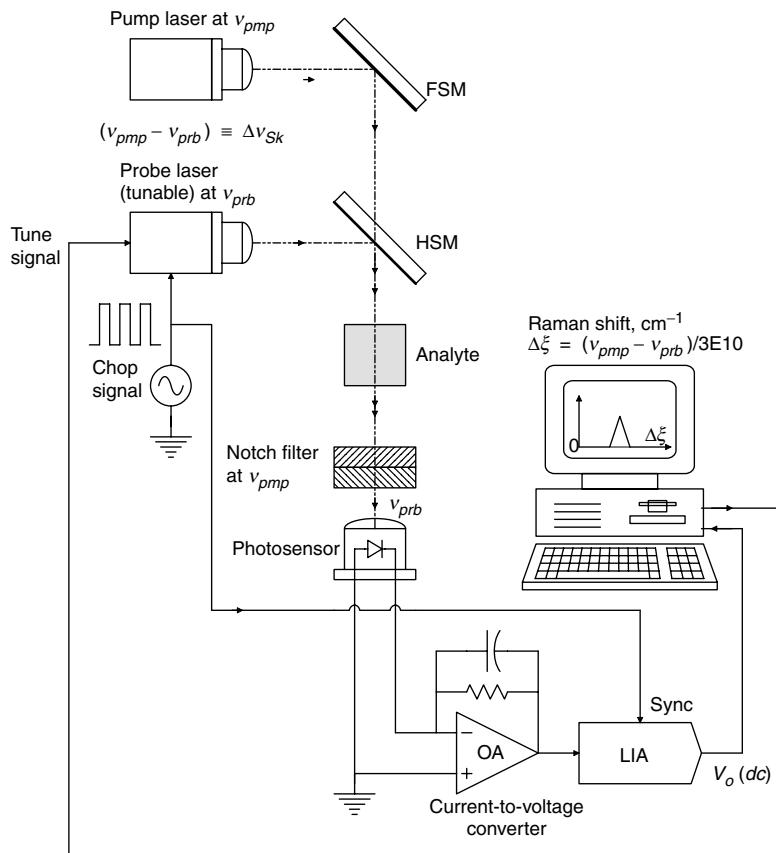
**FIGURE 7.85**

Diagram of a stimulated Raman spectroscopy system.

laser beam's frequency is not shifted, rather, its amplitude modulated intensity is changed.

Quoting Tarr and Steffes 1998:

“...The stimulated Raman effect requires no phase matching between the two laser beams, only that they be spatially overlapping in the Raman active medium. When the frequency difference of the two lasers equals the vibrational frequency of the Raman active mode (v_{Sk}), a nonlinear optical interaction occurs between the optical field and the medium. The strength of the nonlinear optical interaction depends upon the third-order nonlinear susceptibility, $\chi(\omega_s)$, of the medium. The result of this interaction is a coherent increase in the (transmitted) irradiance of the Stoke's (probe) beam and a corresponding decrease in the irradiance of the (transmitted) pump beam.”

Tarr and Steffes derive a theoretical model for the gain of the Stoke's beam.

We have seen that good *in vitro* results have been obtained with RS and SRS in measuring certain analytes in water, and blood plasma or serum [Northrop, 2002]. Mildly invasive procedures such as cystoscopy, bronchoscopy and colonoscopy may provide a means of using Raman spectroscopy to examine lesions (candidate tumors) on the surface of the epithelium in these locations.

For non-invasive SRS diagnostic applications, the pump or excitation laser's energy must be introduced transdermally and the very weak, backscattered, Raman shifted light

must be collected and analyzed dispersively. Stimulated Raman spectroscopy could be used transdermally by shining the pump and probe lasers through a thin vascular tissue such as a finger web or an earlobe. A serious requirement in all forms of *in vivo* Raman analysis is to be able to do it at laser power levels that will not damage the biological tissues being irradiated (by heating) [Hanlon *et al.*, 2000].

7.6.3 Non-Dispersive Spectroscopy

Non-dispersive spectroscopy (NDS) is a chemical analytical method that avoids using an expensive monochromator and a PMT to quantify a specific analyte. Instead of plotting $\%T(\lambda)$ or $A(\lambda)$ for a sample, and using the resulting peaks and valleys to quantify the analyte, an NDS instrument selects a narrow range of wavelength, $\Delta\lambda$, in which the analyte has a unique peak and valley in its $\%T(\lambda)$ curve. The $\Delta\lambda$ band is generated by passing broadband light through a bandpass filter made from one or more interference filters. Note that there are several possible configurations for a NDS system [Northrop, 2002]. We will consider one of them here. Figure 7.86 illustrates a basic, open loop, two-beam, NDS system that uses manual nulling. Light in the band, $\Delta\lambda$, is first chopped, then passed through a half-silvered, beamsplitter mirror. The direct beam

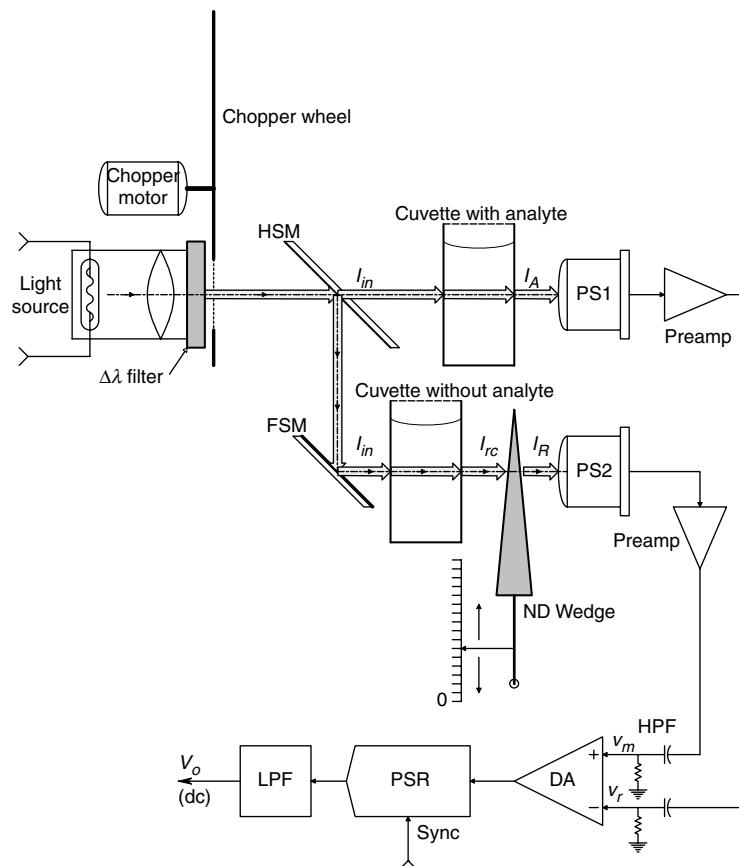


FIGURE 7.86

Diagram of a manually nulled, two-beam, non-dispersive spectroscopy system. The $\Delta\lambda$ filter is a bandpass interference filter that provides light in a $\lambda_1 - \lambda_u$ band.

passes through the sample cuvette, thence to a photosensor, PS₁. The reference beam passes through a blank cuvette filled with solvent, thence through an adjustable, calibrated, neutral density wedge (NDW), thence to photosensor PS₂. The outputs of the photosensors are high-pass filtered to remove dc, then amplified by a difference amplifier (DA). The DA output is a square wave at chopper frequency whose amplitude and phase (0° or 180°) is determined by the relative intensities at PS₁ and PS₂. In system calibration, with no analyte present in the sample cuvette, the NDW is manually set to zero (no extra attenuation in the reference beam). The dc output voltage, V_o , is nulled by adjusting the gain of either pre-amp so that $v_{mo} = v_{ro}$. With the analyte present, more light in the passband, $\Delta\lambda$, in the upper beam is absorbed and the total intensity at PS₁, I_A , is reduced. Now the NDW must be advanced to re-null V_o . The neutral density required to re-null the system can be shown to be proportional to the analyte concentration.

A further refinement of the NDS system is to make it self-nulling (Figure 7.87). In this design, the dc error voltage, V_o , is integrated to make a type 1 control system with zero

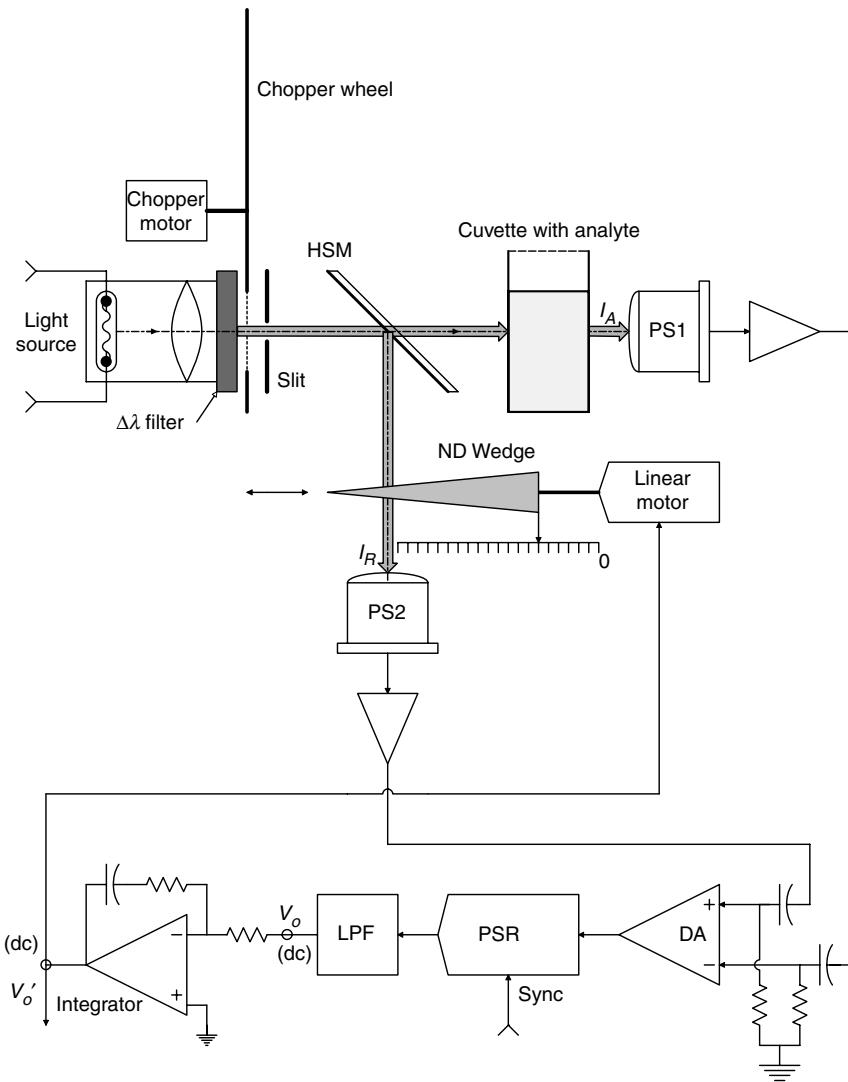


FIGURE 7.87

A servo-nulled NDS system.

steady state error, and the integrator output, V'_o , is conditioned to drive a linear positioning system that moves the NDW. Now, it can be shown that the difference in the voltage V'_o with no analyte, and V'_o with analyte is proportional to the (linear) change in ND, hence in the analyte concentration. We now demonstrate mathematically how an open loop NDS system works (Figure 7.86). Assume that the output of the optical bandpass filter is an ideal rectangular spectrum of width, $\Delta\lambda$. After being split into two beams by the half-silvered mirror (HSM), both source spectrums have width $\Delta\lambda$ and equal magnitudes, $S_1(\lambda)$ W/nm. The reflected reference spectrum is attenuated by the reference cuvette's reflections and the solvent's absorption. The reference cuvette has a transmission $T_{ref} = (I_{rc}/I_{in})$. $I_{in} = (S_1\Delta\lambda)$ W. I_{rc} next passes through the NDW, where it is further attenuated by its transmittance, T_{ND} . In general, $T_{ND} \equiv (I_{out}/I_{in}) = (I_R/I_{rc}) = 10^{-\beta x}$, where x is the displacement of the wedge from 0 on the linear x scale and β is the NDW's attenuation constant. (Note that $ND \equiv \beta x$; also that an ND of 1 means the input light is attenuated by a factor of 0.1, or $I_{out} = I_{in}/10$.) Thus the intensity at the reference photo sensor is:

$$I_R = (S_1\Delta\lambda)T_{ref}T_{ND} \text{ W} \quad (7.253)$$

The spectrum that passes through the HSM also passes through the walls of the sample cuvette twice (in and out) and also through the test solution with analyte. Thus, the measurement spectrum is also attenuated by a factor of $T_{ref} < 1$. We assume that in a narrow band, $\delta\lambda$, around λ_o , the analyte has a strong absorption. Hence the spectrum exiting the cuvette, $S_{1A}(\lambda)$, has a reduced total intensity as shown in Figure 7.88. Using Beer's law, the intensity of $S_{1A}(\lambda)$ is simply:

$$I_A = (S_1\Delta\lambda)T_{ref} - \delta\lambda S_1 T_{ref}(1 - T_A) \quad (7.254)$$

The transmittance minimum of the analyte, T_A , is given by Beer's law:

$$T_A = 10^{-kL[A]} \cong 1 - 2.303kL[A] \quad (7.255)$$

Here, we have assumed that $kL[A] \ll 1$. $[A]$ is the concentration of the analyte, and the value of k is both wavelength and concentration dependent. Thus, after some algebra, we can write the intensity of the light exiting the cuvette as:

$$I_A = (S_1\Delta\lambda)T_{ref}\{1 - 2.303(\delta\lambda/\Delta\lambda)kL[A]\} \text{ W} \quad (7.256)$$

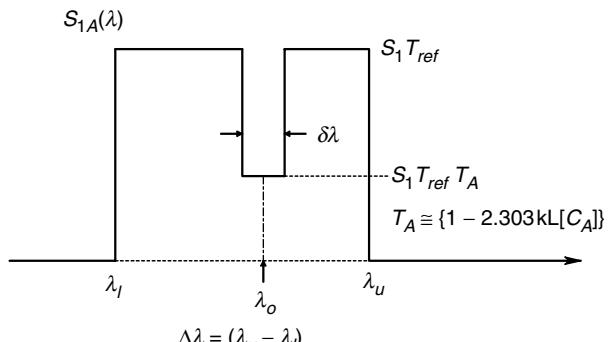


FIGURE 7.88

An idealized light intensity spectrum, $I_A(\lambda)$, from Figure 7.87. The notch represents photon energy absorbed by the analyte.

To null the system, the wedge is advanced until again, $I_R = I_A$ and $V_o = 0$. Thus at null we have:

$$(S_1 \Delta \lambda) T_{ref} T_{ND} = (S_1 \Delta \lambda) T_{ref} \{1 - 2.303(\delta \lambda / \Delta \lambda) kL[A]\} \quad (7.257)$$

So

$$T_{ND} = \{1 - 2.303(\delta \lambda / \Delta \lambda) kL[A]\} \quad (7.258)$$

Now we note that the analyte concentration is small, so we can use $T_{ND} = 10^{-\beta x} \cong 1 - 2.303 \beta x$. Thus:

$$\begin{aligned} (1 - 2.3 \beta x) &= \{1 - 2.3(\delta \lambda / \Delta \lambda) kL[A]\} \\ \downarrow \\ x &\cong (\delta \lambda / \Delta \lambda) kL[A] / \beta \end{aligned} \quad (7.259)$$

Equation 7.259 tells us that the displacement of the NDW by x to renull the NDS when an analyte is present is simply proportional to the analyte's concentration, [A], assuming that Beer's law holds and the NDW's ND is given by $10^{-\beta x}$, where $\beta x \ll 1$. The 2.3 factor comes from approximating $10^{-\beta x}$ by the use of $e^{-\varepsilon} \cong 1 - \varepsilon$. Note that $\log_{10}(a) = \ln(a)/2.303$.

To demonstrate the effectiveness of the NDS approach in quantifying an analyte, Fellows (1997) designed, built and tested a NDS to measure the % oxyhemoglobin, *in vitro*. Percutaneous measurement of the % O₂ saturation of RBC hemoglobin (Hb) is an established NI measurement technique. The pulse oximeter [Northrop, 2002] makes use of the differential absorption of red and near infrared (NIR) light by Hb vs HbO. Pulse oximeters use a red and an NIR light emitting diode (LED) (not a laser) as a 2λ source to make the measurement. Fellows' NDS system was developed as a proof-of-concept design, rather than a competing instrument for the very simple pulse oximeter. Fellows' system architecture closely followed the design shown in Figure 7.87, except instead of a linear, neutral density wedge, she used a rotating, fine mesh screen as a variable, ND attenuator. Unlike the linear wedge, the screen's transmittance is a function of its angle with respect to the beam being attenuated and can be shown to be given by:

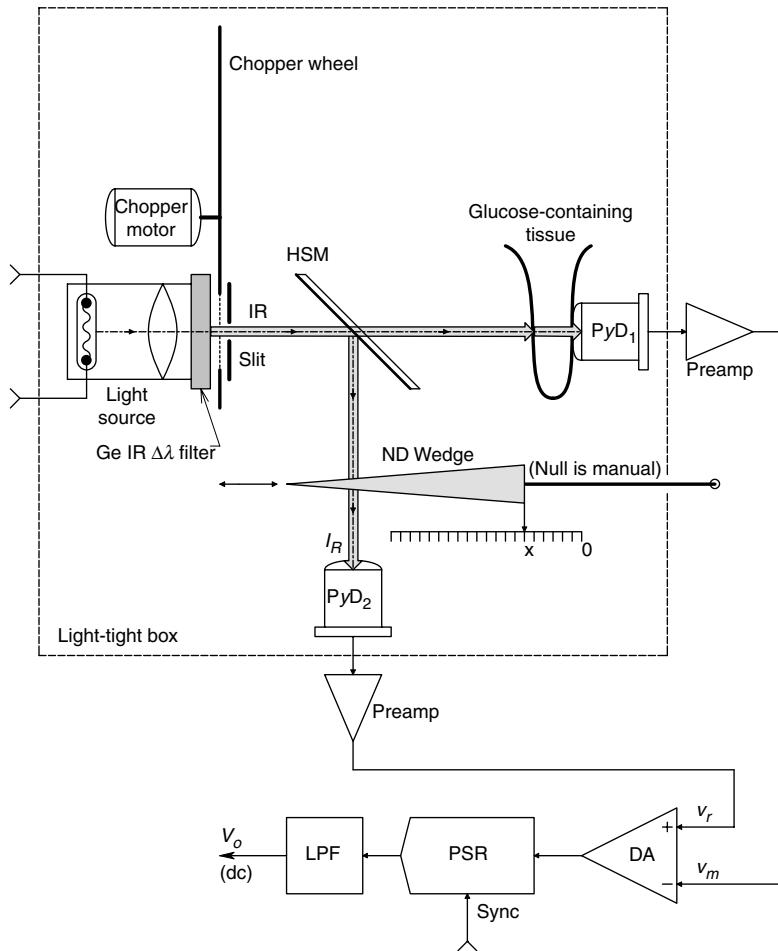
$$T_{SC}(\theta) = (1 - 2r/d)[1 - (2r/d) \sec(\theta)] \quad (7.260)$$

Each wire in the screen has radius, r , and d is the center-to-center spacing of the wires; a square mesh is assumed. The ND screen was rotated by a servo galvanometer whose angle, θ , was proportional to its dc input voltage.

Fellows used a filter that passed light in a 650–750 nm band. One cm cuvettes were used for the test solutions, the standard and the compensation. Sigma® freeze dried human hemoglobin was made up at a concentration of 12 g/L in water buffered to pH 7.4. To make either HbO or Hb, either O₂ gas or CO₂ gas was bubbled through the sample and reference cuvettes, respectively.

Fellows' servo-nulling NDS system was able to measure %HbO, but was difficult to calibrate because of the very nonlinear ND relation. A linear wedge or ND disk would have simplified the operation of her system.

A logical application of the NDS system would be to try to non-invasively and percutaneously measure (tissue) blood glucose in the FIR range. Glucose has three absorption peaks (%T nulls) in the FIR—one at 10.97 μm, one at 11.98 μm and one at

**FIGURE 7.89**

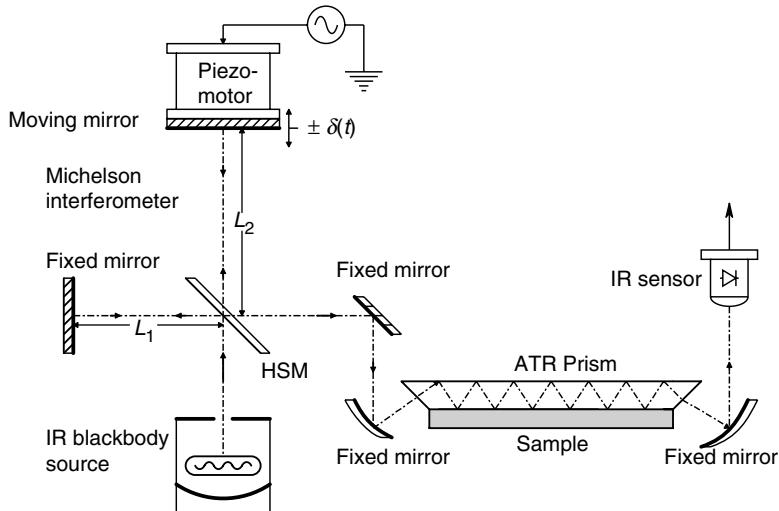
A prototype NDS system that might be able to measure blood glucose by IR absorption. The earlobe (shown) or a finger web might be used.

$12.95 \mu\text{m}$. If the LIR bandpass filter passed from 10.5 to $11.5 \mu\text{m}$, the peak at $10.97 \mu\text{m}$ could be used. Figure 7.89 illustrates a prototype, FIR, NDS system proposed by the author that might use a finger web or ear lobe (always with constant L). The FIR beam is defined by an IR bandpass filter, chopped, and the null is detected using a phase sensitive rectifier. A light tight box excludes stray IR from the pyroelectric photo-sensors (e.g. PVDF). FIR compatible HSM and NDW optical elements must be used. Such a system would be easy to build, but its accurate use would depend on individual calibration with a standard blood glucose test, and the absence of other tissue and blood substances that might overwhelm the glucose absorption peaks.

7.6.4 Fourier Transform IR Spectroscopy

A second very important category of non-dispersive spectrophotometer is the Fourier Transform Infrared (FTIR) System. Some advantages of FTIR spectroscopy are:

1. Better sensitivity and brightness; more light incident on the sample than dispersive IR spectrometers

**FIGURE 7.90**

Schematic of an FTIR spectrometer.

2. High wavenumber accuracy, up to 0.01 cm^{-1}
3. Takes less time to make a spectrogram
4. Relatively insensitive to stray light

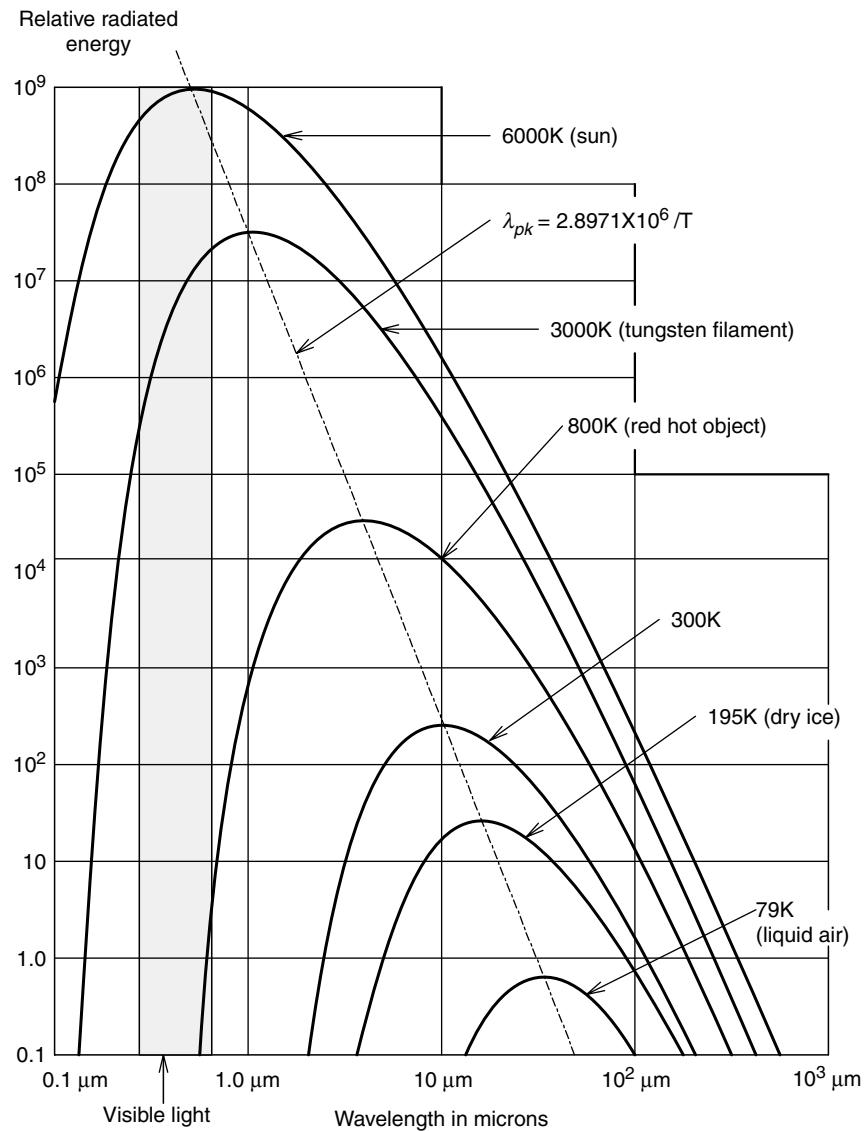
A disadvantage of FTIR spectroscopy, the same as in dispersive IR spectroscopy, is that both H_2O and CO_2 absorb strongly at certain wavenumbers, obscuring analyte peaks.

As shown in Figure 7.90, a blackbody IR source, such as a Globar rod, is used to generate a continuous, broadband, IR spectrum, given by Planck's radiation law (*cf.* Figure 7.91). Collimated radiation from this source is the input to a time modulated, Michelson Interferometer. The interferometer's mirror is periodically displaced by an amount $\delta(t)$, around a center position, L_2 . To understand what happens at the interferometer's output, we need to first consider what happens to a monochromatic input ray with wavelength, λ_1 . The distances L_1 and L_2 are chosen such that there is maximum constructive interference at the interferometer output. Thus, neglecting interfacial losses, the output intensity, I_o , equals the input intensity, I_{in} . Now let the mirror move $\delta = \lambda_1/4$ away from the beam splitter. The light must now travel a total distance of $\lambda_1/2\mu\text{m}$ more to return to the beam splitter, where there is now destructive interference and the output intensity has a null. This extra length the light must travel is called the retardation distance, ρ . In general, $\rho = 2\delta$. We can write an empirical expression for the output intensity as a function of the wave retardation distance, ρ , around L_2 :

$$I_o(\rho, \lambda_1) = I_{in} \left(\frac{1}{2} \right) [1 + \cos(2\pi\rho/\lambda_1)] \text{ W} \quad (7.261)$$

In terms of wavenumber, $\xi_1 \equiv \lambda_1^{-1}\text{cm}^{-1}$, I_o can also be written as:

$$I_o(\rho, \xi_1) = I_{in} \left(\frac{1}{2} \right) [1 + \cos(2\pi\xi_1\rho)] \text{ W} \quad (7.262)$$

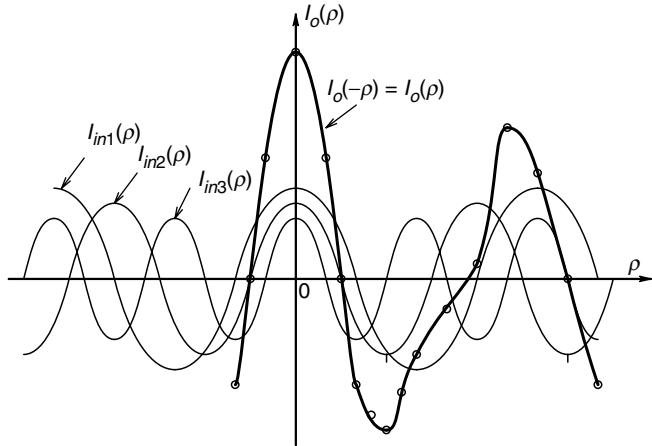
**FIGURE 7.91**

Ideal blackbody radiation spectra. Note how the peaks shift to longer wavelengths at lower object temperatures.

Note that $I_o(\rho, \xi_1)$ is an even function in $\rho\xi_1$. ρ must be in cm in equation 7.262. When two other beams having wavenumbers ξ_2 and ξ_3 are added to the beam of wavenumber ξ_1 , then the output of the time modulated Michelson interferometer can be written, using superposition, at a given retardation, ρ cm as:

$$I_o(\rho) = \sum_{k=1}^3 I_{ink} \left(\frac{1}{2} \right) [1 + \cos(2\pi\xi_k\rho)] W \quad (7.263)$$

When this even interferogram function is plotted vs ρ , we see a strong peak at the origin, surrounded by lower ripple peaks, as shown in Figure 7.92. In the more general

**FIGURE 7.92**

Superposition of the output intensity of the interferometer as a function of wavenumber and retardation distance, ρ , which is time modulated. Note that with just three wavenumbers a peak in the interferogram grow around $\rho = 0$.

case, when the source intensity input to the interferometer is described by a continuous density distribution in terms of wave number, $\xi \text{ cm}^{-1}$, the interferogram intensity is given by the inverse cosine Fourier transform:

$$I_o(\rho) = \int_{-\infty}^{\infty} W_B(\xi) \left(\frac{1}{2}\right) [1 + \cos(2\pi \xi \rho)] d\xi = I_{W_0} + \int_{-\infty}^{\infty} W_B(\xi) \left(\frac{1}{2}\right) \cos(2\pi \xi \rho) d\xi \quad (7.264)$$

where $W_B(\xi)$ is assumed to be even in ξ , so $W_B(\xi) = W_B(-\xi)$. Thus:

$$I_{W_0} = \int_0^{\infty} W_B(\xi) d\xi \quad (7.265)$$

The cosine integral can be considered to be the real, inverse Fourier transform of $W_B(\xi)$. That is:

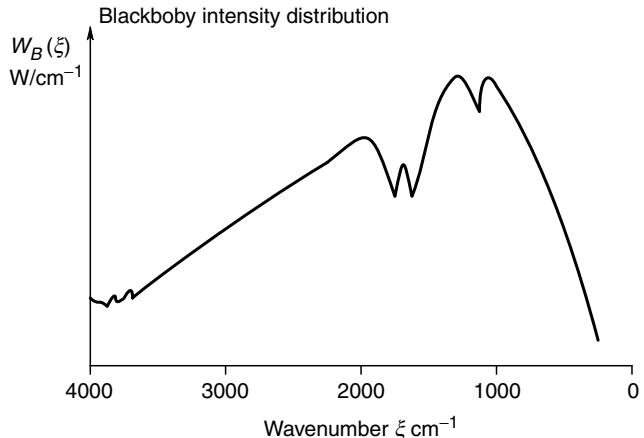
$$I_o(\rho) = \left(\frac{1}{2}\right) F^{-1}\{W_B(\xi)\} \quad (7.266)$$

Note that by the Euler relation, $e^{-j\omega t} \equiv \cos(\omega t) - j \sin(\omega t)$ and one definition of the continuous Fourier transform (CFT) is:

$$F(\omega) = R(\omega) + jX(\omega) \equiv \int_{-\infty}^{\infty} f(t) [\cos(\omega t) - j \sin(\omega t)] dt \quad (7.267)$$

When $f(t)$ is real and even, then it is well-known that:

$$F(\omega) \equiv \int_{-\infty}^{\infty} f(t) \cos(\omega t) dt + 0 \quad (7.268)$$

**FIGURE 7.93**

Plot of the blackbody source's intensity spectrum *vs* wavenumber, as determined by taking the Fourier transform of $I_o(\rho)$ given by equation 7.270.

and the continuous, inverse FT is:

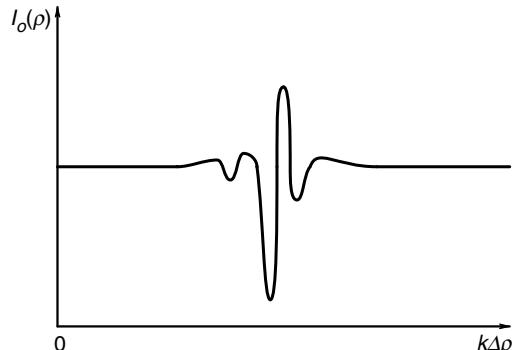
$$(t) \equiv (2\pi)^{-1} \int_{-\infty}^{\infty} F(\omega) \cos(\omega t) d\omega \quad (7.269)$$

where ω is in radians/sec and $X(\omega)=0$. Note that we have defined the blackbody radiation density, $W_B(\xi) \text{ W/cm}^{-1}$ as an even function in wavenumber, $\xi \text{ cm}^{-1}$. Let ξ be analogous to lightwave frequency in Hz and the retardation distance, ρ cm, be analogous to time in equation 7.268. Thus, $F(\omega/2\pi) \rightarrow W_B(\xi)$ and we see that modulation of the source light's spectral distribution, $W_B(\xi)$, by the Michelson interferometer generates an even, inverse Fourier (intensity) function of $W_B(\xi)$, $I_o(\rho)$, at its output, *sic*:

$$I_o(\rho) = \int_0^{\infty} W_B(\xi) \cos(2\pi \xi \rho) d\xi \quad (7.270)$$

Now if we sample the IFT, $I_o(\rho)$, at intervals $\Delta\rho$ and take its DFT, we recover an estimate of the input blackbody intensity distribution, $W_B(\xi)$, as shown in Figure 7.93. This $W_B(\xi)$ estimate is stored in the FTIR system's computer memory. Figure 7.94 illustrates a representative interferogram as a function of the retardation, ρ .

The next step in the operation of an FTIR spectrometer is to pass the interferometer modulated IR source radiation through the sample, where selective absorption

**FIGURE 7.94**

A typical, continuous interferogram output of the photosensor as a function of $k\Delta\rho$.

(attenuation) in certain wavenumber bands occurs. Again, as in equation 7.262, let us look at one particular wavenumber, ξ_1 . The intensity of the light emerging from the sample, $I_e(\rho, \xi_1)$ is given by:

$$I_e(\rho, \xi_1) = I_{in}(\xi_1)T_s(\xi_1)\left(\frac{1}{2}\right)[1 + \cos(2\pi \xi_1 \rho)] \quad (7.271)$$

where $T_s(\xi_1)$ is the wavenumber dependent ($0 \Leftrightarrow 1$) transmittance of the analyte and sample, and $I_{in}(\xi_1)$ is the input intensity at $\xi = \xi_1$. If the input is replaced by the BB distribution, $W_B(\xi)$, we can write:

$$I_e(\rho) = \left(\frac{1}{2}\right) \int_{-\infty}^{\infty} W_B(\xi)T_s(\xi)[1 + \cos(2\pi \xi \rho)] d\xi = I_{ewo} + \left(\frac{1}{2}\right) \int_{-\infty}^{\infty} W_B(\xi)T_s(\xi)\cos(2\pi \xi \rho) d\xi \quad (7.272)$$

Clearly, the 'dc' term is

$$I_{ewo} = \int_0^{\infty} W_B(\xi)T_s(\xi)d\xi \quad (7.273)$$

and the integral is the inverse FT of the frequency domain product, $\{W_B(\xi)T_s(\xi)\}$. When the output interferogram, $I_e(\rho)$, is DFTd at intervals, $\Delta\rho$, we obtain an estimate of the output spectrum:

$$S_{Sa}(\xi) = \left(\frac{1}{2}\right)W_B(\xi)T_s(\xi) \quad (7.274)$$

Since we already know $W_B(\xi)$, the desired sample transmittance is easily found by the relation:

$$T_s(\xi) = \frac{2S_{Sa}(\xi)}{W_B(\xi)} \quad (7.275)$$

It goes without saying that a high speed computer capable of calculating DFTs and generating plots of $T_s(\xi)$ and absorbance $A(\xi)$ is a necessary component of an FTIR spectrometer.

We have seen that all spectrometers pass EMR through a cuvette holding the analyte in solution, or use an ATR plate to interface light with the analyte. In some designs, the incident EMR is broadband ('white'), and the emergent radiation from the sample is passed through a monochromator to examine the transmittance or absorbance at particular wavelengths. In other designs, monochromatic light is first passed through the sample. Spectrometers operating in the FIR require the use of special, IR pass materials for refraction and cuvettes, including ZnSe (0.6–17 μm), ZnS (0.4–12 μm), Ge (2–14 μm), Si (1.2–8 μm), KRS-5 (0.6–50 μm), As₂S₃ (0.8–10 μm) and certain plastics. In nearly all cases, the light beam is chopped so that phase sensitive demodulation can be used (equivalent to a lock-in amplifier) to improve sensitivity and reject noise and stray (dc) light.

Non-dispersive spectrometers use narrow band light, two or more LEDs or laser diodes to avoid the need for an expensive monochromator. FTIR spectrometers also fall

into the class of non-dispersive instruments, as do Stimulated Raman spectrometers. In the FTIR spectrometer, the modulation is provided by the Michelson interferometer.

All sorts of medically important molecules from body fluids can be quantified by IR spectrophotometry. These include, but are not limited to, cholesterol, steroid hormones, thyroxine, theophylline, opiates, opioids, tranquilizers, etc. Knowledge of a drug concentration in blood or urine can be used clinically to adjust the dosage and also to detect drug abuse. Raman spectroscopy is showing promise in the non-invasive diagnosis of certain superficial tumors including skin melanomas, lung cancer, prostate and bladder cancer, etc., in any place an endoscope can reach [Northrop, 2002].

7.6.5 Single and Multiple Bubble Sonoluminescence

In this section, we will examine the phenomena of single bubble sonoluminescence (SBSL) and multiple bubble sonoluminescence (MBSL), caused by cavitation in liquid due to an intense ultrasound field, and consider their potential as chemical analytical tools with possible application to the detection of airborne molecules of high explosive substances such as C4, Semtex and dynamite.

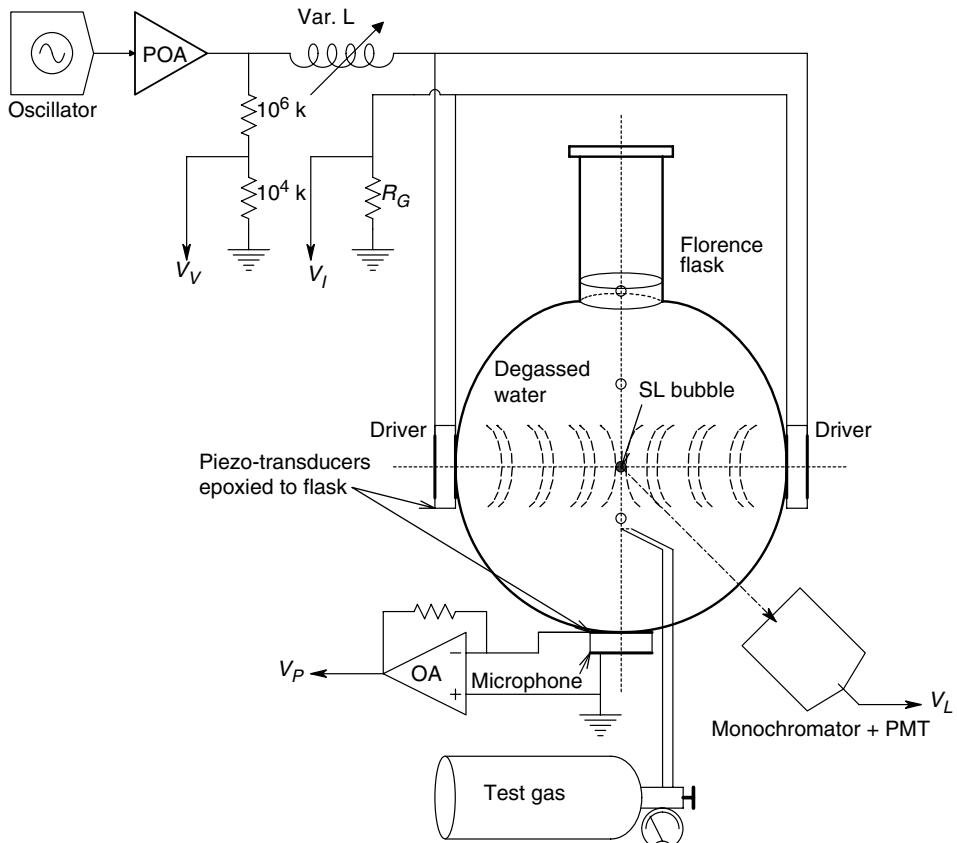
7.6.5.1 Single Bubble Sonoluminescence

Figure 7.95 illustrates the schematic of a basic, easy-to-construct, experimental, SBSL apparatus [Hiller and Barber, 1995]. A 100 ml, round, Florence flask is used as a reaction vessel. Three piezo ceramic disk transducers are used. Electrical connections to the disk transducers should be made using silver epoxy—soldering the leads to the thin silver films on both sides of the transducers can destroy (vaporize) the films if not done carefully.

The transducers are epoxied to the flask in the positions shown (the flask is suspended from its neck by a clamp, not shown). The two larger transducers (*ca.* 2 cm diameter) are the ultrasound drivers, the bottom, smaller (*ca.* 5 mm diameter) transducer is used as a microphone to sense sound in the flask and shock waves from the periodic, rapid collapse of the bubble in the sound field. The microphone signal is conditioned by a simple op-amp circuit and its output, V_P , is sent to an oscilloscope.

A sine wave generator is connected to a power amplifier, whose 32Ω output is connected to a variable inductor having an appropriate current rating. The inductor is connected to the two driver transducers in parallel. Below the transducers' self-resonance frequency (e.g. 1 MHz), they appear electrically like leaky capacitors because of acoustic loading by the flask. Their combined capacitance is of the order of 1 nF. It is found experimentally that the acoustic resonance frequency, f_A , for the water filled flask is *ca.* 20–30 kHz. At f_A , the maximum acoustic energy is coupled to the flask. An easy way to observe f_A is to monitor V_P as the frequency of the function generator is varied with the inductor shorted out. At f_A , V_P is maximum. To produce maximum driving power to the transducers, they must be electrically resonated by adjusting the variable inductance. L is varied at constant f_A until zero phase shift is seen between V_V and V_I . R_G is made 1Ω , so V_I equals the current into the transducers. Note that at electrical resonance, the power amplifier sees a real impedance of $(R_G + R_X)$, where R_X is the transducers' equivalent acoustic load into the flask. Small test gas bubbles are introduced into the flask by a thin glass tube connected to a gas supply.

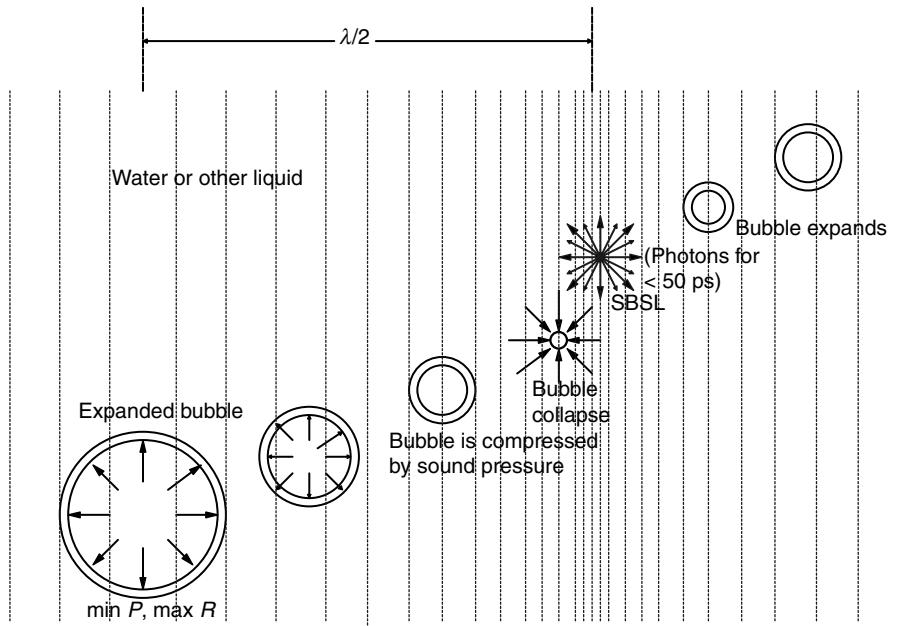
SBSL is an interesting physical phenomenon, first described in literature about 1990 [Lohse, 2002; Crum, 1994]. If a single small bubble of gas in water, having a resting diameter of *ca.* 10 μm , is trapped in an ultrasound field of appropriate intensity, the bubble

**FIGURE 7.95**

The single bubble sonoluminescence system of Hiller and Barber (1995).

undergoes expansion and compression at the ultrasound frequency. When the bubble is compressed by the sinusoidal sound pressure waves, at a critical pressure, its volume abruptly collapses at a speed of about Mach 4 to about $1\text{ }\mu\text{m}$ diameter. Its maximum radius at minimum sound pressure is *ca.* $100\text{ }\mu\text{m}$. Thus, the volume compression ratio is an enormous 10^6 [Moss *et al*, 1997a and b]. This abrupt compression causes quasi adiabatic heating of the gas volume, causing the internal temperature to rise to an estimated peak of 10^4 – 10^6°C [Crum and Matula, 1997]. A plasma is created inside the collapsed bubble which emits a pulse of photons, *ca.* 200 ps in duration [Lohse, 2002] (Figure 7.96) This flash is probably due to trace Argon gas in the bubble. Even though the individual flashes are extremely short, the fact that they occur at the ultrasound frequency, f_A , (generally 20–30 kHz) causes enough average light flux so that the sonoluminescence is visible as a blue point of light to dark-adapted human eyes.

At the very high temperature and pressure generated when the bubble collapses, a photon emitting plasma forms. The primary photon emission mechanism appears to be electronion or electronatom bremsstrahlung [Suslick *et al*, 1999]. At collapse, various chemical reactions between the nitrogen, oxygen and other trace molecules in the bubble are postulated to occur. These reaction products are gradually absorbed (rectified) by the water, leaving a bubble containing a high concentration of chemically inert argon from the air. The argon gas sonoluminesces. Hiller *et al* (1994) showed that SL does not occur with pure N_2 , O_2 and $\text{N}_2 + \text{O}_2$, Ar (or some other noble gas) is required. On the

**FIGURE 7.96**

Steps illustrating the collapse, sonoluminescence, and re-expansion of a gas bubble in the ultrasound field in water.

other hand, dissolved gases in the water can enter the bubble during its expansion cycle in the aqueous sound wave field.

The emission spectrum of SBSL is largely blackbody, theoretically extending into the far UV because of the high peak plasma temperature. Unfortunately, water and the outer plasma envelope itself absorb the shorter wavelength UV emissions. There are no discernable emission lines from SBSL in water [Suslick *et al*, 1999; Didenko *et al*, 2000].

Didenko *et al* speculated that the lineless SBSL emission in water is:

1. Due to the summation of broadened emissions from multiple molecular and atomic sources
2. Due to bremsstrahlung emission from partial ionization of the heated gas within the bubble

Didenko *et al* studied the SBSL emission spectra from a variety of polar aprotic liquids, including formamide, *N*-methylformamide, *N,N*-dimethylformamide, 1,2-diaminoethane, dimethylsulfoxide and adiponitrile. They subdivided their results into emissions from bubbles held stationary in the sound field (S-SBSL) and bubbles that move around the reaction flask's center and emit M-SBSL. The spectra from the moving bubbles in methylformamide and adiponitrile showed a broad emission line at *ca.* 380 nm that was from —CN, dissociated from the liquid molecules. When DMSO was used, there obviously was no —CN line to dissociate (no N atom in the molecule), but there were also no band emission lines from S₂, —SO or SO₂, which were expected. There were no lines from liquids with stationary bubbles, presumably because the intense blackbody bremsstrahlung radiation intensity swamped them.

Yasui (1998), in a modeling study, examined the effect of the surfactant Triton X-100 on the dynamics of SBSL. Yasui concluded that the dissolved surfactant lowered the

magnitude of the light pulse because it inhibited condensation of water vapor inside the bubble during collapse. Free water vapor lowered the temperature inside the bubble at collapse due to the larger amount of water vapor that undergoes endothermal chemical reactions at collapse.

Yasui states:

"It is predicted, based on the hot-spot model, that the radiation is not thermalized inside a bubble in the case of SBSL in a solution of the surfactant in water and that the [light emission] spectrum of SBSL may deviate from the blackbody spectrum and may have some characteristic lines such as the —OH line (310 nm)."

From the work of Yasui and Didenko *et al*, we can see that if the severe thermodynamic conditions (peak temperature, pressure) inside stationary collapsing bubbles are mitigated, reducing blackbody bremsstrahlung radiation, it may be possible to see individual spectral lines from excited state emissions during sonoluminescence.

7.6.5.2 Multiple Bubble Sonoluminescence

Multiple bubble sonoluminescence (MBSL) is also a phenomenon dependent on ultrasonic energy in a liquid. MBSL occurs as the result of sonic cavitation at very high sound pressure levels. Quoting Suslick *et al* (1999):

"If the acoustic pressure amplitude of a propagating acoustic wave is relatively large (greater than (≈ 0.1 MPa), local inhomogeneities in the liquid can serve as nucleation sites for rapid inertial growth into a cavity of macroscopic dimensions, primarily filled with vapor and dissolved gases. Such a bubble is inherently unstable, and its subsequent collapse can result in an enormous concentration of energy. ... A normal consequence of this unstable growth and subsequent collapse is that the cavitation bubble itself is destroyed. Gas filled remnants from the collapse, however, may serve as nuclei for continuation of the process.

Single bubbles, while creating extreme conditions, simply do not contain sufficient material to be useful for driving chemical reactions in any practical amount. In multi bubble systems ('cavitation clouds'), on the other hand, the interactions between bubbles as they collapse will lead to both substantial asymmetry and the formation of jets during the collapse, which one may well expect to limit the collapse efficiency. Thus, the conditions created during multi bubble cavitation, which is used of necessity for all sonochemical reactions, will be somewhat less extreme than those created during single bubble sonoluminescence."

MBSL creates its own bubbles by cavitation, an external gas source is not needed to form bubbles. Thus, any gas phase reactants must be dissolved in the host liquid if they are to be detected by their spectral signature during MBSL. MBSL hot spots have been measured at *ca.* 5050 K, significantly lower than the peak temperatures reached in SBSL [Suslick *et al*, 1999]. The lower temperatures of the MBSL plasma means far less bremsstrahlung radiation, allowing spectral lines to be seen from the sonochemical reactions occurring inside the cavitation bubbles.

Suslick *et al* (1999) illustrate sharp MBSL line spectra from non-aqueous silicone oil containing metal carbonyl compounds such as $\text{Fe}(\text{CO})_5$, $\text{Cr}(\text{CO})_6$, $\text{W}(\text{CO})_6$ and $\text{Mo}(\text{CO})_6$. The MBSL emission lines for these compounds were almost identical to the emission lines when the compound was burned in a flame at *ca.* 5000 K. MBSL in pure water has been shown to give a strong, broad emission line peak at 310 nm from an excited state of —OH. When metal salts are dissolved in water or alcohol, the MBSL spectra show the metal line spectra.

7.6.5.3 Summary

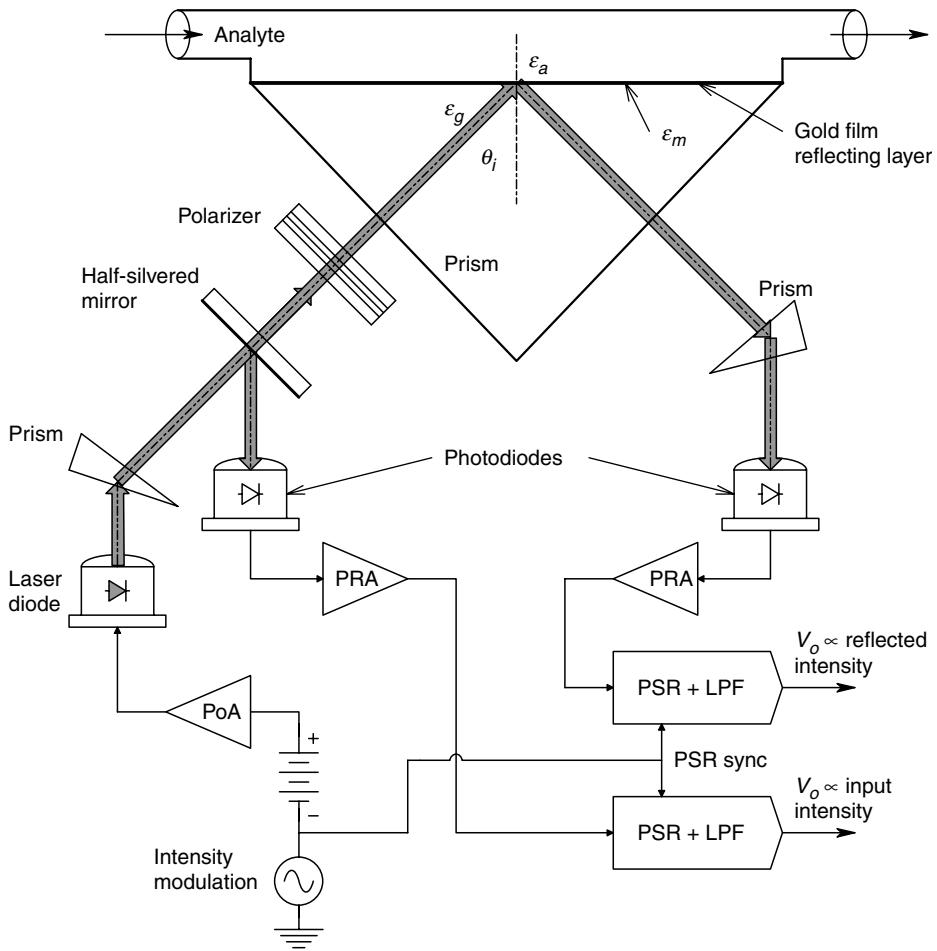
In summary, it appears that MBSL is potentially more useful as an analytical tool than SBSL. Spectral emission lines are lost in SBSL because of the very high plasma temperatures. On the other hand, MBSL is capable of revealing the emission lines of compounds and metal ions in solution. The question remains about its detection sensitivity for organic, nitrogen containing explosive vapors. If MBSL were to be used to detect explosive vapors, the liquid would have to be saturated with a very small concentration of the vapor in air by bubbling the vapor bearing air through the liquid for a long time. Contrast this procedure with concentrating the analyte by passing air through a cold trap, then using a gas chromatograph/mass spectrometer to identify and quantify the explosive vapor. The latter method is accurate but expensive. MBSL, if it works, would be far less expensive and portable. It is certainly worthwhile to pursue MBSL as a means to detect trace vapors from explosives.

7.6.6 Surface Plasmon Resonance

Surface plasmon resonance (SPR) sensors are a relatively new chemical analytical tool. They allow rapid, specific determination of the concentration of a variety of medically, biologically and environmentally important analytes. For example, specific bacteria, antibodies, theophylline, caffeine, NO₂, pesticides, explosives, controlled substances (opioid drugs), etc., have been sensed with SPR. Threshold sensitivities for certain analytes have been reported as low as 0.05 ppb (10^{-11}). SPR technology is relatively simple and inexpensive to implement, compared with analytical systems such as HPLC, mass spectrometry, IR spectrometry and gas chromatography, and it lends itself well to field measurements. Since this text is about non-invasive medical instruments, the analytes are presumably derived from urine or saliva, or smears from mucous membranes, etc. There is no reason, however, why SPR sensors cannot be used on gases, or blood, to look for any analyte (chemical, antibody, bacteria, etc.) therein.

SPR is a quantum phenomenon which occurs when a beam of monochromatic, linear polarized light (LPL) is reflected off a thin metal film, vapor deposited on one side of a glass prism, or when a beam of LPL is incident at a critical angle on a gold-coated diffraction grating. For illustrative purposes, we will examine in detail the prism SPR system using the so-called Kretschmann geometry. This system is shown schematically in Figure 7.97. A thin (50 nm) film of conducting metal such as gold or silver is vapor deposited on one face of a prism, or on the flat face of a half-round rod. A beam of LPL of a known wavelength is directed into the prism or rod so that it strikes the gold film face at an angle of incidence, θ_i . The incident beam's E vector must lie in the plane of incidence (be in the TM mode) for SPR to occur. In intimate contact with the other side of the gold film is a thin film of the analyte, having a permittivity, ϵ_a . It has been experimentally observed that when the incoming beam's wavelength and angle of incidence have unique critical values, the intensity of the reflected beam reaches a minimum. The depth of the null in the output beam intensity is a function of how much energy from the input LPL beam is coupled into the generation of surface plasmons in the metal film. The degree of coupling is a function of λ , θ_i , and most importantly, the dielectric constant of the analyte material.

Surface plasmons can be thought of as induced, wave-like fluctuations in the density of the conduction band electrons in the thin metal film. These fluctuations exist in both space and time (i.e. they can be thought of as traveling waves induced by the incident TM, LPL). The basis for using SPR as an analytical chemical tool is based on the conditions required for resonance. Under conditions of non-SP resonance, the incident

**FIGURE 7.97**

A surface plasmon resonance system using the Kretschmann geometry.

beam of LPL reflects off the metallized surface of the prism and exits the prism to a photo sensor. The metal film effectively forms a conducting mirror surface, where conventional refraction and reflection optical laws apply.

The velocity of light *in vacuo* is $c = 2.998 \times 10^8$ m/s. In transparent liquids or glass, light travels more slowly. The ratio of the speed of light *in vacuo* to the speed of light in the medium is defined as the refractive index of that medium, n_m . That is, $n_m \equiv c/v_m$. From electromagnetic (EM) theory, $c \equiv 1/\sqrt{\epsilon_0\mu_0}$ and $v_m = 1/\sqrt{\kappa_m\epsilon_0\mu_m}$. However, in non-magnetic materials, the magnetic permeability $\mu_m \approx \mu_0$. Thus we can write the refractive index as:

$$n_m \approx \sqrt{\kappa_m} \quad (7.276)$$

v_m (and n) are generally functions of frequency. An EM wave in a medium or free space such as light, or surface plasmon waves which basically exist in two dimensions, can be described by their wave vector, \mathbf{k} . \mathbf{k} is directed along the direction of wave propagation, and its magnitude in vacuo for EM waves is:

$$k_0 = \omega/c = 2\pi\nu/c = 2\pi/\lambda \quad (7.277)$$

In the prism glass with refractive index, n_1 , the magnitude of the EM wave vector is [Krauss, 1953]:

$$k_1 = \omega/v_1 = \omega\sqrt{\kappa_1/c} \quad (7.278)$$

Surface plasmons are generated on the metal film under the condition that the magnitude of the wave vector in the glass incident on the gold film equals the magnitude of the wave vector at the metal/analyte interface. This can be written as:

$$k_i = (2\pi/\lambda) \sin(\theta_i) \sqrt{\kappa_1} = (2\pi/\lambda) \sqrt{(\kappa_a |\mathbf{k}_m| / (\kappa_a + |\mathbf{k}_m|))} \quad (7.279)$$

where κ_1 is the dielectric constant of the glass prism, κ_a is the dielectric constant of the analyte, and $\mathbf{k}_m = \kappa'_m + j\kappa''_m$ is the complex dielectric constant of the gold film. All dielectric constants are generally functions of v (or λ). Canceling like terms and solving for the angle of incidence inside the prism, we have the angle criterion for SPR:

$$\theta_{iR} \cong \sin^{-1}\{\sqrt{(\kappa_a \kappa_m)/[(\kappa_a + \kappa_m) + \kappa_1]}\} \quad (7.280)$$

The resonance condition is due to momentum matching of incident photons with plasmons in the metal. The fact that the permittivity of the analyte layer (typically 250 nm thick) on the other side of the metal affects SPR, may be due to the evanescent field expanding through the metal and coupling into SPs at the analyte surface. The resonance angle, θ_{iR} , is exquisitely sensitive to the dielectric constants of the metal and the analyte in contact with it. Thus any chemical reaction that takes place at the metal surface, such as binding of antibodies to metal bound antigen, will affect ϵ_a and the value of θ_{iR} . Thus, surface reactions of analyte can be used to sense antibodies, or if the antibodies are bound to the metal, it can sense antigen molecules such as those on bacterial and viral surfaces, or analyte molecules in solution or suspension.

Figure 7.98 illustrates the typical SPR curves for a Kretschmann prism system receiving monochromatic light. Note that when the analyte index of refraction increases due to

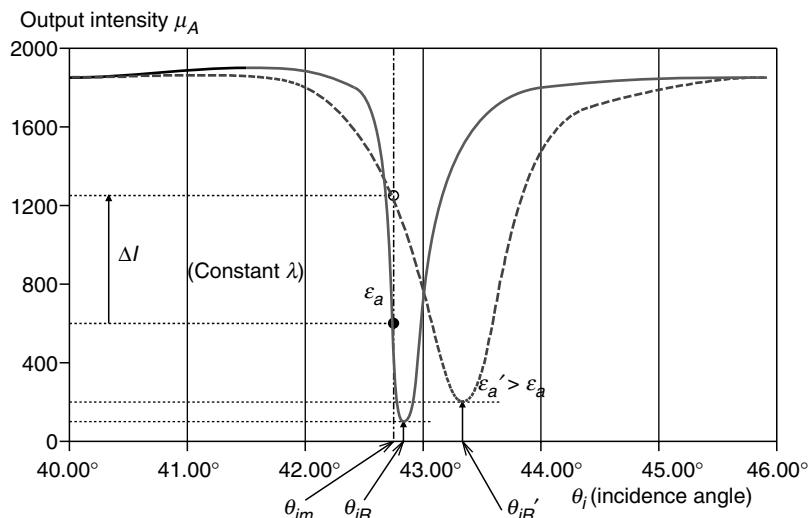
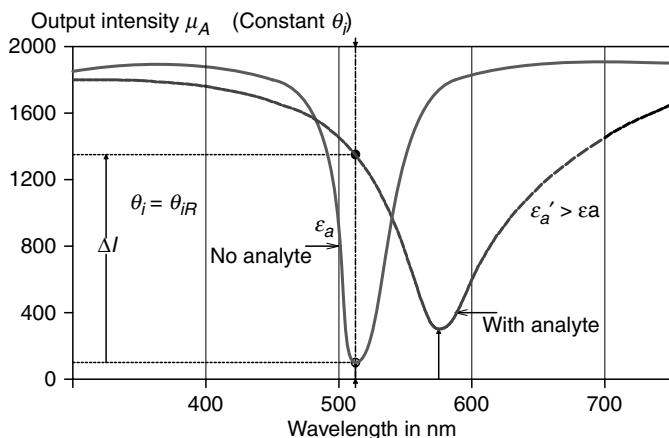


FIGURE 7.98

Reflected light intensity as a function of incidence angle for a Kretschmann SPR system. The solid curve is the system response in the absence of analyte; the dashed curve is obtained with the analyte in intimate molecular contact with the gold film. Monochromatic light is used. Note the peak shift.

**FIGURE 7.99**

An alternate mode of operation of the Kretschmann SPR system is to hold the incidence angle constant and vary the wavelength. Again there is a peak shift with analyte present (dashed curve). At a fixed wavelength (e.g. 512 nm), the difference between output intensity for no analyte (solid curve) and for analyte is an increasing function of analyte concentration.

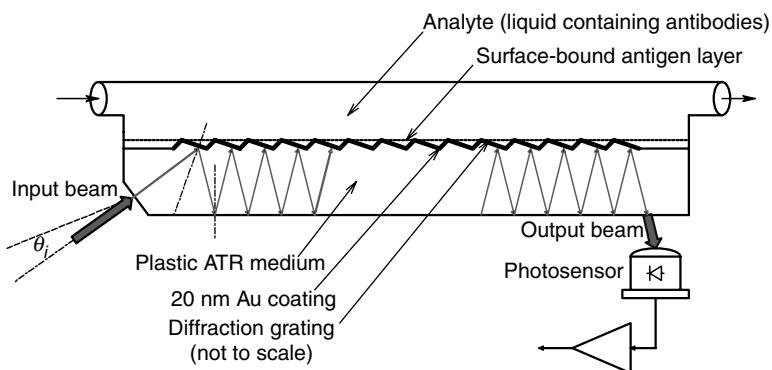
antibody bonding at the metal analyte surface, κ_a also increases to κ'_a and the SPR curve, as a function of incidence angle, shifts to the light and broadens. (Note that $n_a \cong \sqrt{\kappa_a}$.) Thus, the intensity measured at angle θ_{im} , increases, while the intensity minimum moves to θ_{iR}' . Both shifts, either taken together in a formula, or as separate phenomena, can be used to quantify the extent of the binding reaction at the metal surface. The depth of the null at SPR depends in part on the thickness of the metal film [Foster *et al*, 1994].

Another way of using the Kretschmann system is to set the incidence angle at the SPR null for some standard source, λ . The source λ is then varied by a monochromator and the output intensity is plotted as a function of λ , as a surface reaction takes place. Typical intensity $\text{vs } \lambda$ curves are illustrated in Figure 7.99. Since the permittivities of the glass, metal film and analyte are functions of λ , we again see a shifting and broadening of the SPR intensity curves with λ as antibody binding at the metal surface takes place. While varying the source λ can yield good analytical results, it makes the SPR system more expensive because of the need for a precision monochromator.

The reader should be aware that the design of SPR devices for chemical analysis is a rapidly growing field. One alternative configuration of SPR sensor places the analyte solution over the surface of a plastic diffraction grating whose surface has been vapor deposited with gold, silver or aluminum (gold is generally preferred). The plastic top of the grating acts as an attenuated total reflection (ATR) prism, in which light reflected from the grating where the beam strikes it initially is reflected back many times to the grating surface. Such a grating SPR design was proposed by Simon (1998). Figure 7.100 shows a side view of Simon's 'long range' SPR grating system. SPR occurs at a critical beam input angle, θ_i , giving a minimum of output light intensity when the coupling component of the monochromatic, TM-polarized, input beam wave vector satisfies the relation:

$$(2\pi/\lambda) \sin(\theta_i) + (2\pi/b) = (2\pi/\lambda) \sqrt{(\kappa_a \kappa_m)/(\kappa_a + \kappa_m)} \quad (7.281)$$

Here, b is the grating constant, κ_a is the dielectric constant of the analyte and κ_m is the magnitude of the complex dielectric constant of the metal film. A very thin layer of antigen with high affinity to the antibody to be detected is chemically bound to the 20 nm gold film over the grating surface. The antibodies in the analyte liquid bind to the protruding antigen molecules, thus changing κ_a at the gold-solution interface. This change of dielectric constant changes the interfacial refractive index and 'retunes' the SPR to a new input angle, θ_i . Either by changing θ_i to re-attain a minimal light output,

**FIGURE 7.100**

An SPR system can be made more sensitive by using a metal film covered diffraction grating cut into one face of an ATR plate. The metal film is coated with a thin layer of molecules that have affinity to bind with the analyte. SPR systems have been operated with both liquid and gas phase analytes. To make the SPR system an effective analytical instrument, it must be possible to remove all bound analyte between readings.

or by measuring the increase in light output due to $\Delta\kappa_a$, the number or density of bound antibodies can be quantified. A possible disadvantage of Simon's grating SPR configuration is the formation of bubbles on the grating surface and the fact that the measurement is a 'one-shot' event; a new, virgin, antigen surface must be reapplied to the gold film, and the output nulled again before another antibody assay can be done. Simon (1998) claims that the output signal from the ATR layer increases approximately five-fold when about 2 nm of antigen-antibody complex forms at the gold surface. Also, a refractive index change in the analyte of less than 1% causes a 0.5° shift in θ_i to re-null the system.

Other embodiments of the grating SPR system place the analyte over the gold film covered grating—the monochromatic, TM-polarized light is directed through a thin layer of analyte onto the grating. Jory *et al* (1995) reported on an exquisitely sensitive grating SPR system in which an acousto-optical tunable filter (AOTF) element was used to control the wavelength of the incident beam to a precision of 0.0005 nm. They used their system to measure the concentration of NO₂ gas in N₂ (a gas phase analyte). By depositing a thin layer of phthalocyanine over the old coating, a wavelength shift of -0.004 nm renulled the system from zero concentration when 0.01 ppm NO₂ in N₂ was applied. They claimed that the sensitivity of their system allowed detection of changes in the refractive index of the gas of 1×10^{-6} . The fact that very small concentrations of gas are detectable suggests that the grating SPR technology might be developed in the future to detect medical gas concentrations, or even trace gases emanating from controlled drugs and explosives. (At present, trained dogs have the best record for detecting explosives and controlled drugs.)

The potential application of SPR sensors in medical diagnosis is enormous and is just beginning to be realized [Geddes and Lawrence, 1997]. As we have seen, both gas and liquid phase sensing are possible. Practically any analyte that can react with a reactant bound to the gold film surface on a prism or grating, with a strong affinity, can be sensed by SPR. The reaction must cause a change in the refractive index or the permittivity at the gold surface in order to affect the SPR conditions (λ and/or θ_i). SPR antigen-antibody reactions can be used to sense specific antibodies, bacteria, viruses, proteins, hormones, cytokines, etc. Note that specific, monoclonal antibodies can also be bound to the gold film in order to sense any protein or molecule for which they can be made specific.

However, the problem of quick, efficient, SPR sensor regeneration remains to be solved. Once the bound surface reactant has combined with the analyte, the analyte must be totally removed before the next measurement, without affecting the bound reactant, or else, the complex must be removed and the surface reactant layer must be renewed or rejuvenated.

7.7 Temperature Measurements

Many physical and chemical phenomena and physical constants are found to be functions of temperature and thus, can be used to measure temperature. Temperature dependent properties and constants include resistance, dielectric constant, and the magnetic permeability and susceptibility (of paramagnetic salts). Other temperature sensitive phenomena include linear and volume expansion of solids and gases, generation of the Seebeck (thermoelectric EMF) by thermocouples and the generation of Johnson (thermal) white noise by resistors.

Scientific temperature measurements are generally made using the Celsius (centigrade) or Kelvin scales. Absolute zero (thermodynamic zero) occurs at 0 K, or -273.15°C . That is, $K = ^{\circ}\text{C} + 273.15$. While most of the civilized world uses Celsius scale for such mundane things as cooking and weather reports, use of the Fahrenheit scale is dominant in the United States for these applications. The nominal boiling and freezing temperatures of water were originally taken as the two calibration points for linear temperature scales; 100° and 0° are those respective temperatures in the Celsius scale, and 212° and 32° are boiling and freezing in the Fahrenheit scale. It is easy to derive a conversion formula between $^{\circ}\text{F}$ and $^{\circ}\text{C}$:

$$^{\circ}\text{C} = 0.55556 (^{\circ}\text{F} - 32) \quad (7.282)$$

In the following discussions, we will examine the details of some of the common means of temperature measurement and secondary standards for temperature sensor calibration.

7.7.1 Temperature Standards

The SI primary standard for temperature is the triple point of pure water. The triple point of a pure substance is defined as that temperature and pressure at which all three phases (solid, liquid, vapor) are in equilibrium in a closed vessel. The triple point of pure water occurs at $+0.0098^{\circ}\text{C}$ and 4.58 mmHg pressure. This is a single, unique point in the P,T phase diagram for H_2O . The transition (melting) temperature between solid (ice) and liquid water above 4.58 mmHg pressure is a decreasing function of pressure, hence the melting point of ice is slightly pressure dependent at atmospheric pressures (Figure 7.101). The pressure dependence is sufficiently small, so that an ice bath made with double distilled water (both ice and water) can serve as a secondary standard for 0°C .

The phase transition temperatures of other pure substances can be used as secondary temperature standards. In some cases, these transitions can be observed through a microscope, which focuses on a single crystal in a capillary tube. A visible change in the

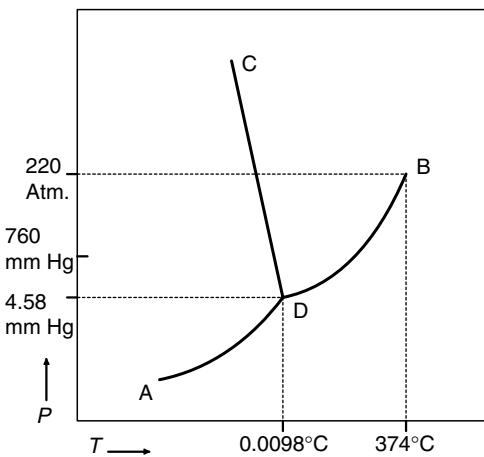


FIGURE 7.101
Phase diagram for water, showing the triple point.

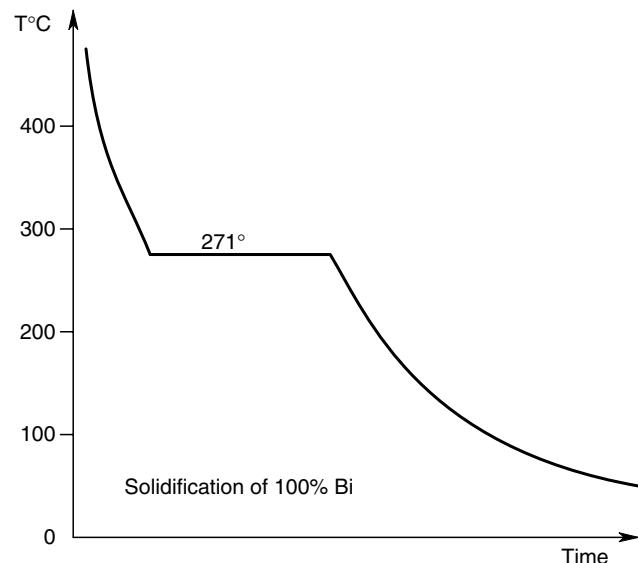


FIGURE 7.102
Cooling diagram for liquid bismuth. The plateau occurs at the freezing temperature of Bismuth.

TABLE 7.3

Melting points of some pure substances (at atmospheric pressure) which can be used for temperature calibration

Substance	Melting Temperature, $^{\circ}\text{C}$
Oxygen	-218.4
Mercury	-38.87
Water	0
Sulfur	112.8
Bismuth	271.3
Cadmium	320.9
Silver	961.9
Gold	1064.4

index of refraction generally accompanies the phase change; often, there is also a change of color, or shape. When the liquid to solid phase change of a large amount of a pure calibrating substance is used, the cooling curve has a plateau at the freezing temperature. This effect is illustrated in Figure 7.102 for the metal bismuth. As a substance freezes, it releases, isothermally at the freezing temperature, a heat of fusion which prevents the substance's temperature from continuing to drop until the phase change is complete. This process is reversed in a melting curve (not shown). The rising temperature plateaus because of the isothermal absorption of the heat of fusion. Some substances which can be used for secondary temperature calibration at atmospheric pressure are listed in Table 7.3 with their melting points.

7.7.2 Some Common Means of Temperature Measurement

7.7.2.1 Mechanical Temperature Sensors

The bending, bimetallic strip is one of the more commonly encountered mechanical temperature sensing systems. This device is used in many common household thermostats and thermometers, and is illustrated in Figure 7.103. At reference temperature, T_o , the strip is perfectly straight and has a length, L . At some higher (or lower) temperature, T , the strip is seen to bend in a section of an arc with radius R , as shown. Beckwith and Buck (1961) show that R is given by:

$$R = \frac{t\{3(1+m)^2 + (1+mn)[m^2 + 1/(mn)]\}}{6(\alpha_B - \alpha_A)(T - T_o)(1+m)^2} \quad (7.283)$$

where t is the total thickness of the strip, m is the ratio of thickness of the top to the bottom strip, n is the ratio of Young's moduli of the top to the bottom strip, α_A and α_B are the coefficients of linear expansion of the top and bottom strips, respectively. We assume $\alpha_B > \alpha_A$. If the strips have the same thickness and their Young's moduli

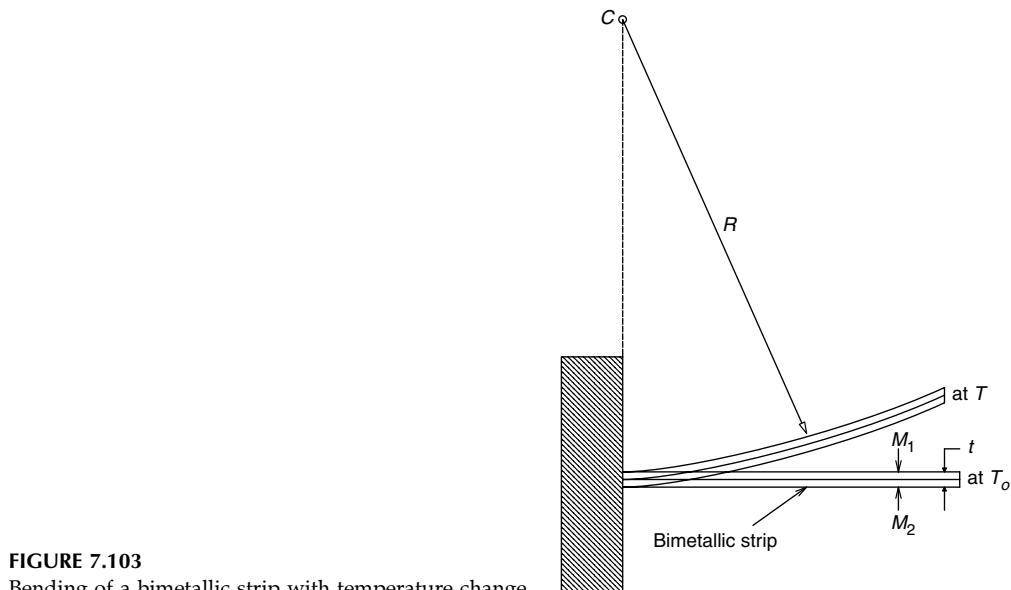
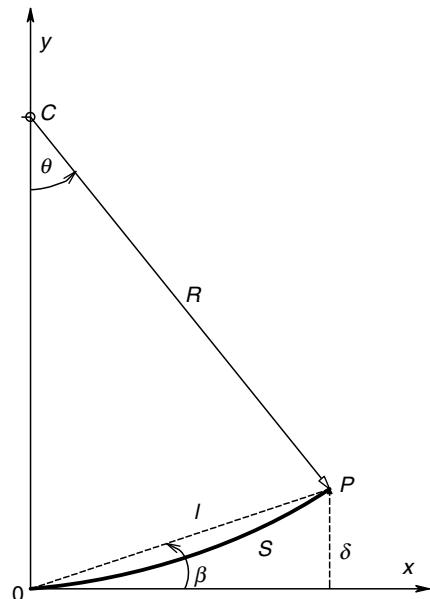


FIGURE 7.103

Bending of a bimetallic strip with temperature change.

**FIGURE 7.104**

Geometrical considerations of the tip deflection of a bimetallic strip thermometer.

are the same, equation 7.283 reduces to:

$$R = \frac{2t}{3(T - T_o)(\alpha_B - \alpha_A)} \quad (7.284)$$

In general, R will be large compared to L . From the system geometry and a knowledge of $R(T)$, we can predict the deflection of the tip of the bimetallic strip, and the angle the tip makes with the equilibrium (horizontal) position of the strip (Figure 7.104). From simple trigonometry, it can be shown that the angle β that the tip of the bimetallic strip makes with its equilibrium position is given by:

$$\beta = \theta/2 = L/2R \quad (7.285)$$

The linear deflection, δx , of the tip P from the x axis is:

$$\delta x = 2R \sin^2(L/2R) \quad (7.286)$$

And if $\beta < 15^\circ$ then, more simply,

$$\delta x \cong L\beta = L^2/2R \quad (7.287)$$

Thus, we see that both the linear tip deflection and the deflection angle are linear functions of the temperature difference, $T - T_o$. In some simple mechanical thermometers, the deflection of the tip of the bimetallic strip is converted to rotary motion of a pointer by means of a linkage. In most home thermostat designs, the tip tilts a mercury filled switch which starts the furnace. Another commonly encountered mechanical temperature sensor is the mercury or colored alcohol filled, sealed glass reservoir and capillary tube. In this common design, the filling liquid expands at a greater rate than the glass, so the column of liquid rises. The space above the expanding liquid in the sealed capillary tube is often filled with an inert gas, or is under vacuum. Most mercury filled, glass laboratory thermometers are designed for 76 mm immersion in the medium whose temperature is being measured. Other thermometers may require total immersion.

If a mercury filled, glass capillary thermometer calibrated for 76 mm immersion depth is used at a different immersion depth, then there will be a small error between the true and indicated temperatures. Beckwith and Buck (1961) discuss the means for correction of this error.

7.7.2.2 Electrical and Electronic Temperature Sensors

There are several types of electronic temperature sensing systems. We have already discussed resistance temperature detectors (RTDs) in Section 6.3.1. Thermocouples and thermopiles were covered in Section 6.4.1. In the following paragraphs, we will describe some of the less well-known electronic temperature measuring systems.

7.7.2.2.1 The Resistance Noise Thermometer

This is a system which uses the fact that the power density spectrum of Johnson (thermal) noise from a resistor is proportional to the Kelvin temperature of the resistor (Section 3.1.3.1). That is:

$$S(f) = 4kTR \text{ MSV/Hz} \quad (7.288)$$

Figure 7.105 illustrates the schematic diagram of a resistance noise thermometer. K_V is a low noise, high gain voltage amplifier, BPF is a unity gain bandpass filter with an equivalent noise bandwidth of B Hz. True RMS is a sensitive, broadband, true RMS voltmeter. R_1 is the probe resistor at the unknown temperature, T_1 . R_2 is the variable reference resistor at known reference temperature, T_2 . With the switch in position A, the mean squared meter voltage is:

$$N_A = (4kT_1R_1 + e_{na}^2)K_V^2B \text{ MSV} \quad (7.289)$$

With the switch in position B the MS meter voltage is:

$$N_B = (4kT_2R_2 + e_{na}^2)K_V^2B \text{ MSV} \quad (7.290)$$

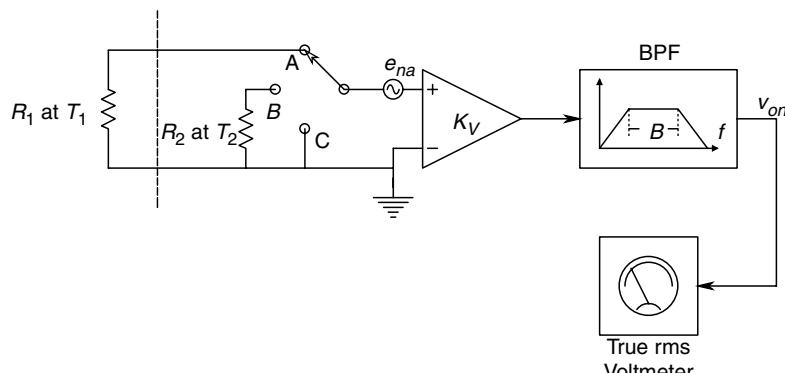


FIGURE 7.105

Schematic of a resistance noise thermometer.

In practice, the resistor R_2 is adjusted until $N_B = N_A$. Equating N_A and N_B , we easily find that at equality,

$$T_1 = T_2 R_2 / R_1 \quad (7.291)$$

The lower limit of this system's range is set by the e_{na} of the amplifier. The T_1 at which the test resistor's noise is equal to e_{na} is given by:

$$T_{1L} = e_{na}^2 / (4kR_1) \text{ K} \quad (7.292)$$

If we let $e_{na} = 3 \text{ nV}/\sqrt{\text{Hz}}$, $R_1 = 3000 \Omega$ and $k = 1.38 \times 10^{-23}$, then $T_{1L} = 54.3 \text{ K}$. The RMS output voltage of the filter when the switch is in position A, $T_1 = 54.3 \text{ K}$ and $K_V = 10^4$, is 13.4 mV, well within the range of a modern, true RMS voltmeter. Since the noise thermometer works by adjusting R_2 so that the thermal noise from R_2 at T_2 equals the noise from R_1 at T_1 , and there is uncertainty in the matching process, system accuracy is improved when the matching is repeated N times and the mean R_2 is used in equation 7.291. The resistance noise thermometer should be useful at temperatures as low as 0.5 K. A noise thermometer made with fine tungsten wire has been used for temperatures as high as 1700 K with an accuracy of 0.1% [Lion, 1959].

Since the Johnson noise from a resistor is a linear function of resistance (at a given temperature), and because the resistance of metals increases with temperature, the noise output of a noise thermometer increases more rapidly with temperature than given by the simplified relations for N_A and N_B above. This nonlinearity in response makes it necessary to calibrate a noise thermometer at several temperatures when it is to be used over a wide range of temperatures. However, the use of special, low tempco alloys for the resistor R_1 , such as manganin, eliminate the noise thermometer's nonlinearity problem in limited ranges of measurement.

7.7.2.2.2 The Resistance Thermometer

This type of thermomenter makes direct use of the fact that a metal's resistance increases with temperature. This effect was discussed in Section 6.3.1, and will be expanded on here. In general, $R(T)$ is a nonlinear function which can be approximated around a given reference temperature, T_o , by a power series:

$$R(T) = R_o(1 + \alpha \Delta T + \beta \Delta T^2 + \gamma \Delta T^3 + \dots) \quad (7.293)$$

where $\Delta T = T - T_o$, and α , β , etc. are temperature coefficients. Generally, when ΔT is small, we neglect the higher order coefficients and use only α to describe $R(T)$. In general, α is defined by:

$$\alpha \equiv \frac{dR(T)/dT}{R(T)} \quad (7.294)$$

We note that the first order temperature coefficient, α , defined by equation 7.294, is in fact, a function of temperature. Resistance thermometers are called resistance temperature detectors, or RTDs. They are generally used in a Wheatstone bridge circuit which is nulled at the reference temperature (e.g. 0°C), and the bridge unbalance voltage, V_o , is an almost linear function of ΔT . Table 7.4 lists the first order tempcos of some of the common metals used for RTDs measured at room temperature, and the useful range of temperatures over which they can be used.

TABLE 7.4
Some metals used for RTDs

Metal	Tempco (α)	Working Range of RTD, °C
Platinum	+0.00392	-190 to 800 -264 to 1000, with corrections
Nickel	0.0067	-100 to 300
Tungsten	0.0048	-100 to 400
Copper	0.0043	-100 to 250

Platinum is the preferred RTD metal because of several factors: It has a high melting point (1775.5°C), it does not oxidize in air at high temperatures, it is relatively chemically inert, and its $R(T)$ characteristic is quite linear from -190 to +400°C, and has a slight negative second derivative above 400°C. In other words, the high order coefficients of the power series, equation 7.293, for platinum are very small. Of course, pure platinum wire is expensive.

Sources of error in using an RTD include self-heating and the effect of lead resistances. Self-heating can be avoided by keeping the power dissipation in the RTD very small and constructing the RTD so that there is low thermal resistance between the wire of the RTD and the medium whose temperature is being measured. A low thermal resistance also means a rapid response time to changes of temperature in the medium. Means of compensating for lead resistances were illustrated in Figure 6.2 of this text.

In the biological temperature range, platinum RTDs are very precise and a resolution of 0.0001°C is not uncommon. However, accuracy decreases with increasing temperature. At 450°, accuracy is several hundredths of a degree, while at around 1000°C, accuracy is around 0.1°C [Lion, 1959].

Thermistors can be subdivided into negative and positive tempco devices. Negative tempco (NTC) thermistors are sintered, amorphous semiconductor devices whose resistance decreases with increasing temperature according to the relation,

$$R(T) = R_o \exp[\beta(1/T - 1/T_o)] \quad (7.295)$$

where T is in K. From the tempco definition of equation 7.294, an NTC thermistor's α is given by:

$$\alpha = -\beta/T^2 \quad (7.296)$$

β is typically 4000 K, and for $T = 300$ K, $\alpha = -0.044$. This relatively large tempco enables the precise measurement of temperatures in the biological range, with 0.0001°C resolution. Consequently, thermistor bridges are used in many applications in ecology and environmental studies, in physiological measurements, and in physical chemistry. Thermistors come in a wide variety of sizes and shapes, including 0.01 in diameter, glass covered balls. As in the case of RTDs, the power dissipation of a thermistor must be kept low enough so that self-heating does not disturb the temperature measurement. Thermistors can also be used as fluid flow probes. Two, matched NTC thermistors are used—one in still fluid for temperature compensation, the other in the moving stream. Both devices are operated at average power dissipations high enough to raise

their temperatures well above ambient temperature. The moving fluid extracts more heat from the measurement thermistor, cooling it and causing its resistance to increase. This increase of resistance is a nonlinear function of fluid velocity (Section 7.3.2.1 gives a description of conventional hot wire anemometers). We often see thermistors used as frequency determining resistors in oscillators, the frequency of which is a function of temperature. Such oscillators have been used in biotelemetry applications. The thermal response time constant of a thermistor depends on its mass, its insulation and the medium in which it is immersed. Thermistor time constants range from 100 ms to several minutes.

Thermocouples and Thermopiles are discussed in depth in Section 6.4.1. Traditionally, a thermocouple system is used with a precision potentiometer to read its EMF and the temperature is found from a look-up table for the particular couple materials used, such as copper and constantan. Potentiometers are expensive, slow to use, require a standard EMF and a reference temperature source. Consequently, where possible, it is more economical and time-saving to use calibrated RTDs or thermistors for temperature measurements formerly done with thermocouples.

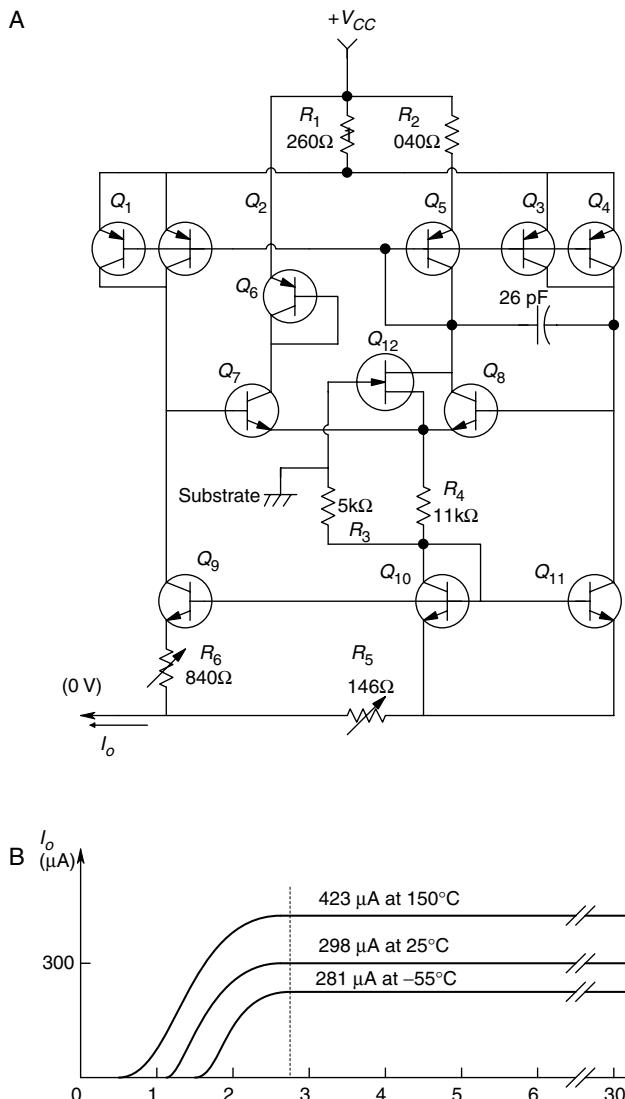
7.7.2.2.3 Electronic IC Temperature Sensors

These are specialized integrated circuits used for sensing temperatures in the -55 to $+150^\circ\text{C}$ range. Physically, they can be packaged in the form of a small metal (TO-52) can, a flat pack (F-2A) or a plastic (TO-92) case. Analog Devices offers the AD590 and AD592, two-terminal, temperature-to-current sensors. The AD sensors operate from a 4–30 V dc supply, and provide an output current which is a linear function of the sensor's kelvin temperature. That is, at 0°C , the output current is a nominal $273.15\ \mu\text{A}$, while at 25°C , the output current is $298.15\ \mu\text{A}$, etc. Over the entire operating temperature range, the AD590M has a maximum nonlinearity of $\pm 0.3^\circ\text{C}$, maximum repeatability and long term drift are $\pm 0.1^\circ\text{C}$ and the absolute error with the sensor calibrated for zero error at 25°C is no more than $\pm 1.0^\circ\text{C}$. The AD temperature-to-current sensors are generally used with an op-amp current-to-voltage amplifier, which allows generation of a voltage output proportional to the Fahrenheit or Celsius scales. Figure 7.106 shows the internal schematic of an AD590 sensor and its current/voltage/temperature curves. An op-amp conditioning circuit for an AD590 is shown in Figure 7.107.

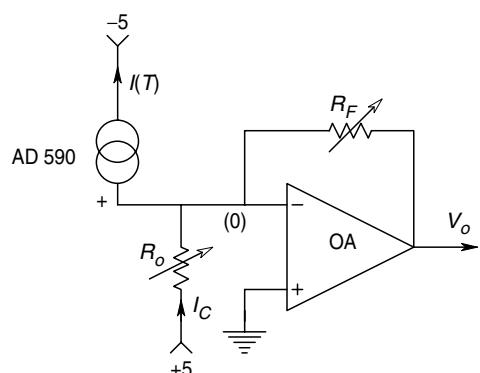
National semiconductor also makes IC temperature sensors—the LM34, LM35, LM134, LM135, LM234, LM235, LM335 series. These sensors are three terminal devices, which produce an output voltage proportional to the temperature. For example, the AD34CA has a $10\ \text{mV}/^\circ\text{F}$ output voltage gain, a range from -50 to $+300^\circ\text{F}$, and a nonlinearity of $\pm 0.3^\circ\text{F}$. The LM35 series has a $10\ \text{mV}/^\circ\text{C}$ output voltage gain over a -55 to $+150^\circ\text{C}$ range, and $\pm 0.15^\circ\text{C}$ nonlinearity. IC temperature sensors are ideally suited for environmental monitoring and control applications within their temperature ranges.

7.7.2.2.4 Optical Pyrometers

These instruments provide a no-touch means of estimating the surface temperatures of hot objects in the range of 775 – 4200°C , such as metals being hot-worked, molten metals, gas plasmas and furnace interiors. Optical pyrometers make use of the fact that all objects at temperatures above $0\ \text{K}$ radiate heat in the form of broadband, electromagnetic energy. The range of the electromagnetic spectrum, generally considered to be thermal radiation, lies in the range 0.01 – $100\ \mu\text{m}$ wavelength. Objects which are radiating heat are characterized by three parameters which describe what happens to long wave,

**FIGURE 7.106**

A. Simplified schematic of an Analog Devices' AD590 electronic temperature sensor. B. Output current *vs* applied voltage and temperature for the AD590 sensor.

**FIGURE 7.107**

A simple op-amp circuit to condition the current output of the AD590 temperature sensor. For dc stability, the op-amp should be a chopper or auto-zeroing type. R_o sets $V_o = 0$ at 0°C , and R_F sets the output scale, so 100°C gives $V_o = 10\text{ V}$.

electromagnetic radiation (heat) at their surfaces. This relation is:

$$e = a = 1 - r \quad (7.297)$$

where e is the surface emissivity, which is always equal to its absorbtivity. r is the reflectivity of the surface. An ideal blackbody radiator has $e=a=1$ and $r=0$ (that is, all radiant energy striking its surface is absorbed, and none is reflected). For a non-ideal radiator, a finite fraction of the incident energy is reflected. The same properties exist for emitted radiation from a blackbody. For a hot object which is not an ideal radiator, the radiated heat output is given by:

$$W = e W_{bb} \quad (7.298)$$

where W is the total radiant emittance (in MKS units, W/m^2) from the hot surface, e is the emissivity and W_{bb} is the radiant emittance from an ideal blackbody.

Max Planck, using the newly developed quantum theory in the early 1900s, developed an expression for the power spectrum of heat radiation from a blackbody [Sears, 1949]:

$$W_f = \frac{2\pi h f_3}{c^2 \exp(hf/kT) - 1} = \frac{dW_{bb}}{df} \quad (7.299)$$

where W_f is in $\text{W}/(\text{m}^2 \text{ Hz})$, c is the speed of light, k is Boltzmann's constant, h is Planck's constant and W_{bb} is the radiant emittance of the blackbody in W/m^2 . From an experimental and practical point of view, it is useful to write Planck's equation in terms of wavelength instead of frequency. Substituting $f=c/\lambda$, we find:

$$W_\lambda = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda T) - 1} = \frac{dW_{bb}}{d\lambda} \quad (7.300)$$

where $C_1 = 2\pi c^2 h = 3.740 \times 10^{20}$, $C_2 = hc/k = 1.4385 \times 10^7$ and W_λ is in $\text{W}/(\text{m}^2 \text{ m}\mu)$. A plot of equation 7.300 is shown in Figure 7.91. There are two important observations to make about this figure:

1. As the temperature of the blackbody increases, the peak of maximum spectral emittance shifts systematically to shorter wavelengths
2. As the temperature increases, the area under the spectral emittance curves increases

Subjectively, this means that when we look at a hot object, the color of the object's surface shifts from a dim, dull red to a bright orange to a brighter yellow, and so on. The peaks of the spectral emittance curves can be found by differentiating the expression for W_λ and setting the derivative equal to zero. This gives a transcendental equation which can be solved numerically to yield the wavelength at the peak, λ_{PK} .

$$\lambda_{PK} = 2.8971 \times 10^6 / T \quad (7.301)$$

Here, λ_{PK} is in nanometers and T in K. This relation is called Wien's displacement law.

If we integrate the expression for spectral emittance, we obtain an expression for the total radiated power per m^2 from the blackbody's surface:

$$W_{bb} = \int dW_{bb} = \int W_\lambda d\lambda = \frac{\pi^4 C_1 T^4}{15 C_2^4} = \sigma T^4 \quad (7.302)$$

The constants have been defined above, and $\sigma = 5.672 \times 10^{-8}$ for MKS units. Equation 7.302 is known as the Stefan-Boltzmann equation.

Practical hot surfaces have non-unity emissivities which are generally a function of wavelength. Thus, the practical, spectral emittance curve for a hot object may have many peaks and valleys. If these irregularities are averaged out, we can often fit a scaled down, blackbody spectral emittance curve to the practical curve so that their peaks occur at the same temperature, T . Such a scaled blackbody, W_λ , is called a graybody curve.

In one form of optical pyrometer, shown in Figure 7.108, a human operator makes a subjective color comparison of a glowing tungsten filament with the hot surface under measurement. The color comparison is made easier by optically superimposing the image of the filament on that of the hot object. When the filament's color, determined by its current, matches that of the object, it disappears on the background of the object, and the filament current is read. Since the filament may be considered to be a blackbody, its spectral emittance closely follows that of the object when its temperature is the same as the object's. Often, a red filter is used to convert the color matching task to one that involves brightness matching. The filament ammeter is calibrated in temperature.

In another form of optical pyrometer, shown in Figure 7.109, the filament is run at a constant current and brightness. The intensity of the image of the hot object is then varied with a neutral density wedge. Again, a red filter is used to make the null process an exercise in monochromatic intensity matching. The wedge position is calibrated in object temperature, assuming blackbody emission.

Total radiation pyrometers are also known as radiometers, and are electronic instruments that measure the integral of the W_λ curve. That is, their output is proportional to the integral of W_{bb} over the area of the instrument's aperture. In one form of radiometer, the detector is a thermopile in which half of the thermocouple junctions must be kept at a reference temperature. The thermopile is designed to be a nearly 100% blackbody absorber. Thus, regardless of the shape of W_λ , all of the incoming

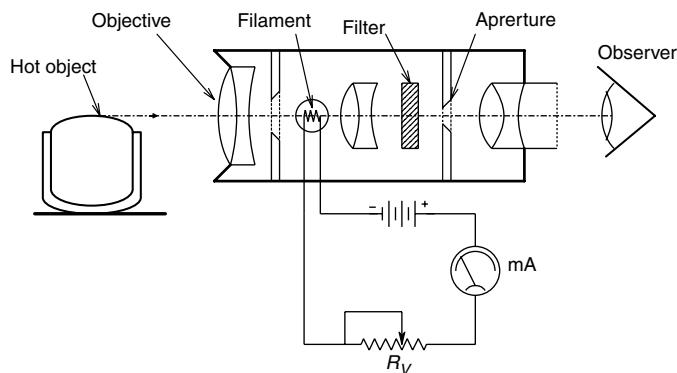
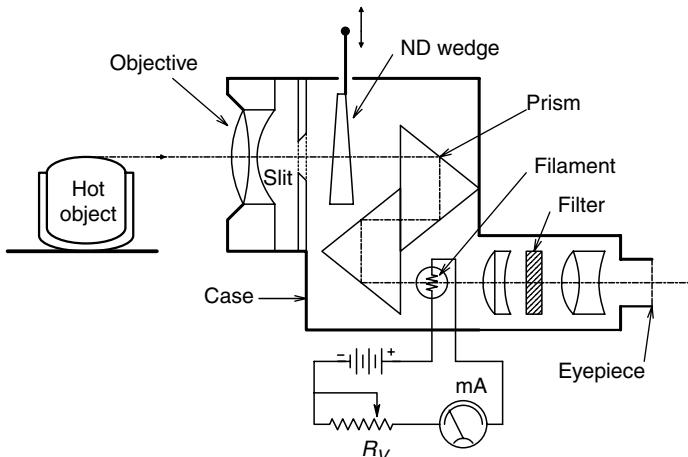
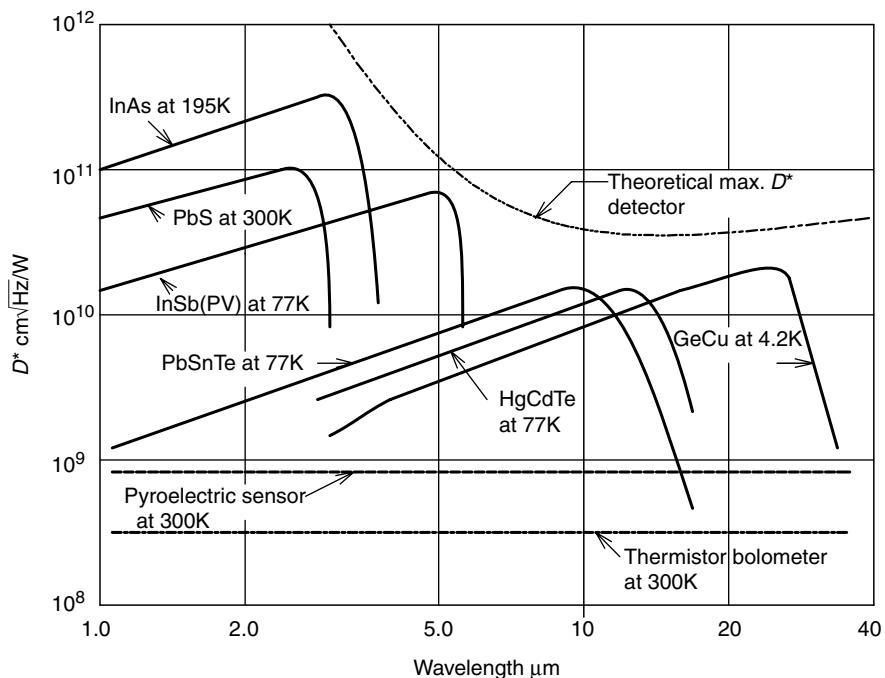


FIGURE 7.108

Diagram of an optical pyrometer. The operator adjusts R_V until the color temperature of the filament matches that of the object. The human operator must have reliable color vision in the red-yellow end of the spectrum.

**FIGURE 7.109**

In this form of optical pyrometer, the operator matches monochromatic intensities of the filament and object. The neutral density wedge position is calibrated in object temperature.

**FIGURE 7.110**

Spectral sensitivities of some common IR sensors. Sensitivity can generally be improved by chilling the sensor with liquid nitrogen. Note: Solid lines = photoconductors, Dashed line = Pyroelectric sensors, Dash-dot line = Thermistor bolometers. Vertical axis = D^* , the specific detectivity. (Source: Adapted from Barnes' Handbook of Infrared Radiation Measurement, 1983)

energy is captured and converted to a temperature rise of the sensing junctions. Total radiation pyrometers can be used with blackbody radiation, as well as coherent sources such as lasers.

Other detectors used in radiation pyrometers may include photoconductors, photodiodes and pyroelectric detectors. Pyroelectric detectors absorb thermal energy and generate electrical signals. One example of a pyroelectric detector material is the

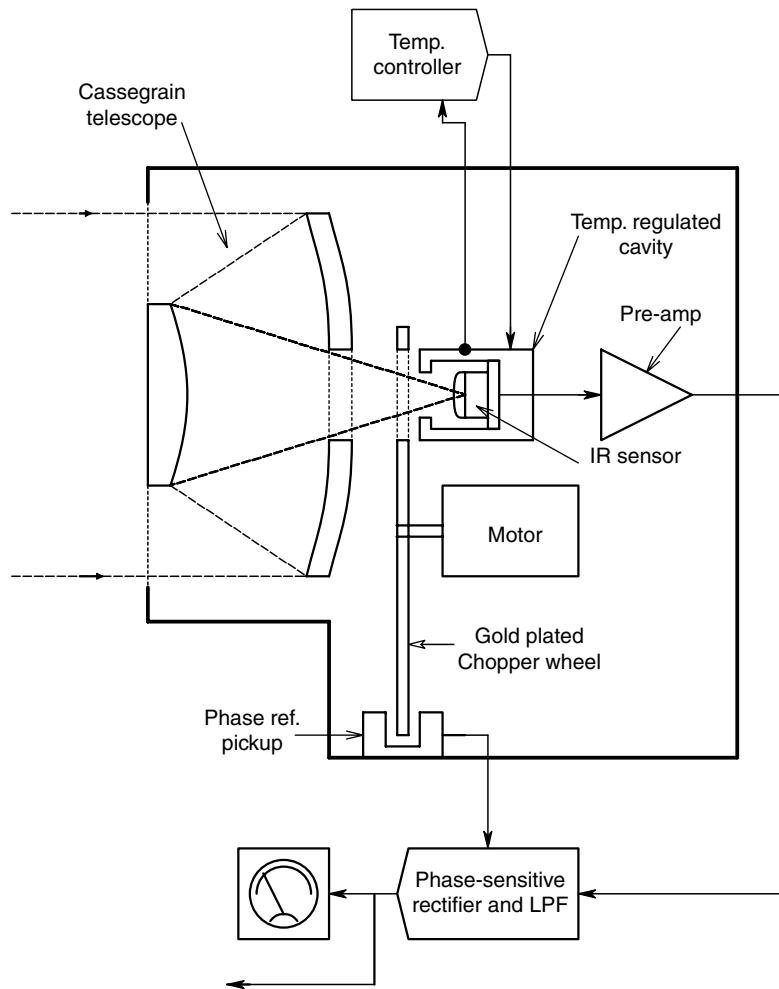
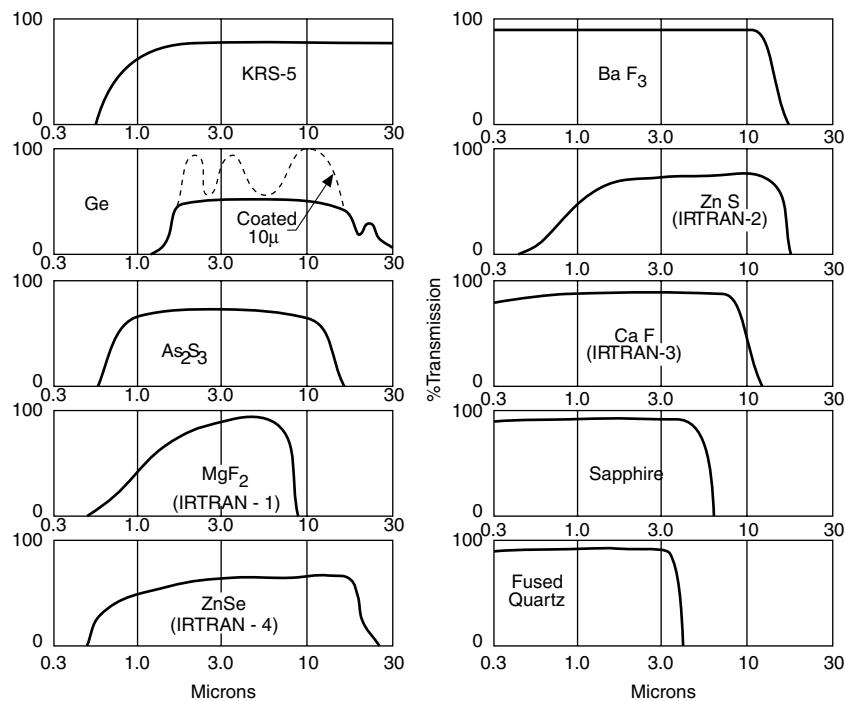
**FIGURE 7.111**

Diagram of a basic IR radiometer. Note that chopping is necessary to avoid $1/f$ noise from the sensor and pre-amp. (Source: Courtesy of Barnes Engineering Co.)

polarized polymer film, polyvinylidene difluoride (PVDF). PVDF film absorbs strongly in the IR, and a free mounted, 28 μm PVDF film has a response time constant of about five seconds to a step temperature change. All metallic radiation detectors have spectral response characteristics. The broadest response characteristic is that of the blackbody thermopile—this sensor responds to radiation from UV to FIR (0.25–20 μm). PVDF pyroelectric sensors respond primarily in the range 6.6–66 μm . LiTaO₃ crystals are also used in faster responding pyroelectric sensors. Most photodiodes and photoconductors respond maximally in the NIR, and are relatively narrow band devices compared with PVDF and thermopiles. Figure 7.110 illustrates the detectivity (D^*) of some common photoconductive IR detectors, pyroelectric detectors, thermopiles and thermistor bolometers. Note that several of the photoconductive sensors are operated at LN₂ temperature (e.g. HgCdTe, GeCu at 77 K) to reduce noise. The detectivity, D^* , is defined as $D^* \equiv \sqrt{(\text{sensor area})/\text{NEP}}$. Its units are $(\text{cm}\sqrt{\text{Hz}})/\text{W}$, and at a given wavelength, a higher D^* is better. NEP is the sensor's noise equivalent power. The NEP is

**FIGURE 7.112**

IR transmission characteristics of some materials used in IR optical systems. (Source: Courtesy of Barnes Engineering Co.)

the sensor's optical input power in watts required to give an output SNR of unity. The sensor's total noise can be from several sources, and is assumed to have a white power density spectrum. NEP is also a function of wavelength, and a low NEP is better.

The optics of pyrometers also present a problem, as conventional glasses do not transmit effectively at wavelengths beyond $2.6\text{ }\mu\text{m}$. Thus, special materials must be used in IR radiometers for windows, mirrors, and lenses. A broadband, IR radiometer/pyrometer uses front surface mirrors coated with aluminum or gold to focus the radiation on the sensor. A chopper wheel is used to 0,1 modulate the radiation and permit ac amplification of the sensor output signal out of the $1/f$ noise band. Phase sensitive rectification is then used to recover the amplified signal. A typical IR radiometer system is shown in Figure 7.111. When lenses and windows are used in radiometer design, they must be able to transmit in the IR. Figure 7.112 summarizes the IR transmission characteristics of commonly used IR optic materials. Note that pure germanium is now widely used for IR optics [Barnes, 1983]. The passband of germanium is from about 2–30 microns. Pure germanium can be considered to be a dielectric and has a high refractive index of $n=4$. The reflection coefficient of germanium, r , is given by the Fresnel reflection equation:

$$r = \frac{(n - 1)^2}{(n + 1)^2} = 0.36 \quad (7.303)$$

Thus, assuming zero absorption for a Ge window, only $(1 - 0.36)^2 = 0.41$ of the incident IR would be transmitted [Barnes, 1983].

The optics of some radiometers make use of a chopping scheme in which the IR sensor is alternately exposed to the radiation under measurement, dark, and a standard blackbody source of known temperature. Thus, calibration and correction for dark response can be done simultaneously by a computer attached to the radiometer.

7.8 Chapter Summary

In this chapter, we have presented a survey of some of the common (and not so common) means of measuring certain physical quantities, including angular acceleration, velocity and position, linear acceleration, velocity and position, force, torque, pressure and vacuum, and temperature.

Mechanical rate gyroscopes (including classical rotating mass, vibrating mass, air jet and micromachined oscillating disk) gyroscopes were described and analyzed using the LaGrange method. Sagnac fiber optic rate gyros were presented as systems for sensing the inertial angular velocity and position of vehicles. Clinometers using pendulums and liquid filling were also described and their dynamics analyzed.

Various kinds of linear and rotational accelerometers were described, including servo-nullled, Newtonian moving mass systems, piezoelectric and micromachined units.

The Doppler effect was introduced as the basis for many instruments, using sound or electromagnetic radiation to measure the linear velocity of vehicles or particles suspended in a moving fluid. Laser Doppler velocimetry (LDV) was described.

A new section on the Global Positioning System (7.3.3.1) and its upgrades as a means for navigation on the Earth's surface and in its atmosphere was presented.

Optical interferometry was introduced and analyzed as a means to measure small displacements. Closed loop, optical and ultrasonic phase lock, velocity and ranging, and systems developed by the author were described as a means of continuously measuring vehicle motion.

Sensors used to measure force and pressure were seen to be related, in that a pressure acting on the area of a diaphragm or piston produces a force. Forces can produce changes in resistance, capacitance, inductance, charge displacement (in piezoelectric transducers), as well as the optical transmission properties of fiber optic waveguides. These physical changes can be sensed as functions of force or pressure. Some torque sensors also make use of sensed mechanical strain caused by the torque.

The principles of spectrometry were introduced in Section 7.6. Dispersive, non-dispersive and Fourier transform spectrometers were described as a means to quantify chemicals in gas liquid or solid states. The additional chemical analytical usefulness of single and multiple bubble sonoluminescence, and surface plasmon resonance were also reviewed.

We saw that many physical processes respond to changes in temperature. These include chemical reactions, conduction in semiconductors, resistance of conductors, dielectric constant, speed of sound, refractive index of optical fibers, etc. In fact, many other physical measurements require compensation for temperature changes. Optical pyrometry was introduced as a means of remote sensing the temperature of objects.

The basis for many measurement systems and sensors has been found to lie in several, basic physical laws. These laws include the Doppler effect, the Sagnac effect, the piezoelectric effect, the Hall effect and the piezoresistive effect (strain gauges), to name a few.

Problems

- 7.1 A conventional pendulum clinometer is subject to error in its static tilt angle reading when the case is subject to a linear acceleration component, \ddot{x} in the $+x$ direction, perpendicular to the gravity vector, as shown in Figure P7.1. The \bullet symbol shows the center of mass of the pendulum. In the Figure, a $\theta_c = +17^\circ$ is shown. If there were zero acceleration of the case in the x direction, the pot wiper (and pendulum) would be vertical. Due to the acceleration to the right, the pendulum angle is $\theta_p = 26^\circ$, producing a serious error in the tilt reading.

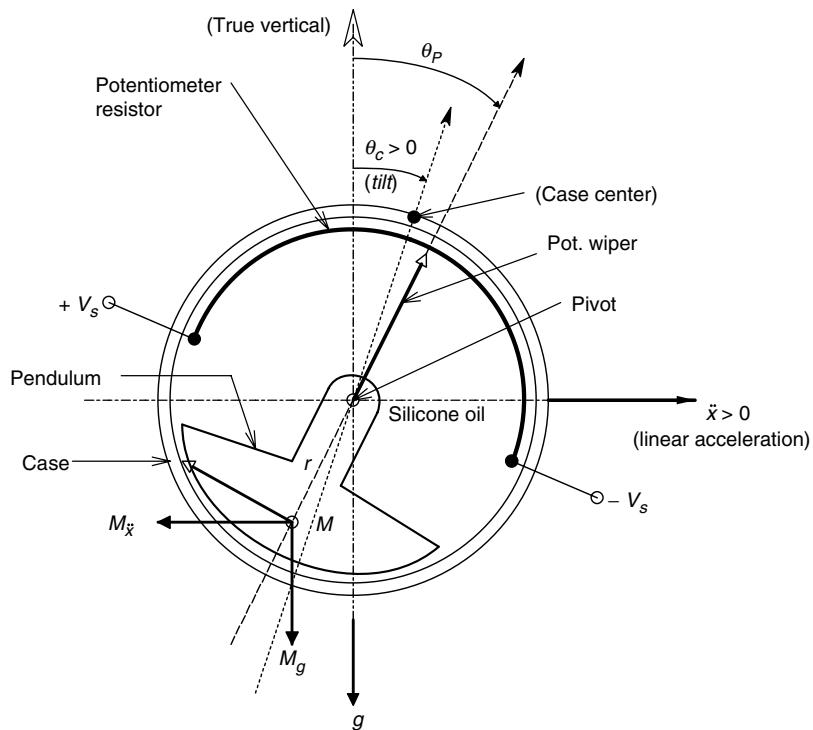


FIGURE P7.1.

- (A) Derive a transfer function for V_o as a function of the linear acceleration, \ddot{x} . Assume $\theta_c = 0$, $|\theta_p| < 15^\circ$, so $\sin(\theta_p) \cong \theta_p$ in radians and $\cos(\theta_p) \cong 1$. Also, $V_o = K_o(\theta_c - \theta_p)$. The pendulum is subject to a viscous friction torque from the silicone oil, $\tau_f = D(\dot{\theta}_p - \dot{\theta}_c) \text{ Nm}$. τ_f acts to retard the clockwise swing of the pendulum, as shown in the figure.
- (B) Assume a constant linear acceleration, \ddot{x} , as shown. In the steady state, how much would the case have to be rotated, and in what direction, so that $V_o = 0$?
- 7.2 An active, electromagnetic, differential pressure sensor is shown in Figure P7.2. Assume that the net force acting on the diaphragm is given by:

$$0 = P_o A - P_i A - F_M - M \Delta \ddot{x}$$

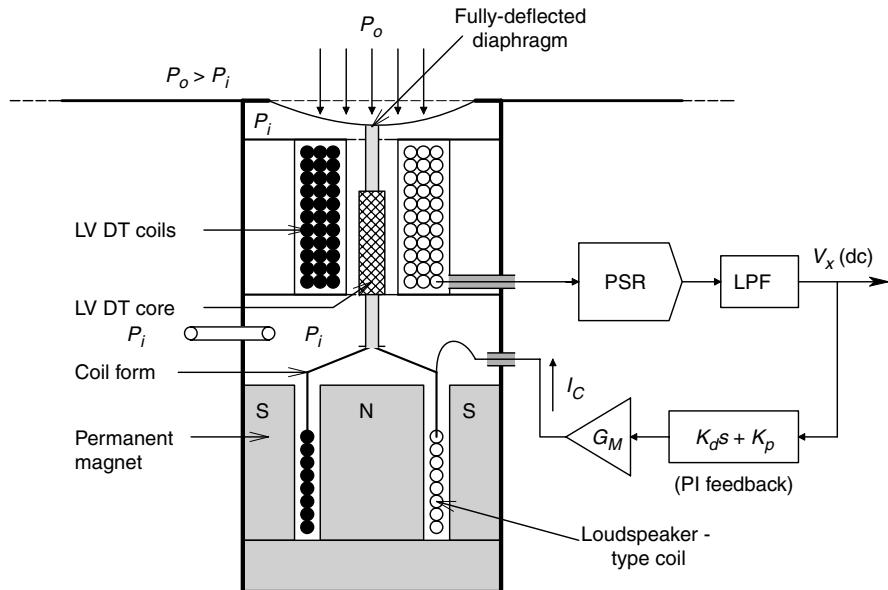


FIGURE P7.2.

Forces with a minus sign act upward. Diaphragm deflection is measured with an LVDT whose output is given by $V_x = K_L \Delta x$. The diaphragm stiffness is neglected. The upward electromagnetic force from the coil is $F_M = K_M I_C$.

From the figure: $I_C = G_M [K_d \dot{V}_x + K_p V_x]$ and $\Delta P = P_o - P_i$

- (A) Derive an expression for the transfer function, $\frac{V_x}{\Delta P}(s)$, in time constant form.
- (B) Find expressions for the sensor's steady state compliance, $\Delta x/\Delta P$ and the sensor's ω_n .

7.3 The circuit of Figure P7.3 is used to generate a DSBSCM square wave at $v_o(t)$, modulated by the ΔC of the differential capacitor sensor. The square wave period T satisfies $RC_o \ll T \ll R_F C_F$. For example, let $RC_o = 10^{-5}$ s, $T = 0.001$ s and $R_F C_F = 0.1$ s.

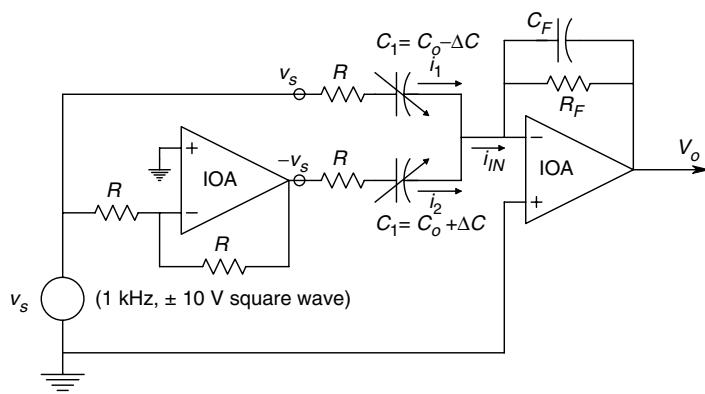
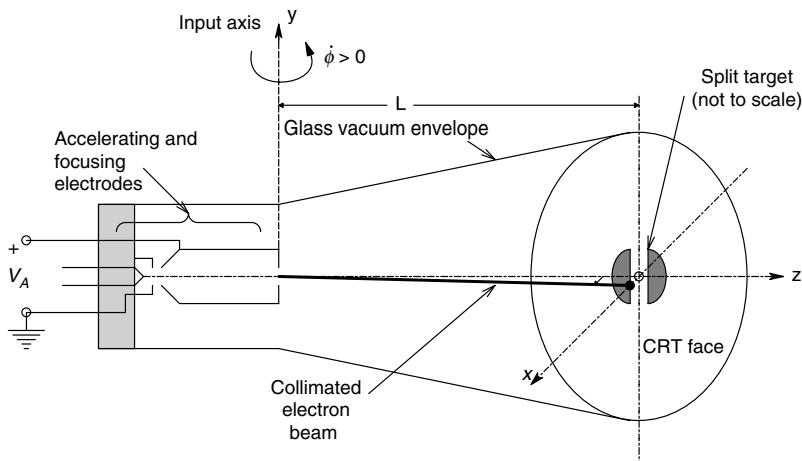


FIGURE P7.3.

(A) Plot and dimension the steady state currents $i_1(t)$ and $i_2(t)$ flowing into the op-amp's summing junction. Give an expression for the area under $i_1(t)$, A_1 and under $i_2(t)$, A_2 , over a half cycle ($0 < t < T/2$). Note that both currents effectively tend towards 0 before $T/2$, so that ∞ can be used for the upper limit of integration in finding the areas to simplify the math. Note the summing junction is at virtual ground.

(B) Since the duration of the net input current transient is very short compared to $\tau_F = R_F C_F$, we can consider the inputs to the op-amp's SJ to be impulse functions with areas of $+(A_1 - A_2)$ and $-(A_1 - A_2)$ occurring at the positive and negative switching times of the input square wave, respectively. That is, at $t=0$, $i_{in}(t) = (A_1 - A_2)\delta(t)$ etc. Find, plot and dimension $v_o(t)$ over two square wave cycles for $\Delta C > 0$. Note $e^{-x} \approx 1 - x$ for $x \ll 1$.

7.4 In Section 7.2.2.4, we examined the operation of the novel Humphrey air jet rate gyro. In this problem, we will examine the feasibility of using a modified cathode ray tube (CRT) as a rate gyro. The system is shown schematically in Figure P7.4. Electrons are accelerated through a potential V_A , and then a focused beam of these moving electrons leaves the accelerating/focusing plates with a velocity v_x , traveling to the CRT face *in vacuo*. On the inside of the face is a circular tungsten target, 1 cm in diameter. The target is split into two halves by a narrow, 1 μm gap. With the CRT stationary ($\dot{\phi} = 0$), the beam is focused on the stationary CRT face, so that one-half of its current is collected by each half of the target. Assume the accelerating voltage is $V_A = 100$ V, the free flight distance is $L = 31$ cm, the electron beam has a circular cross-section, and is 1 mm in diameter at the target. Assume further that a 1 μm lateral deflection of the beam at the target is just detectable electronically.



- (A) Find the threshold rotation rate, $\dot{\phi}_t$, that will produce this detectable beam deflection.
- (B) Now assume Hg^{++} ions form the beam, other conditions being the same as in part (A). Find the threshold rotation rate for a mercury ion beam.
- (C) Design an analog circuit to convert the differential electron currents to or from the target electrodes to an output voltage proportional to $\dot{\phi}$.

7.5 The circuit of Figure P7.5 uses an AD532 temperature-to-current sensor instead of an ice bath as a temperature reference for an Iron-Constantan thermocouple. For $V_{SS} > 4V$, the dc current sourced by the AD532 IC is given by: $I(T) = (T + 273)10^{-6} A$. T is in $^{\circ}\text{C}$. Assume the op-amp is ideal. $R_F = 9\text{k}\Omega$, $R_1 = 1\text{k}\Omega$, $S_{\text{Fe}} = 18.5\text{ }\mu\text{V}/^{\circ}\text{C}$, $S_{\text{Con}} = -35.0\text{ }\mu\text{V}/^{\circ}\text{C}$, $S_{\text{Cu}} = 6.5\text{ }\mu\text{V}/^{\circ}\text{C}$.

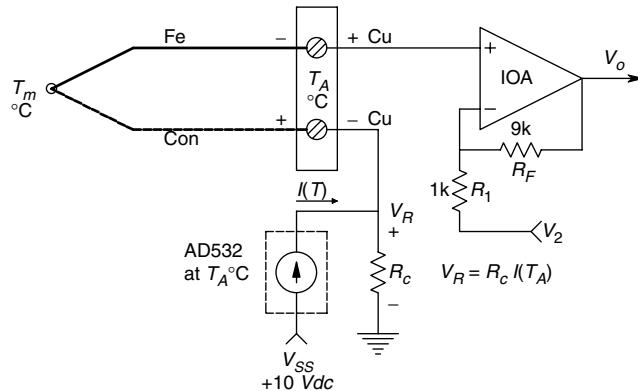


FIGURE P7.5.

- (A) Derive a general algebraic expression for $V_o = f(T_m, T_A, \text{circuit parameters})$. Use superposition.
- (B) Find numerical values for R_C and V_2 that will make $V_o = KT_m$ exactly. Find the numerical value of K .

7.6 A Sagnac effect FOG, shown in Figure P7.6, uses a non-reciprocal optical phase shift (NRPS) device to null the interference fringe shift pattern that is caused by the angular velocity of the FO ring. The system relations are: $\theta_s = K_s\dot{\phi}$, $\theta_e = \theta_s - \theta_f$, optical power at the detector is $P_D = (P_o/2)[1 + \sin(\theta_e)] W$, $V_1 = K_D P_D$, NRPS gain is $\theta_f = K_C V_3$. Under normal operating conditions, $|\theta_e| < 3^{\circ}$, so $(\theta_e) \cong \theta_e$ radians.

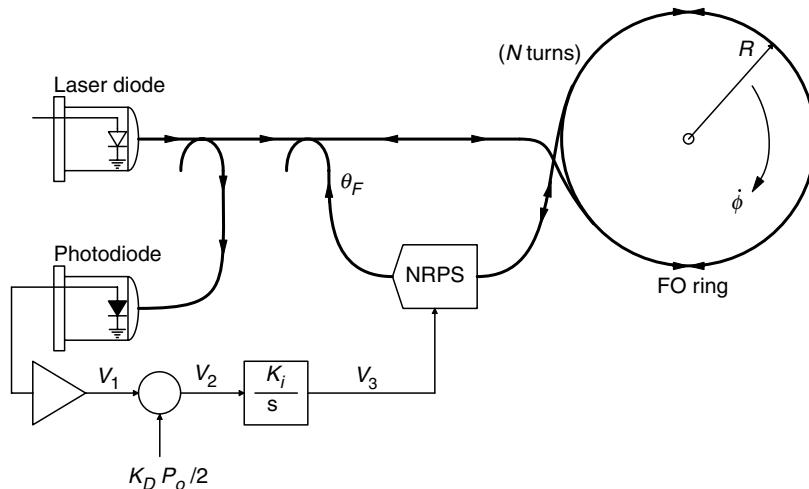


FIGURE P7.6.

- (A) Draw a linearized block diagram of the system.
- (B) Find an expression for $\frac{V_o}{\phi}(s)$ in time constant form. What advantage does this system have over a simple FOG?
- 7.7 A proposed design for a vibrating ball rate gyro is shown in Figure P7.7. A ball bearing with mass M is forced to oscillate radially in a groove milled in a disk 'platform'. The ball's displacement in the groove is forced to be $r(t) = r_0 \sin(\omega_0 t)$, regardless of the platform input rotation rate, $\dot{\phi}$. The disk, groove and ball are supported by a post that has negligible compliance. The axial torque on the post, M_p , is measured by a semiconductor strain gauge system so $V_o = K_T M_p$. Use the Lagrange method to find the instantaneous, perpendicular force, F_ϕ , on the ball due to gyro rotation. Note that $\frac{d}{dt} \frac{\partial L}{\partial \dot{\phi}} - \frac{\partial L}{\partial \phi} = F_\phi$, where ϕ is the input angle, $\dot{\phi}$ is the input rate, L is the Lagrange function and $L = T - U$. T is the total system kinetic energy, and U is the total potential energy ($U=0$ in this system). Note that there are two orthogonal components of velocity on the ball, $v_\phi = r\dot{\phi}$ and $v_r\dot{r}$. In determining F_ϕ , neglect terms containing $\ddot{\phi}$.

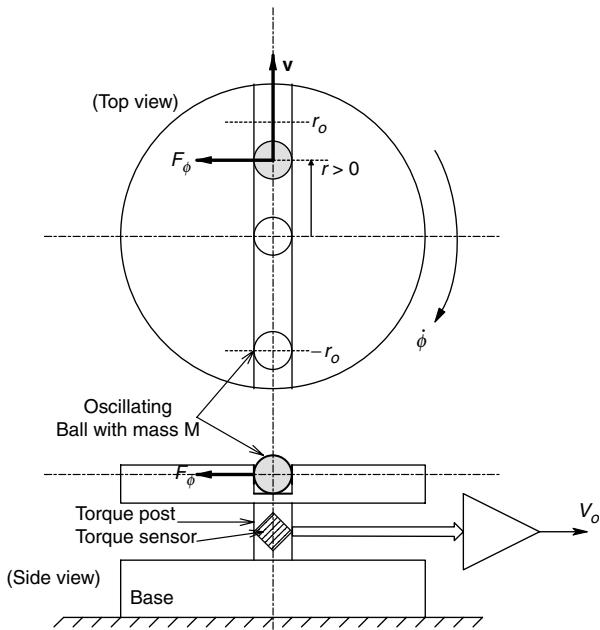


FIGURE P7.7.

- (A) Give expressions for L , $\frac{\partial L}{\partial \dot{\phi}}$, $\frac{\partial L}{\partial \phi}$ and $\frac{d}{dt} \frac{\partial L}{\partial \dot{\phi}}$. Note that $r(t) = r_0 \sin(\omega_0 t)$.
- (B) Derive an expression for $M_p(t)$ as a function of $\dot{\phi}$.
- (C) Describe the type of signal processing required to get $V_o = K_o \dot{\phi}$.
- 7.8 Instead of using a voltage-to-period converter, a constant phase LADAR system is configured using a voltage-to-frequency converter in which $f_o = K_V V_o$, as shown in Figure P7.8. The distance to the stationary, reflecting target is L_o meters. In the steady state, $V_v = 0$ and $V_\varphi = V_{set}$.
- (A) Derive an expression for V_o in the steady state.

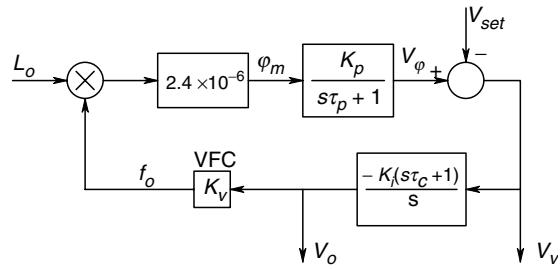


FIGURE P7.8.

- (B) Now assume the target is receding at constant velocity $v = \dot{L} > 0$. Assume steady state conditions so that $\dot{V}_v = 0$. Use the integrator transfer function to derive an expression for the steady state V_v . Note that \dot{V}_o can be found by assuming $L(t)$ and differentiating the V_o expression of part (A). Also note that V_v and V_o are both nonlinear functions of L .

- 7.9 A double gimbal mechanical gyro is spun up and initially set so $\theta = 0^\circ$ and $\phi = 0^\circ$. The linearized gyro torque equations are:

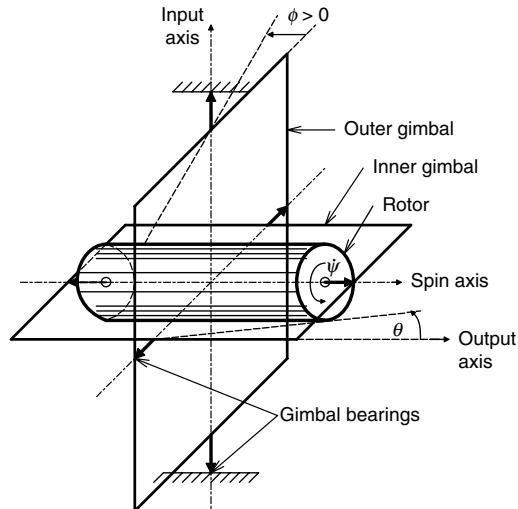


FIGURE P7.9.

$$M_\theta = J_1 s^2 \Theta(s) - H_r s \Phi(s)$$

$$M_\phi = H_r s \Theta(s) + J_1 s^2 \Phi(s)$$

Assume $M_\theta = 0$ (unrestrained inner gimbal), and let $M_\phi(t) = M_{\phi_0} \delta(t)$ (impulse of torque) be applied around the outer gimbal.

- (A) Find expressions for $\Theta(s)$ and $\theta(t)$. Sketch and dimension $\theta(t)$.
 (B) Repeat (A) for $\Phi(s)$ and $\phi(t)$.

- 7.10 The system shown in Figure P7.10 is a constant phase, closed loop, laser ranging (LAVERA) system. Short, periodic pulses of frequency f are emitted from a collimated, 910 nm laser diode source. The light reflects off a target at distance L ; the reflected light propagates back to the receiving optics which focus it on an avalanche photodiode sensor. In the steady state, $V_i = 0$.

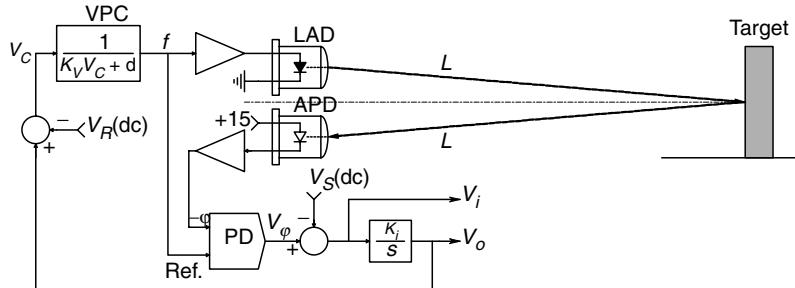


FIGURE P7.10.

- (A) Derive a simple expression for the steady state phase, φ , between the transmitted and received pulses. Neglect the delays associated with the logic and analog signal processing. In the system, a digital phase detector (PD) converts φ to a voltage: $V_\varphi = K_D \varphi$. V_C is the input to a VPC oscillator in which the period is given by $T = K_V V_C + d$. The oscillator output is narrow pulses that drive the LAD to flash at a rate $f = 1/T$. Assume that in the steady state, $dL/dt = V_i = 0$.
- (B) Derive expressions for φ_{ss} , f_{ss} and V_{oss} .
- (C) Now assume the target is moving at a constant velocity, $\dot{L} > 0$, away from the system. Derive expressions for $V_i = f(\dot{L}, \text{other parameters})$ and $f = g(L, \dot{L}, \text{other parameters})$.

8

Basic Electrical Measurements

8.1 Introduction

Electrical measurements are defined in this chapter as measurement of the traditional electrical parameters of voltage, current, electric field strength, magnetic fields, resistance, capacitance, inductance, the ac steady state parameters of impedance and admittance, power, frequency and phase.

8.2 DC Voltage Measurements

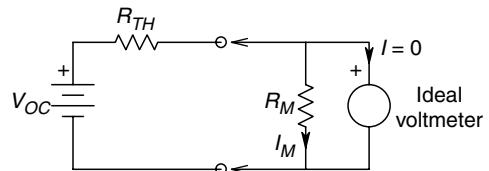
DC voltage measurements can be made over an enormous range, from nanovolts to thousands of kilovolts. The practical limit to low voltage measurement is noise, while the practical limits to high voltage measurements involve circuit loading and insulation. Of course, the specialized voltmeters that allow measurement of nanovolt potentials are generally unsuitable for high voltage measurements. All voltmeters, regardless of their specialization, can be represented by an equivalent circuit consisting of a parallel impedance by which the voltmeter loads the circuit under test (CUT), and an ‘ideal’, infinite impedance voltmeter. Thus, we see that every voltmeter takes some power from the CUT. An ideal voltmeter would take zero power and not load the CUT. The CUT itself, by Thévenin’s theorem, can be represented by an open circuit voltage and equivalent series impedance. Obviously, connecting a practical voltmeter to the Thevenin equivalent circuit will result in current flowing in the circuit, and the (indicated) voltage at the voltmeter’s terminals being less than the open circuit voltage by the amount of the voltage drop across the Thevenin resistor. This situation is illustrated in Figure 8.1. In mathematical terms, assuming resistors rather than impedances, the voltmeter reads:

$$V_M = V_{OC}R_M/(R_M + R_{TH}) \quad (8.1)$$

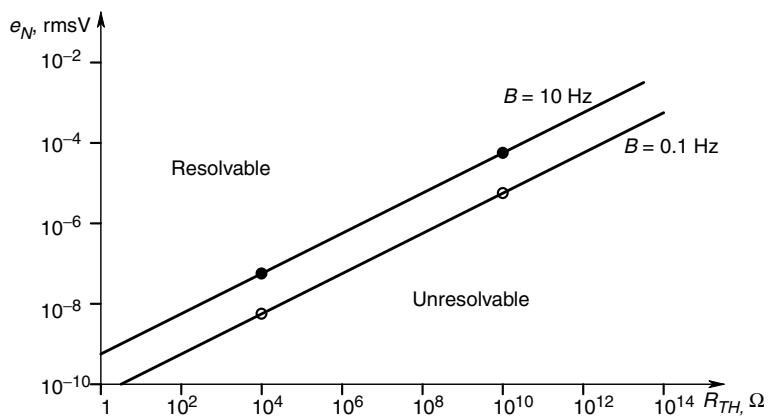
Under normal circumstances, $R_M \gg R_{TH}$.

Voltmeter loading of the CUT is related to a parameter called the *voltmeter sensitivity*, η . η is defined as the full scale voltage of the meter divided by the power dissipated in the meter at full scale voltage, V_{FS} :

$$\eta = V_{FS}/P_{FS} = \frac{V_{FS}}{V_{FS}^2/R_M} = \frac{R_M}{V_{FS}} = 1/I_{MFS} \Omega/V \quad (8.2)$$

**FIGURE 8.1**

Simple circuit showing a non-ideal voltmeter attached to a Thevenin equivalent source.

**FIGURE 8.2**

Graph illustrating the role of thermal noise from the Thevenin source resistor in limiting the resolution of low level dc voltage measurements. 10 Hz and 0.1 Hz noise bandwidths are considered.

where I_{MFS} is the current drawn by the voltmeter when at its full scale voltage. Typical analog dc voltmeters, such as found in multimeters, have η values of $20,000 \Omega/V$, which means that the voltmeter draws $50 \mu A$ from the CUT at full scale voltage. The equivalent resistance of any dc voltmeter can always be found by multiplying V_{FS} by η . Some electronic voltmeters have a fixed input resistance, regardless of scale. This resistance is typically 10 or $11 M\Omega$, but some modern electronic voltmeters have R_{MS} as large as $10^{10} \Omega$.

Figure 8.2 illustrates the role of Johnson (thermal) noise from R_{TH} in limiting the resolution of dc voltage measurements. Recall from Section 3.2.3.1 that the Johnson noise power density spectrum from a resistor is given by $S(f) = 4kTR MS V/Hz$. The effective Hz bandwidth of a voltmeter may be considered to be the reciprocal of the time required to take a reading of an applied step of voltage.

This response time generally ranges $0.1\text{--}10 s$, hence the effective bandwidth ranges $10\text{--}0.1 Hz$. If we assume that the dc open circuit voltage, V_{OC} , must be greater than the Johnson noise from R_{TH} , $R_M \gg R_{TH}$ and R_{TH} is the sole source of noise in the circuit, then we can write:

$$V_{OC} \geq \sqrt{4kTR_{TH}B} RMS V \quad (8.3)$$

This function is plotted for $B = 0.1$ and $10 Hz$ in Figure 8.2. Note that dc voltages below the lines are unresolvable because of noise.

8.2.1 Electromechanical DC Voltmeters

There are many types of dc voltmeters, some in fact, are effective as power line frequency ac voltmeters as well. In this section, we examine the designs of some dc voltmeters, both

analog and digital. Included are D'Arsonval meters, dynamometer meters, capacitor voltmeters, electrometer voltmeters, chopper type nanovoltmeters, and thermocouple voltmeters.

8.2.1.1 D'Arsonval DC Voltmeters

These voltmeters, also known as permanent magnet moving coil voltmeters, are the most common design of electromechanical dc voltmeter. A typical D'Arsonval microammeter meter movement, illustrated in Figure 8.3, is basically a current-to-angular position transducer. Current flowing through the turns of the rectangular coil suspended in the uniform magnetic field in the air gap generates an orthogonal component of force on each conductor on each side (but not the top or bottom) of the coil. The magnetic field is generated by a strong permanent magnet. The resultant torque acts against a helical spring, or in some meters, a flat torsion spring, used to suspend the coil. When the magnetically produced torque equals the spring torque in the steady state, the meter pointer then stops moving.

The force per unit length of a current carrying conductor in a uniform magnetic field is given by the vector equation:

$$dF = I(dI \times B) \text{ N/m} \quad (8.4)$$

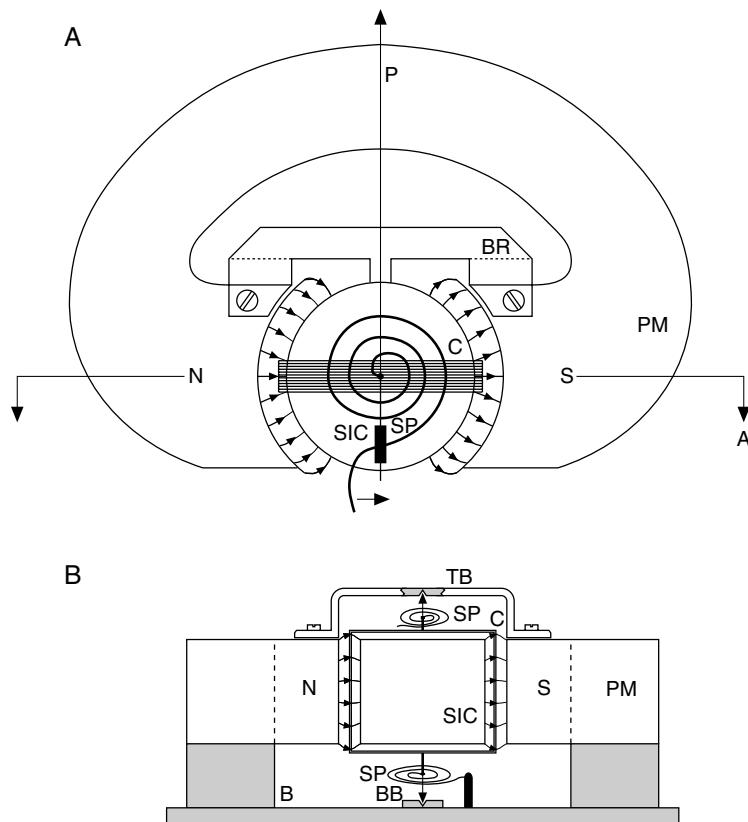
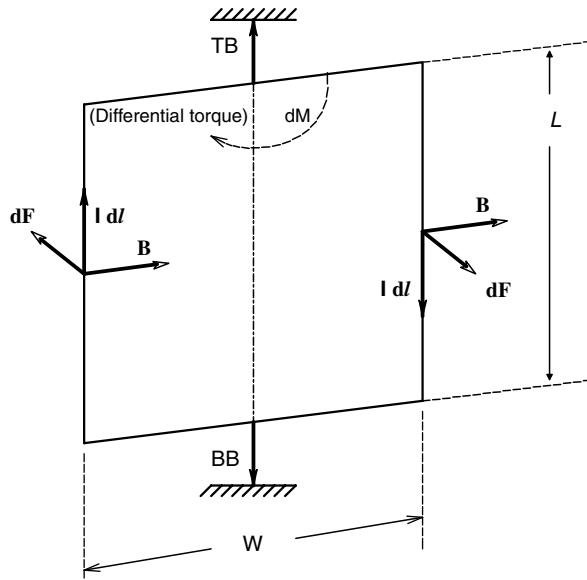


FIGURE 8.3

A. Face view of the innards of a D'Arsonval meter movement. Note: P = pointer, PM = permanent magnet. C = pivoted rectangular coil, SP = helical spring, SIC = soft iron core to concentrate magnetic flux. B. Vertical section (AA') through D'Arsonval meter. Cross hatched material is bakelite. Note: TB, BB = top and bottom bearings, B = base.

**FIGURE 8.4**

Magnetic (Faraday) vector forces on one turn of the rectangular, current carrying coil of a D'Arsonval meter.

If the vector \mathbf{Idl} is rotated into the vector \mathbf{B} , the direction of \mathbf{dF} is given by the right hand screw rule. Accordingly, the magnitude of \mathbf{dF} can be written:

$$\mathbf{dF} = \mathbf{IB} \sin(\varphi) \mathbf{dl} \quad (8.5)$$

Now from Figure 8.4 we see that the length of each side of the coil in the perpendicular magnetic field is L . The angle φ between the current and the \mathbf{B} field is 90° . Also, the radial distance from each side to the pivot is $W/2$, and there are N turns of wire in the pivoted coil. Thus, the mechanical torque generated by passing current through the N turns of the coil in the magnetic field is given by:

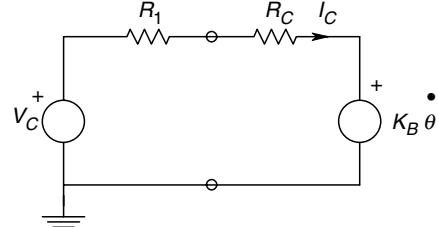
$$T_M = 2 BLNIW/2 = BNAI = K_T I \quad (8.6)$$

where A is the area of the rectangular coil ($A = LW$) and $K_T = BNA$ is the meter's torque constant. Unfortunately, the dynamic behavior of a D'Arsonval meter movement is not simply described by the balance between T_M and the spring torque. If the coil is moving, an EMF is induced in it, given by:

$$E_B = 2 NLB(\dot{\theta}W/2) = NBA\dot{\theta} = K_B\dot{\theta} \quad (8.7)$$

where θ is the angle of rotation (pointer angle) of the coil, $\dot{\theta}W/2$ is the tangential (linear) velocity of a side of the coil and K_B is the coil's back EMF constant. A complete circuit of the D'Arsonval meter coil is shown in Figure 8.5. L_C is the self-inductance of the coil, which is generally negligible at low frequencies. From Figure 8.5 we see that the current in the coil is given by:

$$I_C = (V_C - E_B)/(R_C + R_1) = (V_C - E_B)/R_T \quad (8.8)$$

**FIGURE 8.5**

Equivalent circuit of the D'Arsonval meter coil. The EMF due to coil motion in the magnetic field is $V_B = K_B \dot{\theta}$.

The Newtonian torque balance equation can be written:

$$T_M = K_T I_C = J\ddot{\theta} + D\dot{\theta} + K_S \theta \quad (8.9)$$

or

$$K_T[(V_C - K_B \theta)/R_T] = J\ddot{\theta} + D\dot{\theta} + K_S \theta \quad (8.10)$$

which reduces to:

$$V_C(t) = \ddot{\theta} J R_T / K_T + \dot{\theta} (D R_T / K_T + K_B) + \theta K_S R_T / K_T \quad (8.11)$$

where D is the viscous damping torque constant, J is the moment of inertia of the coil and pointer and K_S is the torque constant of the torsion spring.

When this second order, linear ODE is Laplace transformed, we finally get a transfer function relating pointer angle, θ , with applied voltage:

$$\frac{\Theta(s)}{V_C(s)} = \frac{K_T/(R_T K_S)}{s^2 J / K_S + s[(B + K_T K_B / R_T) / K_S] + 1} \quad (8.12)$$

Note that the mechanical natural frequency of the D'Arsonval meter movement is:

$$\omega_n = \sqrt{K_S/J} \text{ r/s} \quad (8.13)$$

and its damping factor is given by:

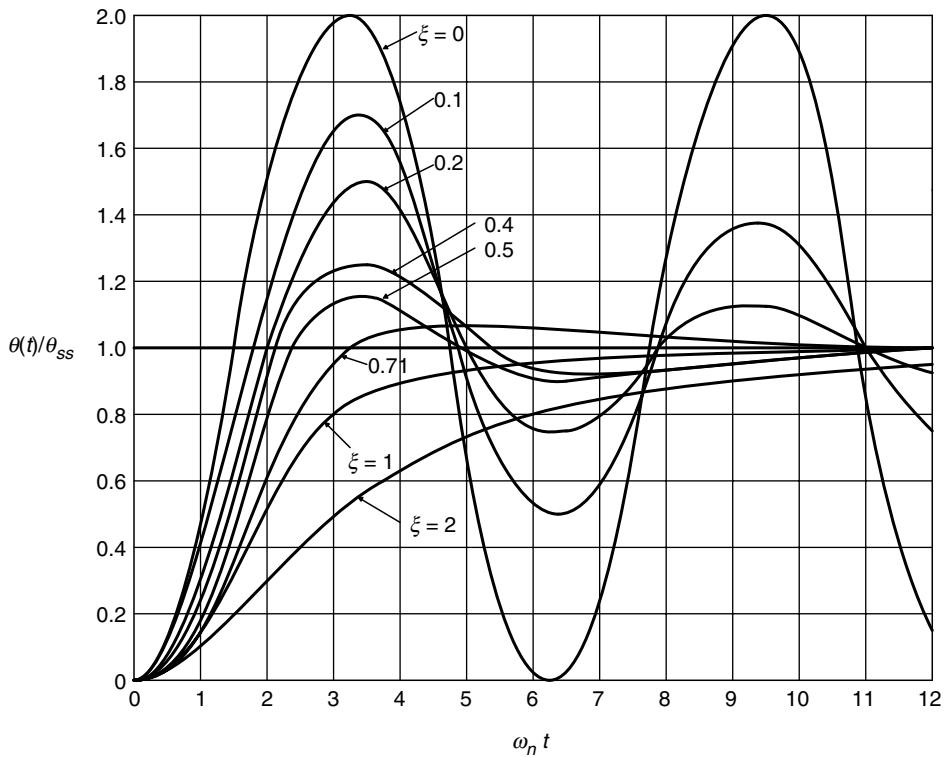
$$\zeta = \frac{D + K_T K_B / R_T}{2\sqrt{J K_S}} \quad (8.14)$$

Equation 8.12 above is a classic example of the low-pass transfer function of a second order, linear electromechanical system. The step response of the D'Arsonval meter is easily found by setting:

$$V_C(s) = V_C/s \quad (8.15)$$

and using a table of Laplace transforms, the meter deflection as a function of time is found to be:

$$\theta(t) = V_C \frac{K_T}{R_T K_S} \left[1 - \frac{e^{-\omega_n t}}{\sqrt{1 - \zeta^2}} \sin \left(\left\{ \omega_n \sqrt{1 - \zeta^2} \right\} t + \tan^{-1} \frac{\sqrt{1 - \zeta^2}}{\zeta} \right) \right] \quad (8.16)$$

**FIGURE 8.6**

Mechanical response of a D'Arsonval meter to an applied step of voltage. Different damping factors are shown. $0.7 \leq \xi \leq 1$ gives the best compromise between rise time and no overshoots.

The mechanical response of a D'Arsonval meter to a step of applied voltage is shown in Figure 8.6. Note that as the damping factor, ζ , varies from 1 to 0, the response becomes more and more oscillatory. Obviously, it is inconvenient to have to wait for a highly underdamped meter to settle down to its steady state reading of:

$$\theta_{ss} = V_C K_T / (R_T K_S) \quad (8.17)$$

Experience tells us that the best dynamic performance at a fixed ω_n for a second order system's step response occurs when the damping factor, ζ , lies somewhere in the range 0.5–0.707.

From the analysis above, we observe several important properties of D'Arsonval meters used as voltmeters. First, they require a steady state current to produce a reading. This current must come from the circuit under test. Practical full scale current is typically 20–50 μA , although special, sensitive D'Arsonval meter movements with torsion spring suspensions have been built with full scale currents of 1 μA or less. To obtain greater voltmeter sensitivity, we see from equation 8.17 that we must increase $K_T = BNA$. Meter size and cost are design considerations. B is set by the state-of-the-art and cost of permanent magnets available. A is set by size considerations and N by the maximum tolerable mass and moment of inertia of the coil. Too high a mass will give problems with bearings and coil suspension, a high J will result in an underdamped transient response and long settling time. Likewise, increased sensitivity reached by making the torsion spring constant, K_S , small will also result in low damping. Second, the total

resistance in the circuit, R_T , is set by the voltmeter's sensitivity and the desired full scale voltage. From the relations above, we see that:

$$V_{FS} = (R_S + R_C)/\eta = R_T/\eta = R_T I_{MFS} \quad (8.18)$$

Obviously, for any given full scale voltage, meter loading by a D'Arsonval analog dc voltmeter will be less with a meter with a high η . Also, for very large V_{FS} , R_T will be very large. For example, R_T for a 500 VFS with a 20,000 Ω/V movement will be 10 M Ω . Multi-range, dc D'Arsonval voltmeters have a switch that switches various R_S s in series with the meter, satisfying equation 8.18 above.

8.2.1.2 The Capacitor, or Electrostatic Voltmeter

This voltmeter is an unusual type of voltmeter, which is best suited for dc or line frequency ac, high voltage measurements. Capacitor voltmeters have been used from about 0.5 kV to over 200 kV with special insulation. A front view schematic diagram of a capacitor voltmeter is shown in Figure 8.7. The design is very much like a parallel plate, radio tuning capacitor. Rotor plates pivot on low friction bearings and are restrained by a linear torsion spring. The capacitance between rotor and stator plates is a function of the rotation angle, θ , (pointer angle) and, in a linear capacitor, assumed here for convenience, can be written as:

$$C(\theta) = C_o + K_C\theta \quad (8.19)$$

Here, C_o is the capacitance between the plates at $\theta = 0^\circ$ and K_C is the capacitance constant. $C(\theta_{MAX})$ ranges from about 20 pF up to several hundred pF, depending on design. It can be shown that at a given voltage, there is an attractive torque acting to rotate the rotor

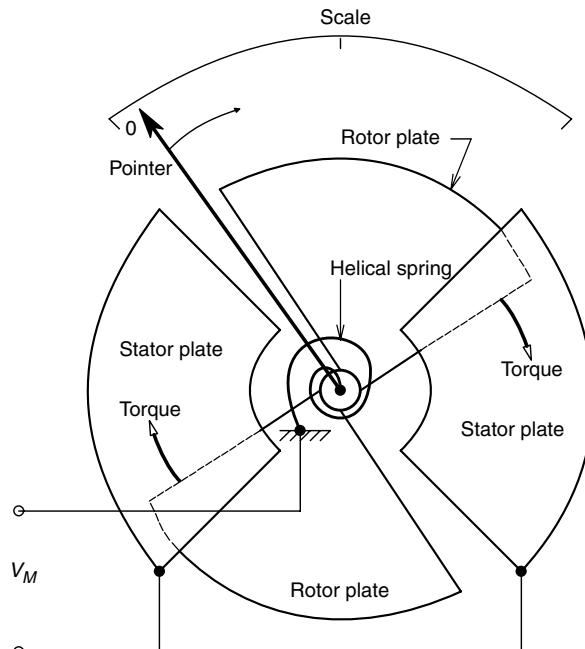


FIGURE 8.7

Transparent face view of an electrostatic, capacitor voltmeter. The electric field between the plates generates the deflection torque. The scale is nonlinear.

into the stator to produce maximum capacitance and maximum stored electric field energy in the capacitance. The energy stored in the capacitor at steady state is:

$$W = \frac{1}{2} C(\theta) V_C^2 \quad (8.20)$$

The torque acting to rotate the rotor into the stator plates may be shown to be given by:

$$T_M = \left. \frac{\partial W}{\partial \theta} \right|_{V_C} = \frac{V_C^2}{2} \frac{dC(\theta)}{d\theta} = \frac{V_C^2}{2} K_C \quad (8.21)$$

The electrostatic derived torque must equal the Newtonian torques:

$$T_M = J \ddot{\theta} + D \dot{\theta} + K_S \theta \quad (8.22)$$

After substituting the expression for T_M into the second order ODE above and Laplace transforming, we obtain the transfer function:

$$\frac{\Theta}{V_C^2}(s) = \frac{K_C/2K_S}{s^2 J/K_S + s D/K_S + 1} \quad (8.23)$$

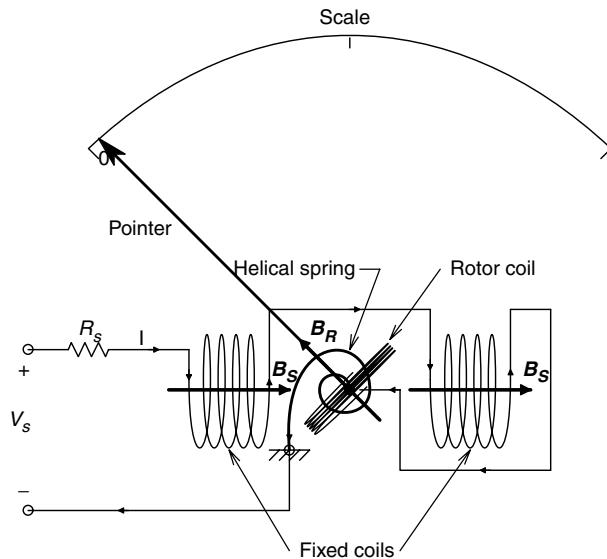
The mechanical properties of the capacitor voltmeter, similar to those of the D'Arsonval meter movement, act as a mechanical low-pass filter above $\omega_n = \sqrt{K_S/J}$ r/s. Hence for steady state conditions,

$$\theta_{SS} = \overline{V_C^2} K_C / 2K_S \quad (8.24)$$

The capacitor voltmeter is an example of a *square law meter movement*. The scale has nonlinear calibration from equation 8.24 and also because $C(\theta)$ is, in general, not linear. After the meter reaches steady state deflection, no current flows in the circuit for an applied dc voltage step. However, if a capacitor voltmeter is suddenly placed directly across an energized, high voltage circuit, it appears initially as a short circuit and heavy current flows into the meter. This may damage the CUT, so often a series resistor is used with the electrostatic voltmeter to limit the initial charging current. A typical electrostatic voltmeter, such as the Sensitive Research model ESD, has a scale calibrated for 1–5 kV, with the most sensitive (expanded) scale at the center of the meter ($\theta_{MAX}/2$) at about 2.75 kV. This meter has a guaranteed accuracy of 1% of full scale voltage, a capacitance of 18 pF at full scale, and an insulation resistance of $10^{15} \Omega$. θ_{MAX} is about 65° for this meter, which is less than the 90° typical of most D'Arsonval meter movements.

8.2.1.3 The Electrodynamometer Meter

This movement is another voltmeter design, which as we will see below, is square law, and can be used for the measurement dc or power line frequency ac voltage, current and power. Here, we will examine the electrodynamometer as a dc voltmeter. Figure 8.8 shows a diagram of the basic electrodynamometer movement. Note that there are two, symmetrically placed stator coils around a rotor coil which, as in the case of other analog meters, is constrained by a torsion spring. When an electrodynamometer meter is connected as a voltmeter, the three coils are wired in series, as shown. Note that initially, the flux density vectors repel each other and create a torque that rotates the pointer

**FIGURE 8.8**

Face view schematic of an electrodynamometer meter movement connected as a voltmeter. The scale is nonlinear.

clockwise. Here, as in the case of the D'Arsonval dc voltmeter, a series resistor is used to limit the meter's full scale current. The energy stored in the magnetic field of the electrodynamometer's coils is given by:

$$W = i_M^2 [L_S/2 + L_R/2 + M_{SR}(\theta)] \quad (8.25)$$

where i_M is the current in the coils, L_R is the self-inductance of the rotor coil, L_S is the self-inductance of the two stator coils together and $M_{SR}(\theta)$ is the mutual inductance between the stator coils and the rotor coil. Note that $M_{SR} = M_{RS}$. Due to meter coil geometry, $M_{SR}(\theta)$ may be approximated by:

$$M_{SR}(\theta) = K_M(\theta - \theta_{MAX}/2) \quad (8.26)$$

Note that at half maximum deflection, the axis of the rotor coil is perpendicular to the axis of the stator coils and $M_{SR} = 0$. The electromagnetically produced torque is given by:

$$T_M = \frac{\partial W}{\partial \theta} \Big|_{i_M} = i_M^2 \frac{dM_{SR}(\theta)}{d\theta} i_M^2 K_M = [V_M/(R_S + R_R + R_1)]^2 K_M \quad (8.27)$$

As in the previous examples, we equate the electromagnetic torque with the Newtonian reaction torques and form a low-pass transfer function:

$$\frac{\Theta}{V_M^2}(s) = \frac{K_M/[(R_S + R_R + R_1)^2 K_S]}{s^2 J/K_S + s D/K_S + 1} \quad (8.28)$$

In the steady state, the meter deflection is:

$$\theta_{SS} = \overline{V_M^2} K_M/[(R_S + R_R + R_1)^2 K_S] \quad (8.29)$$

Here, as in the case of the electrostatic voltmeter, the deflection is square law (i.e., proportional to the mean squared applied voltage). The inertia of the meter movement effectively averages the squared voltage. Electrodynamometer voltmeters are seldom used above power line frequency because the coil inductive reactance begins to become more than 1% of $(R_S + R_R + R_1)$, producing frequency dependent calibration errors. Electrodynamometer voltmeters generally require far more current for steady state, full scale deflection than do D'Arsonval voltmeters. Typical electrodynamometer voltmeter sensitivities range 10–50 Ω/V , or full scale currents range 0.1–0.02 A. It appears that this type of meter is truly a 'wattsucker' and is best suited for measurements on power systems, rather than on electronic circuits. Dynamometer voltmeters, ammeters and wattmeters generally have 80–90° arc scales, with scale lengths of 6.5–7.0 in and a mirror to eliminate parallax when reading the pointer. They are calibrated for dc, or ac, voltages or currents ranging about 25–500 Hz, depending on design. Accuracies are in the range of 0.1–0.25% of full scale reading. Due to the mass of the moving coil, the bearings must be rugged and must have low friction.

8.2.2 Electronic DC Voltmeters

In this section, we will consider the design, applications and limitations of various types of electronic dc voltmeters, beginning at the low end of the voltage scale, with nanovoltmeters which are designed to work with extremely low dc input potentials through to Thevenin resistors on the order of tens of ohms. Figure 8.9 illustrates the design of a *chopper type nanovoltmeter*. A special, low noise electromechanical chopper running at 60 Hz is used to convert the low level dc potential, V_S , to an ac signal, V_2 , which is amplified by an RC amplifier, A_1 , with a gain of -10^4 . The ac signal's frequency, 60 Hz, is above the $1/f$ noise portion of the RC amplifier's input voltage noise

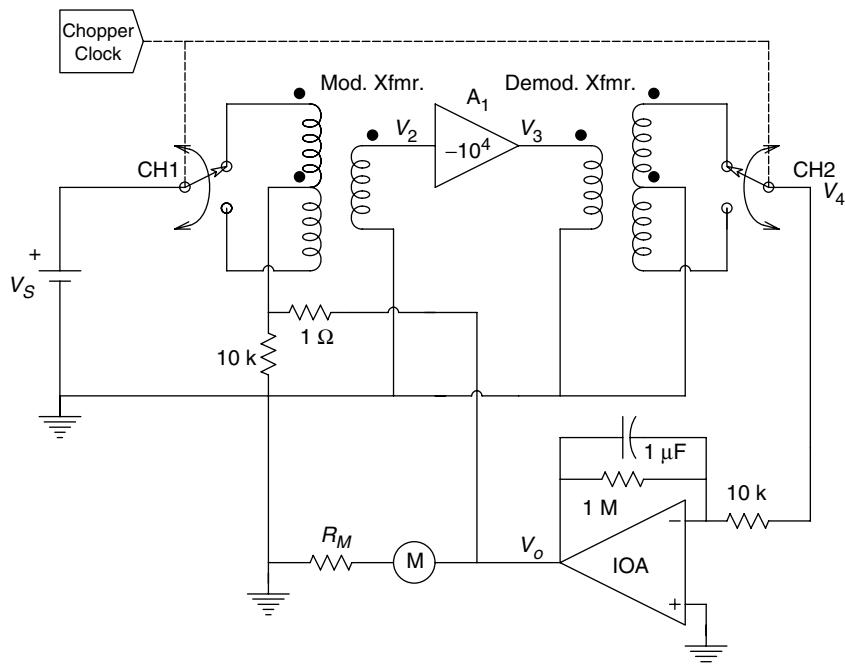
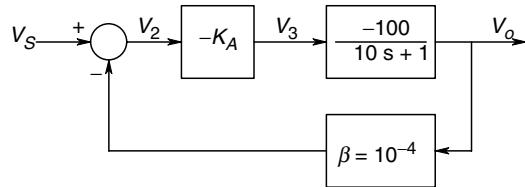


FIGURE 8.9

Schematic diagram of a chopper type, dc nanovoltmeter.

**FIGURE 8.10**

Block diagram of the chopper nanovoltmeter.

spectrum, e_{na}^2 . The equivalent short circuit voltage noise spectrum of the input amplifier is of the form:

$$e_{na}^2(f) = \eta + b/f \text{ MS V/Hz} \quad (8.30)$$

In the flat (white) part of the noise's power density spectrum, η can be as low as 49×10^{-18} MS V/Hz (Toshiba 2SK146 JFET). Following amplification by A1, V_3 is synchronously demodulated by a second chopper. The output of the second chopper, V_4 , has a dc component which is conditioned by the op-amp low-pass filter, A2. A block diagram summarizing these operations is shown in Figure 8.10. With the feedback resistors shown, the nanovoltmeter's transfer function can be written as:

$$\frac{V_o}{V_S}(s) = \frac{(-10^4)[-100/(10s + 1)]}{1 - [-(10^6)(10^{-4})(10s + 1)]} = \frac{9.901 \times 10^3}{s10/101 + 1} \quad (8.31)$$

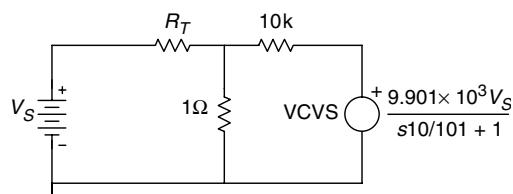
Thus, there is a low noise amplification of the dc input voltage by 79.9 dB, and single pole, low-pass filtering, with a break frequency at 1.61 Hz. When operating on the lowest range (typically 10 nV full scale) it is necessary to further amplify V_o to a level where an analog D'Arsonval microammeter can serve as an output indicator. An additional dc gain stage of 10^2 will boost a 10 nV input to a 10 mV output. Now, if a 20 μ A full scale, dc, D'Arsonval meter is used, it will read full scale if its internal resistance is 500 Ω (or less).

The input resistance of the nanovoltmeter can be calculated using the equivalent input circuit of Figure 8.11. At dc, R_{IN} can easily be shown to be:

$$R_{IN} = (R_T + 1)101.01 \Omega \quad (8.32)$$

where R_T is the dc series resistance of either side of the input chopper's coil. The dc input resistance of the nanovoltmeter amplifier may be on the order of 1 k Ω because of the action of the negative feedback.

The specifications of two commercially available nanovoltmeters are outlined below. The Keithley Model 148 nanovoltmeter has a design similar to that described above. Its output is read on a center scale analog microammeter. Full scale ranges begin at 10 nV and go to 0.1 V in 18 overlapping ranges. On the 10 nVFS range, the input resistance is over 1 k Ω , and the dc nV source must not have a Thevenin resistance of over 10 Ω for noise and loading reasons. On the 0.1 V_{FS} range, the input resistance is greater than 1 M Ω and the source resistance should be less than 10 k Ω . Resolution is better than 1 nV on the 10 nV range and accuracy is $\pm 2\%$ of full scale on all ranges. The line frequency rejection is 3000:1 on the 10 nV range and the CMRR is 160 dB at 60 Hz on all ranges.

**FIGURE 8.11**

Equivalent input circuit of the chopper nanovoltmeter.

The Keithley Model 181 nanovoltmeter has a 6 1/2 digit, digital readout and is IEEE-488 bus compatible. It uses a special low noise JFET headstage to obtain an input resistance of over $10^9 \Omega$ on the low ranges. There are seven full scale ranges, 2 mV–1 kV. Meter noise is claimed to be less than 30 nV peak-to-peak on the 2 mV scale with the output low-pass filter on. The low-pass filter is a 3-pole digital design, with an equivalent time constant of 0.5, 1 or 2 seconds, depending on meter range. This meter can be run with 5 1/2 digit resolution when inputs are noisy.

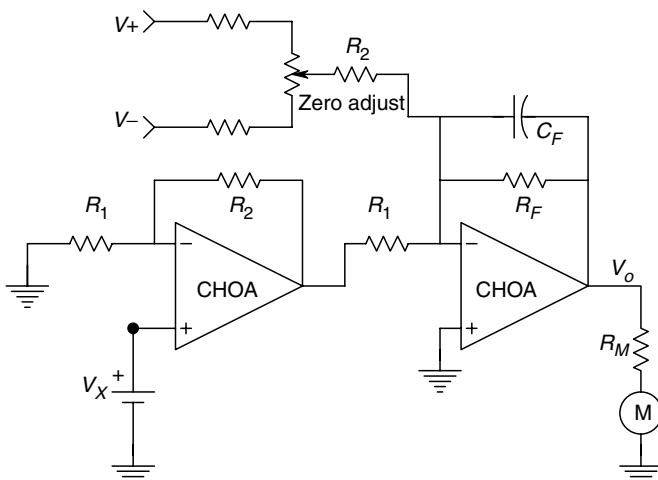
Due to the special challenge in measuring dc voltages in the nanovolt range, meters must be allowed to warm up at least an hour before use to reach thermal equilibrium. The dc nanovolt measurements are confounded by factors which are normally neglected in high level measurements. These include the generation of thermoelectric (Seebeck) EMFs by junctions between dissimilar metals at different temperatures, the induction of low frequency EMFs by time varying magnetic fields whose flux passes through the circuit from the meter to the source, low frequency and dc potentials due to ground loops, low frequency EMFs due to triboelectric effects on connecting coaxial cables, and piezoelectric EMF artifacts. A good discussion of sources of error in low level, dc measurements can be found in the Keithley handbook on *Low Level Measurements*, 5th edition.

In order to operate either an analog meter or a digital interface, the dc voltage under measurement must be amplified to a level useful to drive the analog indicating meter (about 100 mV) or the analog-to-digital converter (1–10 V). This means that microvolt level signals need to be amplified by factors of 10^5 – 10^7 . Inherent in all dc amplification is the addition of noise from the amplifier, generally consisting of a mixture of white noise and $1/f$ noise, dc offset and drift due to temperature changes. It was seen in Section 3.7 that the noise performance (as well as the dc drift performance) of a high gain amplifier is set by the headstage which must have a gain greater than five. Thus the amplifiers used in any sensitive, low level dc voltmeter must have low noise, and also very low dc drift and offset voltage. Op-amps suitable for high gain, dc signal conditioning include chopper stabilized and commutating auto zero amplifiers, the latter being described in Section 2.5.3.

The Intersil ICL7600/ICL7601 commutating autozero op-amp is well suited for dc voltage amplification system headstages. The ICL7600 has a low, input offset voltage of $2 \mu\text{V}$ a very low input offset voltage drift of $0.2 \mu\text{V}/\text{year}$ and a very low input offset voltage tempco of $\pm 0.005 \mu\text{V}/^\circ\text{C}$ and a dc input bias current of 300 pA . The switching (commutating) frequency of the ICL7600 system capacitors can be set in the range 160–5200 Hz.

Another approach to low noise, drift-free dc amplification is to use a chopper stabilized op-amp, such as the ICL7650. This op-amp has a dc input offset voltage of $1 \mu\text{V}$ over the operating temperature range, an offset voltage tempco of $\pm 0.01 \mu\text{V}/^\circ\text{C}$, a dc input bias current of 35 pA , a dc open loop gain of 134 dB and a gain bandwidth product of 2 MHz. The internal chopping frequency is 200 Hz. Many other manufacturers offer integrated circuit, chopper stabilized or auto zero dc amplifiers. Tempcos of $50 \text{ nV}/^\circ\text{C}$ are pretty much state-of-the-art and input dc offset voltages range 1 – $5 \mu\text{V}$.

A low-pass filter is generally used at the output of the dc amplifier stages, before the ADC or the DVM module. If a D'Arsonval analog meter is used, the meter movement itself acts as an electromechanical low-pass filter (equation 8.11). However, additional low-pass filtering may be required to improve the dc microvoltmeter's resolution. A simplified schematic of a dc voltmeter with electronic amplification is shown in Figure 8.12. If an analog meter is used, precision is seldom more than $\pm 0.25\%$ of full scale because of the visual uncertainties of reading the pointer position. Properly designed digital meters can commonly reach precisions of $\pm 0.005\%$, or 50 ppm. Noise, dc drift

**FIGURE 8.12**

Schematic diagram of an analog, dc millivoltmeter. Two, chopper stabilized, op-amps are used for amplification. The second op-amp is used as a low-pass filter and for zeroing. The meter is a dc D'Arsonval microammeter.

and quantization error are still the ultimate determiners of resolution in instruments with digital readouts, however.

Electronic dc voltage measurements above the single volt range generally do not require amplification, either for an analog indicating meter or a digital output indicator. In fact, electronic voltmeters used in the range 10–1,000 V generally make use of internal input voltage dividers to attenuate the voltage under measurement to a value useful to the meter amplifier or ADC. An external resistive voltage divider probe is generally used to measure dc voltages from 1 kV to 40 kV. Such a probe makes an external voltage divider with the dc voltmeter's input resistance. For example, the Fluke model 80K-40 high voltage probe is designed to be used with any dc voltmeter having an input resistance of $10\text{ M}\Omega$. It makes a 1000:1 voltage divider, so 1 kV applied to the probe gives 1 V at the meter. The probe and meter present a $1000\text{ M}\Omega$ load to the high voltage circuit under test, hence the probe and meter draw $1\text{ }\mu\text{A}/\text{kV}$ from the CUT.

At dc voltages above 40 kV, special measurement techniques must be used because of insulation problems associated with the resistive voltage divider. At very high voltages, resistors may behave erratically due to the generation of corona discharge, or leakage due to the absorption of water from the atmosphere. In fixed installations, these problems can be overcome by using proper insulators and couplings, and housing the resistors in an inert, insulating gas such as SF_6 or Freon, or in high pressure (≥ 10 atm), dry air or nitrogen.

In order to measure dc voltages in excess of 200 kV electronically, use is often made of electrometer based, electric field measuring instruments. Such techniques make use of the fact that the geometry is fixed in super high voltage systems, hence the electric field geometry is fixed, and at any point in the field, the field strength is proportional to the dc voltage generating the field. We discuss instruments used to measure electric fields in the next section.

8.3 Static Electric Fields and the Potential of Charged Surfaces Measurement

Static (dc) electric fields can arise from a variety of physical and electrical causes, which include the presence of bound or mobile surface charges (electrons, ions) on conductors

or insulators. The surfaces can become charged by being bombarded with charged particles such as sand or raindrops, or from triboelectric effects where charges become separated from the surfaces of insulators when they rub on conductors or other insulators. Of course, static fields exist around conductors operating at high dc voltages with respect to ground and also near the faces of certain monitors using cathode ray tubes.

Static charges and fields can create problems ranging from annoying ('static cling') to extreme danger (static discharge sparks ignite solvent vapors or combustible dust, causing an explosion). Other problems caused by static electricity include the attraction of fine dust particles to clean, charged surfaces, attraction of charged dust particles to clean, grounded surfaces, the physical attraction of charged materials such as paper or plastic film to other objects and when the electric field strength exceeds a critical value (about 3×10^6 V/m in air at STP), there can be an abrupt discharge of charged persons or objects to static-sensitive semiconductor devices or electronic equipment, causing damage. Clearly, it is important to be able to measure the dc voltages associated with the production of dc fields, to give warning of potentially damaging or dangerous conditions (pun intended).

The first method of measuring the dc potential of a charged object field makes use of a dc, electrometer voltmeter connected to a special probe as shown in Figure 8.13 A. To measure the potential of the object in question, which may be a charged insulating surface or conductor, the probe is brought to a standard distance, d , from the object. The standard distance is required to set the capacitance, C_2 , to a known value. Figure 8.13B shows the equivalent circuit for the dc electrometer field meter. To measure V_S , capacitance C_1 is initially shorted to ground, and a grounded metal plate is used to cover the proximal plate of capacitor C_2 . (C_1 is the input capacitance of the electrometer voltmeter plus the capacitance to ground of the wire connecting the proximal plate of C_2 to the electrometer input. C_2 is the capacitance between the proximal plate of C_2 and the charged object.) Next, the metal plate is slid aside, fully exposing the plate of C_2 . C_2 and C_1 are in series, and acquire charge from the field from the potential on C_3 . Initially, if C_3 is isolated and is at potential V_S , some charge flows from C_3 into C_2 and C_1 , causing the voltage on C_3 to drop to V'_S . In the steady state, this can be written as:

$$(V_S - V'_S)C_3 = V'_S \frac{C_1 C_2}{C_1 + C_2} \quad (8.33)$$

The new voltage on the isolated object with capacitance C_3 is found to be:

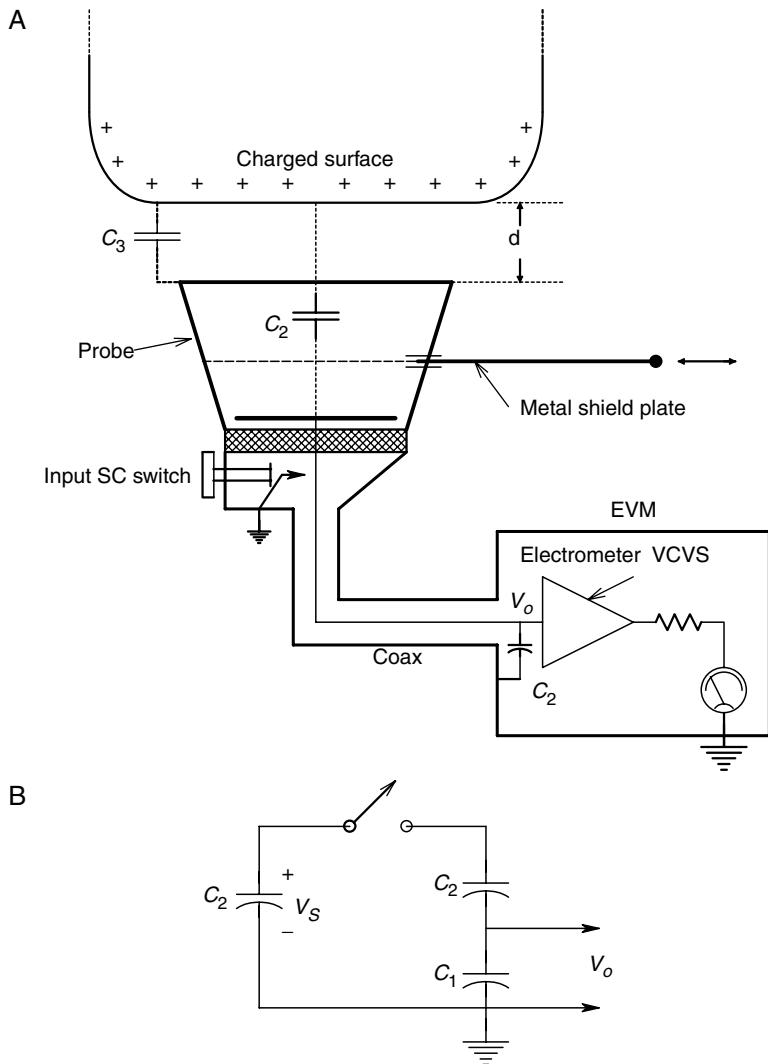
$$V'_S = \frac{V_S C_3}{\frac{C_1 C_2}{C_1 + C_2} + C_3} \quad (8.34)$$

Now, because the same current flows in C_3 , C_2 and C_1 , the charge that resides in C_2 is also present in C_1 . Thus from the equation

$$V_o C_1 = (V'_S - V_o) C_2 \quad (8.35)$$

we find that:

$$V_o = \frac{V'_S C_2}{C_1 + C_2} \quad (8.36)$$

**FIGURE 8.13**

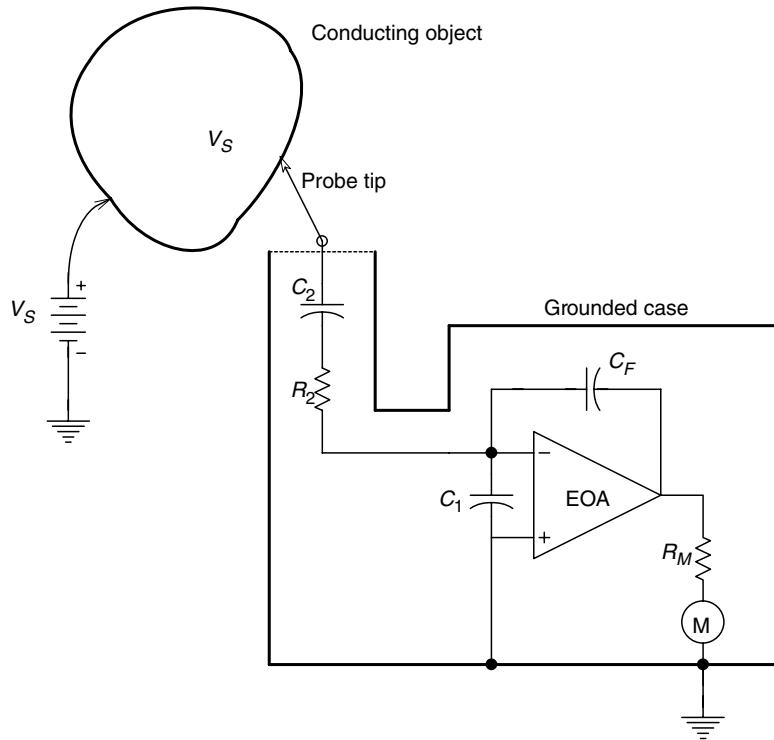
A. Section through a capacitance probe used to measure high, dc, electrostatic potentials on charged objects. The probe acts as a voltage divider. B. Equivalent circuit of the high voltage capacitance probe.

If the expression for V_S' in equation 8.34 is substituted in equation 8.36 above, we finally obtain:

$$\frac{V_o}{V_S} = \frac{C_2 C_3}{C_1 C_2 + C_3(C_1 + C_2)} \quad (8.37)$$

If $C_3(C_1 + C_2) \gg C_1 C_2$ then equation 8.37 reduces to

$$\frac{V_o}{V_S} \cong \frac{C_1}{C_1 + C_2} \quad (8.38)$$

**FIGURE 8.14**

An electrometer op-amp is used as a charge amplifier in this circuit to measure high dc voltages.

For example, if $C_2 = 10^{-11} \text{ F}$ and $C_1 = 10^{-7} \text{ F}$, then $V_o = 10^{-4} V_S$. Hence, an initial 10 kV charge on C_3 will result in a 1 V reading on the electrometer.

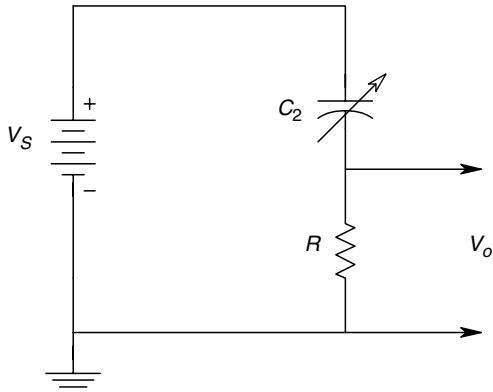
In the system described above, the value of C_1 depends on the capacitance to ground of the cable connecting the probe with the electrometer, as well as the input capacitance of the electrometer. In order to eliminate the system's requirement for a fixed input geometry to determine C_1 , the charge amplifier configuration shown in Figure 8.14 can be used. Now C_1 appears between the summing junction, which is at virtual ground, and actual ground. Hence, negligible current flows through C_1 , and C_F determines the output voltage of the system. R_2 limits the initial charging current through C_2 and C_F , so that the electrometer op-amp will not go into current saturation. Now, it is easy to show that the charge amplifier's output is:

$$V_Q = -V_S(C_2/C_F) \quad (8.39)$$

We wish C_2/C_F to be 10^{-3} or 10^{-4} , so the op-amp will not saturate for large V_S .

Another approach to the problem of measuring dc voltages in the kilovolt range through their fields is to use a time modulated C_2 and to measure the ac current which flows in resistor R , shown in Figure 8.15. In this form of dc, high voltage meter, the modulation of C_2 may be expressed mathematically as

$$C_2(t) = C_{20} + \Delta C \sin(\omega t), \quad \Delta C \ll C_{20} \quad (8.40)$$

**FIGURE 8.15**

Equivalent circuit of a field mill, dc voltmeter. The value of C_2 is modulated sinusoidally around an average value.

There are several ways in which C_2 can be modulated. One common method is to have the blades of a grounded, rotating chopper wheel alternately cover and uncover the proximal plate of C_2 . Another method physically modulates the distance between the proximal plate and the charged surface. As a result of the modulation of C_2 , an ac current of frequency ω will flow through resistor R to ground. This current may be found by writing a loop equation for the system of Figure 8.15:

$$V_S = \dot{q}R + q/C_2(t) \quad (8.41)$$

or

$$V_S = \dot{q}R + q(1 - \Delta C/C_2)/C_{20} \quad (8.42)$$

The ODE of equation 8.42 can be put in the standard form:

$$\dot{q} + q \frac{1}{C_{20}R} \left(1 - \frac{\Delta C \sin(\omega t)}{C_{20}} \right) = V_S/R \quad (8.43)$$

This ODE has the standard solution:

$$y \exp\left(\int P dx\right) = \int (\exp\left(\int P dx\right) Q dx + c) \quad (8.44)$$

in which $x = t$, $y = q$, $Q = V_S/R$, and $P(t) = [1 - \Delta C \sin(\omega t)]/RC_{20}$. Performing the integrations indicated in the standard solution, we find that:

$$q(t) = \frac{V_S C_{20}}{1 - [\Delta C \sin(\omega t)]/C_{20}} + c \quad (8.45)$$

The steady-state sinusoidal current in the series circuit is:

$$i_{SS}(t) = \dot{q} = V_S \Delta C \omega \cos(\omega t) \quad (8.46)$$

Thus, the ac voltage across R is proportional to V_S :

$$v_o(t) = V_S R \Delta C \omega \sin(\omega t) \quad (8.47)$$

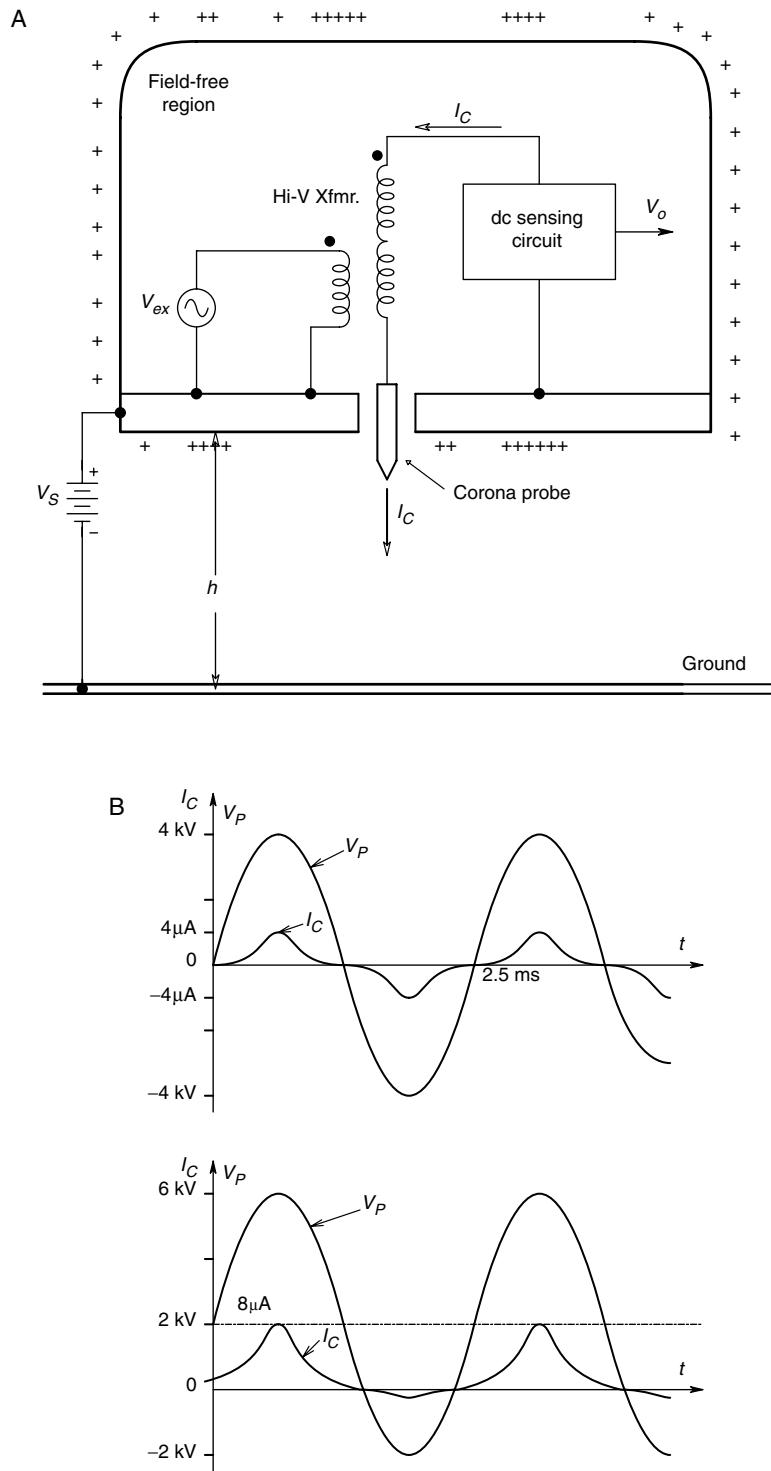
For example, if $\omega = 100 \text{ r/s}$, $V_s = 100 \text{ kV}$, $R = 10^5 \Omega$ ohms and $\Delta C = 10^{-11} \text{ F}$, then the peak value of v_o is 10 V. Of course, the input resistance of the ac voltmeter used to measure v_o should be over $10^8 \Omega$.

The two methods of measuring the high dc voltage, V_s , described above do not require direct contact with the charged surface. However, their accuracy is limited by the accuracy in determining either C_2 or ΔC , which in turn depends on accurate determination of the distance from the proximal plate of C_2 to the charged surface. At least one commercially available 'digital field meter' designed to have a working distance of 4.0 in from the charged surface, uses an ultrasonic ranging system similar to that used in popular cameras which use self-developing, 'instant' film, to permit precise adjustment of the working distance. The Semtronics model EN 235 Auto-DFM reads $0 \leq V_s \leq 20 \text{ kV}$ with $\pm 5\%$ accuracy of the reading at 4 in distance. Correct distance is indicated by LEDs. Other field voltmeters require a physical measurement of their working distance, presumably with an insulated ruler. Voltage ranges of some portable, battery operated field meters can have ranges of 0–500 V to 0–200 kV. Such meters are typically used to measure static charges accumulated in industrial processes such as paper and plastic sheet manufacturing.

Another severe static electricity problem occurs when helicopters fly through charged particles such as sand or water droplets. The metal airframe of the helicopter can acquire a potential of thousands of volts. If not discharged before landing, this charge can be dangerous in fueling and cargo handling operations. Bradford (1975) described a system which automatically senses the magnitude and sign of the helicopter's charge, and automatically discharges the airframe by spraying charged droplets of isopropyl alcohol from the helicopter's fuselage. The helicopter's fuselage has a capacitance of about 2,200 pF when on the ground. Its body capacitance when hovering at 25 m is about 500 pF. To measure the helicopter's body potential, Bradford used a corona field sensor manufactured by Dayton Aircraft Products. A block diagram of the Dayton sensor is shown in Figure 8.16A. A 400 Hz ac voltage is amplified to 4000 peak volts and applied to a wire brush corona probe which protrudes through a hole in the fuselage into the high dc field region next to the fuselage. As seen in Figure 8.16B, if the fuselage is at zero potential with respect to ground, the ac current due to corona flowing from the probe is symmetrical and has zero average. If the helicopter body is at some positive potential, V_s , the dc field from V_s causes extra corona current to flow on the positive cycles and less to flow during the negative cycles, giving the corona current an average (dc) component which is found to be proportional to E , hence V_s . The sensor electronics and static charge discharge controller were located inside the fuselage where, of course, the electric field is zero.

Calibration of the field sensor was made difficult by the fact that helicopter body capacitance decreases nonlinearly with hover height. As this decrease occurs, the body potential increases if the charge remains constant. Offsetting this increase in V_s is the fact that the field strength in the vicinity of the corona probe tends to decrease from geometrical considerations as altitude increases. Bradford (1975) used a large, parallel plate capacitor (plate separation of 1 foot) to generate a uniform field to calibrate the corona probe field sensor. He obtained a conditioned sensor (dc) output of about 1 V per kV on the capacitor.

It is generally difficult to measure an actual electric field's strength in V/m without disturbing the field with the measurement apparatus. This is true of field measurements in conducting media as well as in free space. If the measurement sensor is a dielectric with a low dielectric constant, and not grounded, the disturbance of the field will be minimal.

**FIGURE 8.16**

A. Diagram of a Dayton Aircraft corona field sensor, used to measure the potential accumulated on helicopter fuselages. B. Current waveforms in the corona field sensor. In the figure on top, $V_S = 0$, while in the figure at the bottom, $V_S = 2$ kV.

There are several, possible electro-optical approaches to the problem of measuring dc electric fields. One is to measure the change in the linear polarization of light transmitted through a quartz crystal subject to an electric field applied across a preferred axis. This phenomenon is called the *electrogyration effect* [Rogers, 1977]. Another electro-optical method proposed for measuring dc electric field strength makes use of the *Pockels effect*. Certain crystalline substances, such as ammonium dihydrogen phosphate (ADP), potassium dihydrogen phosphate (KDP), cuprous chloride (CuCl), cadmium telluride (CdTe) and gallium arsenide (GaAs) exhibit the Pockels effect, in which application of an electric field along the *privileged axis* of the crystal causes changes in the refractive indices along two, orthogonal axes (x and y).

To examine how the Pockels effect might be used to measure an electric field, we orient a Pockels crystal so that the privileged axis is perpendicular to the gradient of the field to be measured. An internal E field is set up in the crystal, which is proportional to the external field we wish to measure. The internal electrical field, if correctly aligned, causes a symmetric change in the optical index of refraction along the x and y axes. Circularly polarized light is directed at the input surface of the crystal and propagates within the crystal along the z axis. The light is acted on by the birefringence induced by the internal E field and emerges elliptically polarized. The emergent light is passed through a linear polarizer analyzer plate which is oriented at 45° with respect to the x and y axes. The intensity of light from the output of the linear polarizer can be shown to be a linear function of the crystal's internal E field. Referring to Figure 8.17, we see that the x and y components of the electric vector of the circularly polarized input light at the entrance face of the crystal can be written as:

$$E_{Xi} = E_i \cos(\omega t) \quad (8.48a)$$

$$E_{Yi} = E_i \cos(\omega t - \pi/2) \quad (8.48b)$$

Now the refractive indices of the Pockels crystal are affected by its internal electrical field, E , according to the model:

$$n_X = n_0 + K_P E \quad (8.49a)$$

$$n_Y = n_0 - K_P E \quad (8.49b)$$

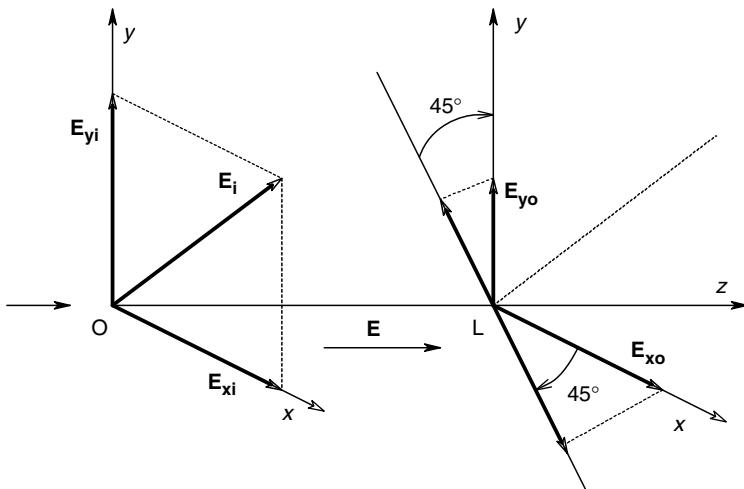


FIGURE 8.17

Vector diagram relevant to the use of a Pockel's cell to measure electric fields.

After propagating a distance L meters through the crystal, the x and y components of the exiting light ray's \mathbf{E} vector can be written:

$$E_{XO} = E_i \cos(\omega t + \phi_X) \quad (8.50a)$$

$$E_{YO} = E_i \cos(\omega t + \phi_Y - \pi/2) \quad (8.50b)$$

The phase lags due to orthogonal changes in propagation velocity are assumed to be linearly proportional to the internal \mathbf{E} field and can be written as:

$$\phi_X = 2\pi L n_X / \lambda = (2\pi L n_0 / \lambda + 2\pi L K_P E / \lambda) = \phi_0 + 2\pi L K_P E / \lambda \quad (8.51a)$$

$$\phi_Y = 2\pi L n_Y / \lambda = (2\pi L n_0 / \lambda - 2\pi L K_P E / \lambda) = \phi_0 - 2\pi L K_P E / \lambda \quad (8.51b)$$

Propagation through the linear polarizer selects those lightwave \mathbf{E} field components along the line, LAN, which define the polarizer's preferred axis. Thus, the lightwave \mathbf{E} field amplitude along the LAN axis exiting the polarizer can be written as:

$$E_{AN} = E_{XO} / \sqrt{2} - E_{YO} / \sqrt{2} = (E_i / \sqrt{2}) [\cos(\omega t + \phi_X) - \cos(\omega t + \phi_Y - \pi/2)] \quad (8.52)$$

We note that the second term in equation 8.52 is of the form $\cos(\alpha - \pi/2)$ which, by trigonometric identity can be written as $\sin(\alpha)$. It is well-known that the intensity of an EM plane wave is related to its \mathbf{E} field peak amplitude by:

$$I_{AN} = E_{AN}^2 (\varepsilon_0 c / 2) \quad (8.53)$$

By substituting equation 8.52 into equation 8.53, and assuming that $\sin(x) \approx x$ for small x , we find that the average intensity is given by:

$$I_{AN} = \frac{E_i^2}{2} \left[1 + \frac{4\pi L K_P E}{\lambda} \right] (\varepsilon_0 c / 2) \quad (8.54)$$

In the foregoing analysis, we have assumed that there was zero neutral density and reflection losses in E_i and that $2\pi L K_P E / \lambda \ll 0.25$ radians. Other detection schemes are possible, but the important result is that output intensity varies linearly with the Pockel crystal's internal \mathbf{E} field. For a numerical example, consider a 1.0 cm crystal of potassium dihydrogen phosphate (KDP, KH_2PO_4). We evaluate KDP's sensitivity to be:

$$\frac{4\pi L K_P}{\lambda} = \frac{4\pi \times 0.01 \text{ m} \times 29 \times 10^{-12} \text{ m/V}}{632.8 \times 10^{-9} \text{ m}} = 5.7576 \times 10^{-6} \text{ m/V} \quad (8.55)$$

Thus, an internal field on the crystal of 10^4 V/m or 10 V/mm would produce a 5.76% increase in the detected intensity. A dc field measuring instrument based on the Pockels effect would need a means of correcting for intensity changes in the laser light source.

8.4 DC Current Measurements

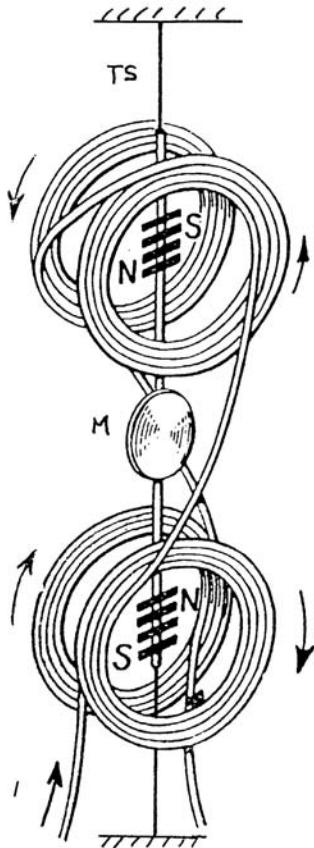


FIGURE 8.18

Diagram of Lord Kelvin's four-coil, permanent magnet, mirror galvanometer.

The technology of analog dc current measurements may be traced back to the early 19th century [Drysdale, Jolley and Tagg, 1952]. Oersted discovered in 1819 that a wire carrying dc current, when held near a compass needle, caused the compass to deflect, implying that the current carrying wire generated a magnetic field. This effect was put to use by Lord Kelvin, who, in 1858, developed a four-coil mirror galvanometer. In Kelvin's instrument, shown in Figure 8.18, four coils were arranged in two opposing pairs; their fields acted on two arrays of permanent magnets with opposing polarities, attached to a vertical torsion spring suspension, to which was also attached a small mirror. As the magnetic fields of the current carrying coils interacted with the magnetic fields of the suspended permanent magnets, torque was developed causing rotation of the suspension and mirror. A collimated beam of light directed at the mirror was reflected onto a distant scale. Even a slight rotation of the suspension is seen as a linear deflection of the spot of light. Galvanometers of this type were used more to detect very small currents (on the order of nA), rather than to measure large currents accurately. Such galvanometers found a major application as null detectors for dc Wheatstone and Kelvin bridges, as well as for dc potentiometers. In 1836, Sturgeon employed a coil suspended in the field of a permanent magnet as a galvanometer. This design evolved and in 1882, D'Arsonval built a moving coil, permanent magnet, mirror galvanometer with a torsion spring suspension (Figure 8.19). Notable in the D'Arsonval design was the use of a soft iron cylinder inside the rectangular coil to concentrate the magnetic

flux perpendicular to the sides of the coil. In 1888, Weston developed a similar permanent magnet, moving coil microammeter movement in which the coil was suspended by pivot bearings and the restoring torque was provided by a pair of helical springs. A pointer was attached to the moving coil. The Weston version of the D'Arsonval meter is what is widely used today as the universal, dc microammeter movement which has applications in all analog multimeters and in most instruments with analog readouts. D'Arsonval type dc microammeters using torsion springs, called taught-band meters, are also widely used because of their ruggedness; they have no jeweled bearings which can be damaged by mechanical shocks.

8.4.1 Electromechanical DC Ammeters

Analog dc ammeters are generally of two types, the D'Arsonval permanent magnet moving coil design or the electrodynamic movement. Measurement of currents in the range of 1–10,000 A with either dynamometer or D'Arsonval meter movements

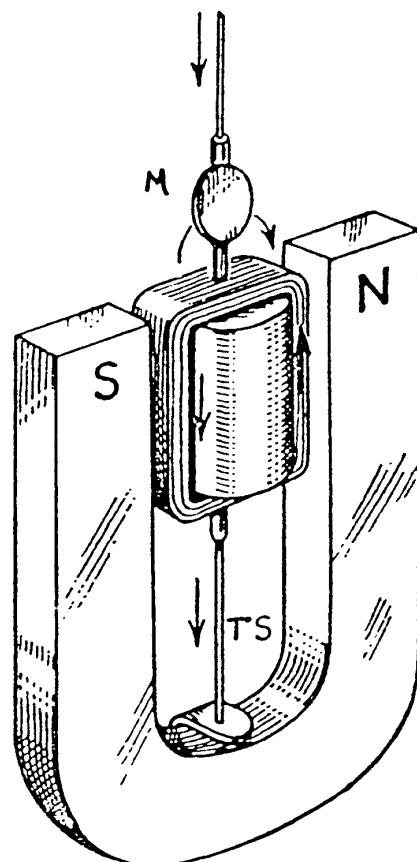
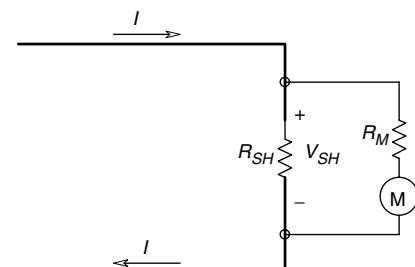
**FIGURE 8.19**

Diagram of D'Arsonval's mirror galvanometer. The coil was suspended by the upper and lower torsion springs. Note: M = mirror, TS = torsion spring.

generally requires an external shunt resistance. A shunt is a precision resistor of very low value designed so that, by Ohm's law, when the full scale current for the meter is flowing, the voltage drop across the resistor and the meter is 50 mV. The lead resistances to the meter are generally not important when measuring I_{FS} in excess of 10 A. In cases where I_{FS} is less than 10 A and an insensitive dynamometer meter movement is used, the meter lead resistance (from shunt to meter) may be important and a specified set of meter leads with known resistance must be used to obtain readings with rated accuracy. The circuit for an ammeter using a shunt is shown in Figure 8.20.

Analog dc ammeters can be subdivided into benchtop (precision) meters and panel meters which are generally used in instruments, systems and communications equipment. Some D'Arsonval microammeters have been miniaturized into 1 in diameter bodies for use as tuning and signal strength meters in radio equipment, tape recorders, etc.

**FIGURE 8.20**

Circuit of a shunt ammeter. The meter is a dc microammeter.

We note that the same basic dc meter movements, which can be used for analog dc voltmeters, are suitable for use as dc ammeters with appropriate shunts. When full scale dc currents less than $20\ \mu\text{A}$ are to be measured, it is generally necessary to employ electronic amplification, discussed in the following section.

8.4.2 Electronic DC Ammeters

To make measurements in the range from 1 fA ($10^{-15}\ \text{A}$) to $50\ \mu\text{A}$, it is necessary to employ dc amplification of some sophistication. In general, current measurements in the range from $10^{-8}\text{--}10^{-15}\ \text{A}$ are made with an *electrometer picoammeter*. Electrometer picoammeters find application in measuring the low current outputs from photomultiplier tubes and ion chambers, as well as specialized measurements in the semiconductor area, etc. As in the case of low level voltage measurements, one fundamental factor limiting precision is the thermal noise from the Norton equivalent conductance seen in parallel with the dc current source being measured. Noise from the electrometer amplifier also confounds resolution. Both e_{na} and i_{na} input noise root power spectrums can be significant in reducing instrument resolution, as well. In addition to problems with noise, there can be measurement errors caused by instrument dc bias current and offset voltage.

Figure 8.21 illustrates the two types of picoammeter circuits commonly used. In Figure 8.21A we see a simple shunt picoammeter in which most of the current to be measured flows through R_{SH} . Ordinarily, $R_{SH} \ll 1/G_N$, so $V_1 = I_S R_{SH}$. V_1 is amplified

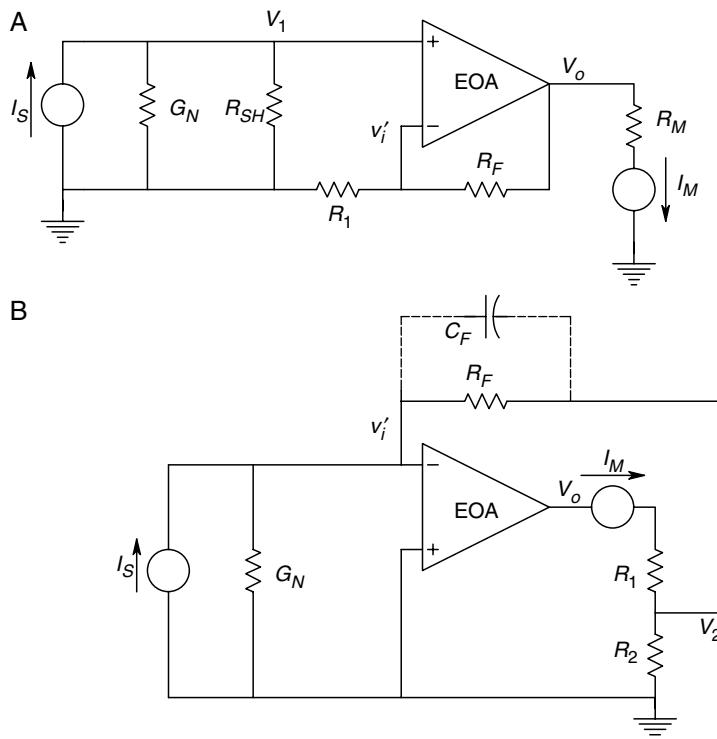


FIGURE 8.21

A. Circuit of an electronic dc ammeter. The op-amp amplifies the voltage developed across the shunt. B. An op-amp used as a transresistor. The summing junction appears at virtual ground.

by the electrometer op-amp (EOA) connected as a non-inverting, dc millivoltmeter. In Figure 8.21B, the dc current under measurement is the input to an operational transresistor circuit, the output of which can easily be shown to be $V_o = -I_S R_F$. To obtain greater sensitivity, it would appear that R_F should be increased. There is a practical limit to the size of R_F , set by its cost and accuracy. Resistors above $10^8 \Omega$ are expensive and as their size increases, their accuracy decreases. Also, when $R_F > 10^9 \Omega$, the effects of shunt capacitance act to slow the overall response time of the operational transresistor (the parallel combination of R_F and C_F act like a low-pass filter).

8.4.2.1 Error Analysis of the Shunt Picoammeter

Three major sources of error in the shunt picoammeter circuit are discussed in this subsection. First, errors arise because of calibration errors in R_{SH} , R_1 and R_2 . Second, errors arise because of the EOA's dc bias current, I_B , and offset voltage, V_{OS} . V_{OS} can be nulled to zero at one temperature, but V_{OS} will drift with temperature. I_B cannot be nulled in solid state EOAs, and we must live with it; fortunately, I_B is very small for modern, IC EOAs—about 40 fA in the AD549L. The V_{OS} tempco is $5 \mu\text{V}/^\circ\text{C}$ for the AD549L. The third source of measurement error comes from amplifier noise and from Johnson (thermal) noise in associated resistors. The short circuit input voltage noise, e_{nar} for an AD549L electrometer op-amp is $4 \mu\text{V}$ peak-to-peak in the 0.1–10 Hz bandwidth, and $35 \text{nV}/\sqrt{\text{Hz}}$ above about 1 kHz. The equivalent input current noise is 0.36 fA ppk in the 0.1–10 Hz bandwidth and can be considered to be $0.11 \text{fA}/\sqrt{\text{Hz}}$ white noise in the range 0.1 Hz–100 MHz.

Figure 8.22 illustrates the dc sources which can cause errors in reading the unknown dc current, I_S . Let us assume that $R_N \gg R_{SH}$. The amplifier gain for the non-inverting input is $K_V = (1 + R_F/R_1)$. Now, the amplifier output voltage due to I_S is just:

$$V_{OM} = I_S R_{SH} K_V \quad (8.56)$$

The output voltage due to V_{OS} , I_B and I'_B is:

$$V_{OE} = V_{OS} K_V + I_B R_{SH} K_V + I'_B R_F \quad (8.57)$$

Thus, we can write an expression for the dc output signal to dc output voltage error ratio as:

$$\frac{V_{OM}}{V_{OE}} = \frac{I_S}{V_{OS}/R_{SH} + I_B + I'_B R_F / (R_{SH} K_V)} \quad (8.58)$$

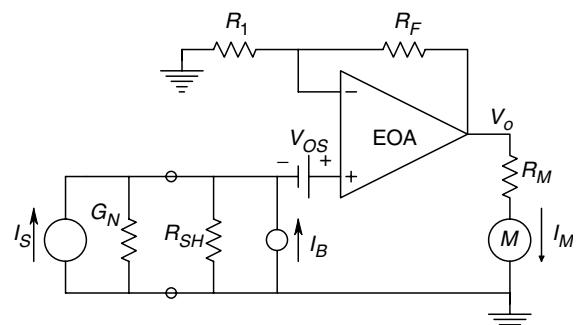


FIGURE 8.22

Circuit showing sources of dc error in the circuit of Figure 8.21A.

Clearly, the meter should be zeroed with input open circuited to null out V_{OE} .

To examine the effect of resistor thermal noise and amplifier noises on the measurement of the dc I_S , we calculate the shunt picoammeter's SNR. This calculation is made more difficult by the fact that the $e_{na}(f)$ root spectrum is in its $1/f$ region. Thus, to evaluate the total RMS output noise, we must assume that all noise sources are independent and uncorrelated, and we must integrate the $1/f$ spectrum over the nominal frequency range used to make dc measurements, 0.1–10 Hz. From the published $e_{na}(f)$ curve on the AD549L data sheet, we can write the short circuit input, noise voltage power density spectrum as:

$$e_{na}^2(f) = \eta + b/f = 1.2250 \times 10^{-15} + (4.0833 \times 10^{-17})/f \text{ MSV/Hz} \quad (8.59)$$

In a JFET input amplifier, $i_{na}(f)$ generally has a flat PDS at low frequencies and can be considered to be white.

Now the total mean squared noise output voltage can be found in the 0.1–10 Hz bandwidth:

$$\begin{aligned} N_O = K_V^2 \int_{0.1}^{10} e_{na}^2(f) df &+ K_V^2 4kT \frac{R_{SH}R_N}{R_{SH} + R_N} B + K_V^2 \frac{(R_{SH}R_N)^2}{(R_{SH} + R_N)^2} i_{na}^2 B \\ &+ 4kT R_1 (R_F/R_1)^2 B + 4kT R_F B \text{ MSV} \end{aligned} \quad (8.60)$$

In equation 8.60, $B = f_H - f_L = 10 - 0.1 = 9.9 \text{ Hz}$, $i_{na} = 0.11 \text{ fA}/\sqrt{\text{Hz}}$ and $K_V = (1 + R_F/R_1)$. Note that we have included Johnson noise from R_1 and R_F . The mean squared output signal is:

$$V_{OS}^2 = I_S^2 \frac{(R_{SH}R_N^2)}{(R_{SH} + R_N)^2} K_V^2 \text{ MSV} \quad (8.61)$$

The MS SNR can be written algebraically as:

$$\text{SNR}_O = \frac{I_S^2/B}{\eta(G_{SH} + G_N)^2 + (b/B) \ln(f_H/f_L)(G_{SH} + G_N)^2 + i_{na}^2 + 4kT(G_{SH} + G_N) + \dots} \quad (8.62)$$

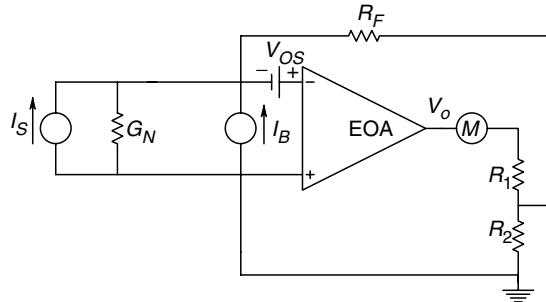
A numerical value for the RMS SNR can be found from the expressions above. For simplicity, assume $R_N \gg R_{SH}$, $R_{SH} = 10^6 \Omega$, $R_F = 100 \text{ K}$, $R_1 = 1 \text{ K}$ and $T = 300 \text{ K}$. The RMS SNR is found to be:

$$\text{SNR}_{RMS} = \frac{I_S}{4.2184 \times 10^{-13}} \text{ A/RMS A} \quad (8.63)$$

Thus, noise limits the resolvable I_S to $3 \times 4.22 \times 10^{-13} \text{ A}$ for the parameters given above, creating random errors.

8.4.2.2 Error Analysis of the Feedback Picoammeter

As in the case of the shunt picoammeter, the electrometer op-amp's offset voltage and bias current cause uncertainty in the measurement of I_S . From the equivalent circuit of

**FIGURE 8.23**

Circuit showing sources of dc error in the feedback picoammeter of Figure 8.21B.

Figure 8.23, we can write an expression for the worst case, dc output signal to dc output error ratio as:

$$\frac{V_{OM}}{V_{OE}} = \frac{I_S}{I_B \pm V_{OS}(G_N + G_F)} \quad (8.64)$$

Here again, the systematic dc error due to V_{OS} and I_B can generally be observed with the input open circuited and compensated for.

The error due to random noise includes the thermal noise from R_N and R_F as well as the amplifier's e_{na} and i_{na} . The output MS SNR can be written as:

$$\text{SNR}_O = \frac{I_S^2 R_F^2}{[i_{na}^2 + 4kT(G_N + G_F)]R_F^2 B + [\eta B + b \ln(f_H/f_L)](R_N + R_F)^2 / R_N^2} \quad (8.65)$$

Equation 8.65 can be reduced to a more interesting form:

$$\text{SNR}_O = \frac{I_S^2}{B\{i_{na}^2 + 4kT(G_N + G_F) + [\eta + (b/B) \ln(f_H/f_L)](G_N + G_F)^2\}} \quad (8.66)$$

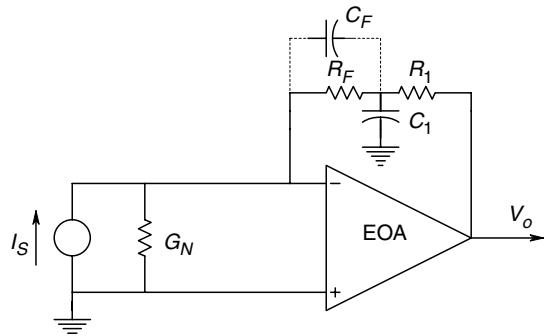
It is instructive to evaluate equation 8.66 above for typical system parameters of $\eta = 1.2250 \times 10^{-15}$ MSV/Hz, $b = 4.0833 \times 10^{-17}$ MSV, $R_F = 10^9 \Omega$, $f_H = 10$ Hz, $f_L = 0.1$ Hz, $i_{na} = 0.11$ fA/ $\sqrt{\text{Hz}}$, $4kT = 1.656 \times 10^{-20}$ and $R_N = 10^6 \Omega$. We find:

$$\text{SNR}_O = \frac{I_S^2}{1.643 \times 10^{-25}} \quad (8.67)$$

So, for reliable low noise measurements, the dc I_S should be over three times the equivalent input RMS noise current, or $I_S > 1.216 \times 10^{-12}$ A. See the problems at the end of this chapter for exercises in calculating noise and dc drift limits to electronic dc picoammeter accuracy.

A problem which often arises in the design of feedback picoammeters is the loss of high frequency response due to the parasitic shunt capacitance which appears in parallel with the very large R_F used. For example, if $R_F = 10^{11} \Omega$, a 1 pF stray capacitance in parallel with R_F will produce a transresistance given by:

$$\frac{V_O}{I_S}(s) = \frac{-R_F}{s R_F C_F + 1} \quad (8.68)$$

**FIGURE 8.24**

A circuit to speed up the response of the feedback picoammeter.

The time constant is $\tau = 0.1$ s in this case. This would give little problem in the measurement of dc currents, but would be unsuitable for the measurement of ac currents. A circuit to speed up the response of the feedback picoammeter is shown in Figure 8.24. It is posed as a problem at the end of the chapter to find the condition on R_1 , C_1 that will cancel the low-pass effect of C_F in parallel with R_F .

8.4.3 Magneto-Optic Current Sensors

The *Faraday magneto-optic effect* has been used in several designs for the measurement of high currents in power distribution systems. Typically, currents can range from tens of amperes to kiloamperes, and the conductors are generally at kilovolts above ground potential, rendering the use of a shunt and millivoltmeter quite impractical. The Faraday magneto-optic effect makes use of the fact that a current carrying conductor is surrounded by a solenoidal magnetic field whose magnitude is proportional to the current. The magnetic field, in turn, induces changes in the optical polarization properties of the glass or fiber optic cable on which it acts. Thus, by measuring the amount of optical rotation induced on linear polarized light emerging from the optical path in the B field, we can indirectly sense the current in the conductor. Details of several Faraday magneto-optic effect current sensors are described in Section 6.7.1 of this text.

8.5 AC Voltage Measurements

Any discussion of ac voltage measurements must necessarily cover a wide range of applications in which the amplitude and frequency of the voltage to be measured are of critical importance. AC voltage measurements can be divided into a number of frequency ranges—audio (including power line frequencies), ranging from around 10 Hz to about 30 kHz; and radio frequencies, including the following bands—low frequency (LF) 30–300 kHz, medium frequency (MF) 300–3000 kHz, high frequency (HF) 3–30 MHz, very high frequency (VHF) 30–300 MHz, ultra high frequency (UHF) 300–3,000 MHz, super high frequency (SHF) 3–30 GHz, extra high frequency (EHF) 30–300 GHz. Resolvable amplitudes range from nanovolts to hundreds of kilovolts. The ac voltages are generally sinusoidal, but situations arise in which one must measure other periodic waveforms including square waves, triangle and sawtooth waves, pulse trains, etc. We also need to measure completely random (noise) voltages with various spectral contents and amplitudes. In the following subsections, we will first describe various non-electronic, electro-mechanical ac voltmeters and then, discuss the designs for electronic ac voltmeters.

8.5.1 Electromechanical AC Voltmeters

Most electromechanical ac voltmeters are designed to work at power line or low audio frequencies, and at voltages ranging from about one volt to tens of kilovolts. As in the case of dc voltmeters, basic electromechanical ac voltmeter design uses a series resistor to limit meter current to give full scale deflection at full scale applied voltage at the meter terminals. Almost all ac voltmeters have their scales calibrated in root mean square volts of a sine wave, regardless of the deflection mechanism of the meter movement. Only true RMS reading voltmeters will remain calibrated when a periodic voltage waveform other than a sine wave is the input to the meter. There are five common types of ac, power line frequency voltmeter movements, some of which we have already described in Section 8.2.1. These include the dynamometer movement, the capacitor (electrostatic) movement, the iron vane movement, the rectifier-D'Arsonval design and the vacuum thermocouple-D'Arsonval design. Of these ac voltmeters, only the thermocouple meter is useful at radio frequencies, up to 50 MHz in some designs.

We defined voltmeter sensitivity for dc voltmeters as the reciprocal of the current required for full scale meter deflection. A similar definition is useful for ac voltmeters, except it is the total meter impedance which is found by multiplying the sensitivity in ohms per volt by the meter's full scale voltage.

8.5.1.1 The Dynamometer AC Voltmeter

This type of ac voltmeter is generally a large, insensitive, bench top instrument used to measure sinusoidal voltages with frequencies from about 20–133 Hz. Accuracy is typically 0.2% of full scale voltage, and meter deflection, assuming a linear spring whose torque, is proportional to the deflection angle, and a mutual inductance which varies linearly with deflection angle, is proportional to the mean squared current in the coils. Typical dynamometer voltmeters can be found with full scale ranges from 30–600 V RMS. About 73.34 mA is required for full scale deflection of the GE type P-3 electrodynamometer voltmeter. Figure 8.8 gives a schematic diagram of a dynamometer voltmeter.

Capacitor voltmeters are also large, insensitive, bench top instruments which can be used to measure sinusoidal voltages with frequencies of 15 Hz–300 kHz. Accuracy is typically 1% of full scale voltage, and meter deflection, assuming linear spring torque with deflection angle and a linear increase of meter capacitance with deflection angle, is proportional to the mean squared voltage across the meter. Typical electrostatic voltmeters can be found with full scale deflections of 300 V RMS–50 kV RMS. It should be noted that the current at a given deflection of an electrostatic meter is given by Ohm's law, $I_M = V_M 2\pi f C_M$, where the meter capacitance, C_M , is itself an increasing function of meter deflection. The innards of an electrostatic voltmeter are shown in Figure 8.7. Note that while the deflection angle of the movement may be proportional to the MS V, the scale is calibrated in RMS V.

8.5.1.2 Iron Vane Voltmeters

These voltmeters are found as large, insensitive, bench top instruments which can be used to measure sinusoidal voltages with frequencies of about 15–133 Hz, with accuracies of about 0.5% of full scale voltage. Iron vane ac panel voltmeters are also found. Iron vane meters can be of the repulsion type, as shown in Figure 8.25, or the attraction type, as shown in Figure 8.26. Typical meter efficiency is about $29 \Omega/V$. This means that the meter draws about 34 mA at full scale deflection. Due to their inefficiency, iron vane voltmeters are generally used to monitor power line voltages, etc. Typical full scale voltages for these meters range 8–750 V RMS, depending on the application. Iron vane

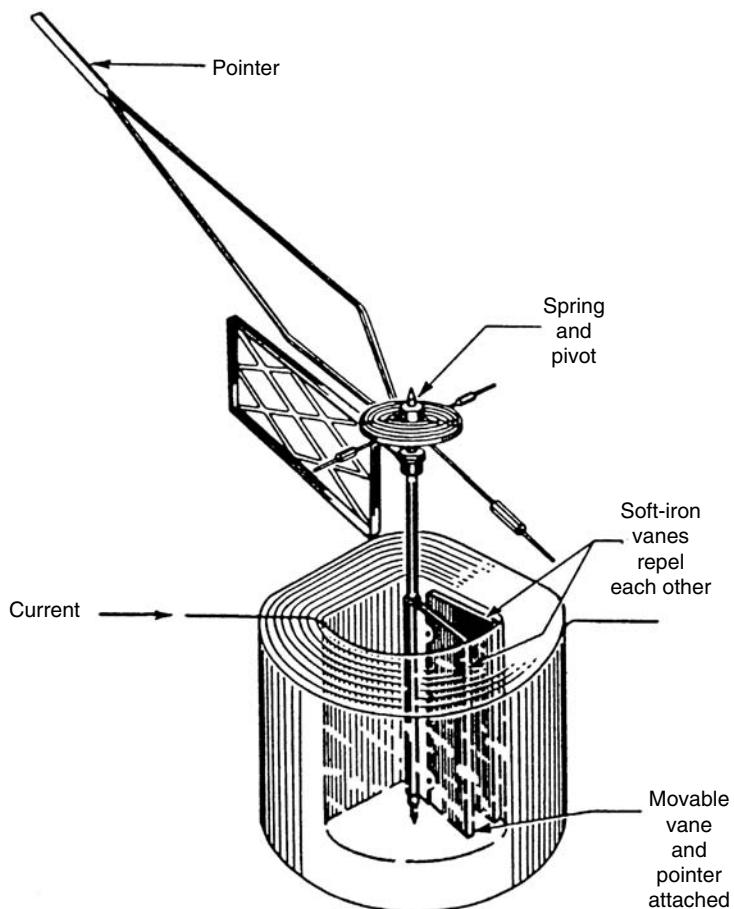
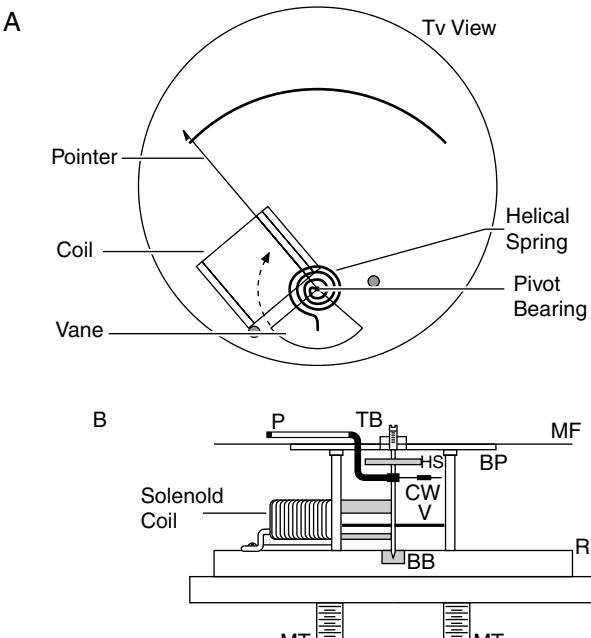
**FIGURE 8.25**

Diagram of an iron vane, ac meter movement of the repulsion type.

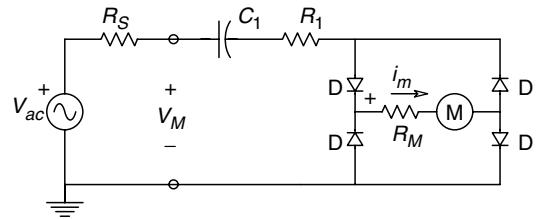
meter scales are generally nonlinear because of the geometry of the vanes and/or coils; the lower quarter of the scale is generally crowded.

8.5.1.3 Rectifier/D'Arsonval AC Voltmeters

These voltmeters generally find application in bench top and pocket volt-ohm-milliammeters (multimeters). Full scale voltages typically range from 3 VRMS to over 1 kVRMS. The circuit of a rectifier/D'Arsonval ac voltmeter is extremely simple, as shown in Figure 8.27. A series resistor, R_1 , limits the meter current for a given full scale voltage. A series capacitor, C_1 , blocks dc from the meter circuit. A bridge rectifier converts ac current in R_1 to a full wave rectified current with a dc average value. The D'Arsonval microammeter movement deflects proportional to the average current through its coil, hence deflection of the rectifier/D'Arsonval meter is proportional to the full wave rectified, average value of the input voltage. The ac voltmeter calibration is, of course, in terms of the RMS of a sine wave input voltage. The D'Arsonval meter movement acts as a mechanical low-pass filter (*cf.* equation 8.12), its deflection being proportional to the average current through its coil. The average value of a full wave rectified sine

**FIGURE 8.26**

A. Face view of an iron vane, ac meter of the attraction type. B. Side view of the attraction type, ac, iron vane meter. Magnetic torque pulls the vane into the current carrying coil.

**FIGURE 8.27**

Schematic of a bridge rectifier, ac voltmeter. A dc, D'Arsonval microammeter movement is used.

wave is easily shown to be:

$$\overline{i_M} = \overline{I_{M(PK)} \sin(\omega t)} = \frac{2}{\pi} I_{M(PK)} = \frac{2\sqrt{2}}{\pi} I_{M(RMS)} \quad (8.69)$$

If we assume at high input voltages that the rectifier diodes are ideal (i.e., they have zero forward resistance and voltage drop, and infinite reverse resistance), then we may write for full scale meter deflection:

$$I_{M(FS)} = \frac{2\sqrt{2} V_{S(RMS, FS)}}{\pi(R_1 + R_M)} \quad (8.70)$$

For a D'Arsonval microammeter with resistance R_M and full scale current, $I_{M(FS)}$, we may solve for the R_1 required to make an ac voltmeter with a full scale voltage reading of $V_{S(RMS, FS)}$, assuming ideal diodes:

$$R_1 = \frac{2\sqrt{2} V_{S(RMS, FS)}}{\pi I_{M(FS)}} - R_M \quad (8.71)$$

The ac, full wave rectifier voltmeter's efficiency can be shown to be:

$$\eta_{ac} = \frac{1}{I_{M(RMS)}} = \frac{R_1 + R_M}{V_{S(RMS, FS)}} = \frac{(2/\sqrt{2\pi})(V_{S(RMS, FS)}/I_{M(FS)})}{V_{S(RMS, FS)}} = 0.9 \eta_{dc} \quad (8.72)$$

The above relations are generally valid for ac input voltages above 10 V RMS. The scales of meters reading below 10 V RMS full scale must generally be corrected for the finite forward and reverse resistances of the real diodes used in the meter rectifier. Also, practical diodes exhibit a voltage dependent, depletion capacitance which is effectively in parallel with their reverse resistance. Thus, at high audio frequencies, considerable reverse diode current can flow, decreasing the average current through the meter movement and causing frequency dependent low meter readings. At power line frequencies and low audio frequencies, diode capacitance is not a problem.

A problem commonly encountered in using ac voltmeters of this sort is their response to non-sinusoidal voltages, such as square waves, triangle waves, etc. To analyze this problem, we note first that any average or dc component in V_S is removed from the voltmeter input current, I_S , by the series capacitor, C_1 . The average (dc) current through the D'Arsonval movement can be written as:

$$\overline{i_M(t)} = \frac{\overline{|v_S(t)|}}{R_1 + R_M} = \frac{\overline{|v_S(t)|}}{(2\sqrt{2/\pi}(V_{S(RMS, FS)}/I_{M(FS)})} \quad (8.73)$$

The actual meter reading can be expressed as a fraction of the full scale voltage:

$$V_{READ} = \frac{\overline{i_M(t)}}{I_{M(FS)}} V_{S(RMS, FS)} = \frac{\overline{|v_S(t)|}}{2\sqrt{2/\pi}} = 1.111 \overline{|v_S(t)|} \quad (8.74)$$

8.5.1.4 Vacuum Thermocouple/D'Arsonval AC Voltmeters

These voltmeters are the only type of electromechanical ac voltmeter which will remain calibrated for input voltage frequencies of 15–20 Hz to about 50 MHz. The reading of a vacuum thermocouple ac voltmeter is proportional to the true root mean square of the voltage being measured.

A schematic of a vacuum thermocouple ac voltmeter is shown in Figure 8.28. The input circuit consists of a series resistor, R_1 , and a heater filament with resistance, R_H .

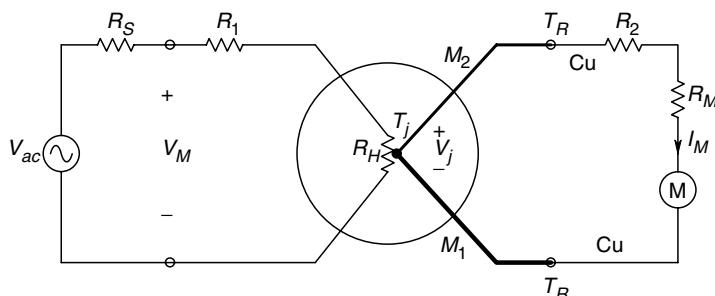


FIGURE 8.28

Schematic of a vacuum thermocouple ac voltmeter. A dc, D'Arsonval microammeter movement is used. Meter deflection is proportional to the mean squared heater current.

For example, R_H for a Western Electric type 20K vacuum thermocouple is given as $750\ \Omega$. The maximum safe heater current is given as 5 mA RMS. A Western Electric type 20D vacuum thermocouple has a heater resistance of $35\ \Omega$, and a safe current of 16 mA RMS. An over-current of as little as 50% is likely to burn out the heater wire, so extreme care must be taken to prevent over-range inputs when using vacuum thermocouple/D'Arsonval ac voltmeters. The maximum rated EMF for the Western Electric vacuum thermocouples mentioned above is 15 mV (dc), and the thermocouple resistance is $12\ \Omega$. If the thermocouple is connected to a 0–20 μA D'Arsonval meter, and the total series resistance in the circuit, $(R_{TC} + R_M + R_2)$, equals $750\ \Omega$, then the meter will read full scale when the heater current is at its rated maximum (3.6 mA RMS in the 20 K vacuum thermocouple's heater).

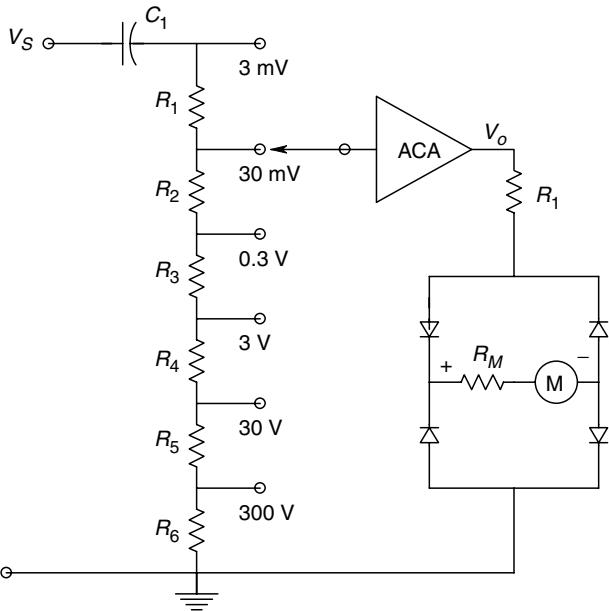
The dc current from the thermocouple, hence meter deflection, is proportional to the mean squared heater current (or heater voltage). This is easy to see from Joule's law—the thermocouple EMF is proportional to heater temperature. Heater temperature is proportional to the electrical power dissipated in the heater, and the power, in turn, is proportional to the mean squared heater current times the heater resistance. In order that the meter be calibrated in RMS volts, a square root scale is painted on the meter face, rather than the linear one commonly found on D'Arsonval meters used in dc applications. Thus, the lower end of the scale is crowded, and calibrations are seldom given below 20 V on a 100 VFS meter. Laboratory quality thermocouple meters commonly have accuracies of 0.5% of full scale voltage.

8.5.2 Analog Electronic AC Voltmeters

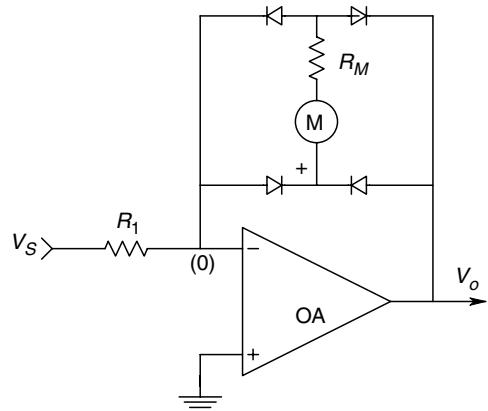
In this section, we will review the more common types of analog ac voltmeters (i.e., ac voltmeters which make use of analog electronic signal conditioning, and which use analog output meters). The ac voltages can be measured effectively from the microvolt range to tens of kV. Frequencies can range from power line to tens of gigahertz. No one type of ac voltmeter is effective over the ranges of frequency and voltage cited. The fundamental limit to the resolution of small ac voltages is noise—thermal noise from resistors, noise from amplifiers, and coherent noise picked up from the environment in spite of our best efforts to shield and guard the circuits being used.

8.5.2.1 AC Amplifier-Rectifier AC Voltmeters

In the first class of ac electronic voltmeter, the ac voltage under measurement is conditioned by a flat bandpass, R-C amplifier of accurately known gain. The amplified ac voltage is then converted to dc, which causes meter indication of the ac voltage. Most ac electronic analog voltmeters make use of the common D'Arsonval movement as the output indicator, either in conjunction with some type of electronic rectifier to convert amplified ac voltage to dc current, or with a vacuum thermocouple to convert the mean squared value of the amplified ac voltage to a dc current to drive the meter. A block diagram of the ac amplifier-rectifier type of ac voltmeter is shown in Figure 8.29. The amplifier's gain is generally fixed and large (e.g., 3.33×10^3), and a precision, frequency compensated attenuator is used to reduce input signals which are above the minimum full scale voltage (e.g., 3 mV) to a level that will not saturate the amplifier. Input resistance is thus set by the attenuator, and is typically 10 or $11\ M\Omega$. Since the amplifier's ac output is rectified to drive the D'Arsonval meter, the rectifier's frequency response generally sets the upper useful frequency of this type of meter. Such meters are typically used at audio frequencies and some versions are useful up to 10 MHz. To overcome the nonlinearity of

**FIGURE 8.29**

Schematic of an ac amplifier-rectifier type ac voltmeter.

**FIGURE 8.30**

A simple operational rectifier circuit that provides linear rectification of the audio frequency ac input current even at low (millivolt) values of V_o .

the rectifier diodes at low output voltage levels, it is possible to use an operational rectifier circuit, shown in Figure 8.30, instead of a passive diode bridge. Such a circuit will improve meter linearity at the low end of the meter scale. However, the reverse capacitance of the diodes, and the finite slew rate and gain bandwidth product of the op-amp provide fundamental limitations to the operating frequency of the operational rectifier. Graeme (1974) gives several circuit modifications for operational rectifiers which improve their performance.

8.5.2.2 Peak Reading Electronic AC Voltmeters

The second type of ac electronic voltmeter (peak reading) uses a low capacitance diode to half-wave rectify the ac voltage under measurement. The rectified ac is smoothed by a capacitor low-pass filter and then, is amplified by a drift free, dc amplifier with known gain, such as a chopper stabilized or commutating auto zero circuit. The dc output of the amplifier causes the deflection of the voltmeter's D'Arsonval meter. The rectifier used

in the peak reading ac voltmeter can be a vacuum, thermionic diode, although low capacitance semiconductor diodes can also be used. The advantage of this design lies in the measurement of RF voltages. Since the ac voltage under measurement is converted to dc at the front end, there is no need for an expensive, high gain, high bandwidth linear amplifier to condition the ac voltage. There is need for a low gain, drift free, dc amplifier, however. Due to the inherent nonlinearity of semiconductor and thermionic diodes at low voltages, peak reading ac voltmeters are generally used for ac input voltages above 1 V_{pk}. Below 1 V_{pk}, special scale calibration must be used to compensate for diode nonlinearity, which can be shown to be square law for input voltages below 50 mV [Oliver and Cage, 1971]. Most peak reading ac voltmeters have input ranges of 1–300 V_{pk} full scale. Their frequency response ranges from power line to over 300 MHz. Indeed, some special microwave diodes have been used to measure peak voltages in waveguides at frequencies up to 40 GHz.

There are two common forms of peak detecting ac voltmeter illustrated in Figure 8.31. In Figure 8.31A, we see a dc peak detector in which the capacitor charges through the diode to a dc voltage equal to the positive peak value of $v_s(t)$. Thus, the output voltage across the diode is $v_1(t) = [v_s(t) - V_{PK+}]$. In the case of a sine wave input with zero mean, the average voltage across the diode is $v_1 = V_{S(PK)}$. This dc level is recovered by low-pass filtering, amplified and used to drive the D'Arsonval indicating meter. Thus, meter deflection is proportional to the peak value of the zero mean sine wave input. In more general terms, for an arbitrary input $v_s(t)$, the mean value of $v_1(t)$ can be shown to be $\bar{v}_1(t) = -(+ \text{peak } v_s - \bar{v}_s)$.

To demonstrate this property, consider the pulse train waveform in Figure 8.32C. Its peak value is V_p and its average value is:

$$\bar{V}_s = \frac{V_p T_1 - V_2 T_2}{T_1 + T_2} \quad (8.75)$$

So, the average value of $v_1(t)$ is the ($-$ peak-above-average) value of $v_s(t)$:

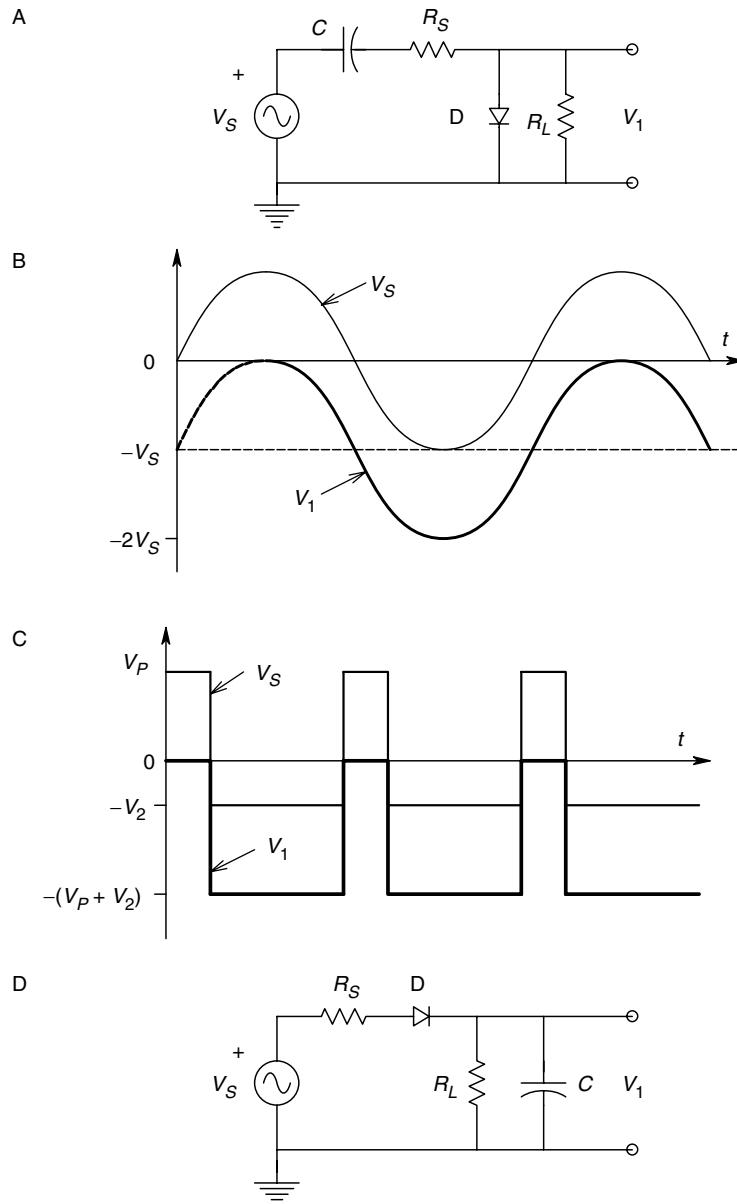
$$\bar{V}_1 = -\left[V_p - \frac{V_p T_1 - V_2 T_2}{T_1 - T_2} \right] = -\frac{T_2(V_p + V_2)}{T_1 + T_2} \quad (8.76)$$

Peak reading voltmeters are generally calibrated in terms of the RMS of a sinusoidal input. The input impedance of the diode probe of a peak reading voltmeter can be shown to be $Z_{IN} = R_L/3$, where in general, $R_S \ll R_L$ [Angelo, 1969].

A second type of peak reading electronic ac voltmeter is shown in Figure 8.31D. In this simple circuit, the capacitor acts as the low-pass filter. The capacitor charges up to the positive peak value of $v_s(t)$. Hence, the average output of the detector is $\bar{v}_1(t) = V_{S(PK)}$. Charge leaks off C through R_L , with a time constant adjusted to be about 0.5 s. The input impedance of the peak rectifier can be shown to be $Z_{IN} = R_L/2$ at frequencies where $1/\omega C \ll R_L$ [Angelo, 1969]. A peak-to-peak reading ac voltmeter can be made from a positive and a negative peak rectifier connected to a dc differential amplifier, as shown in Figure 8.32.

8.5.2.3 True RMS AC Voltmeter of the Feedback Type

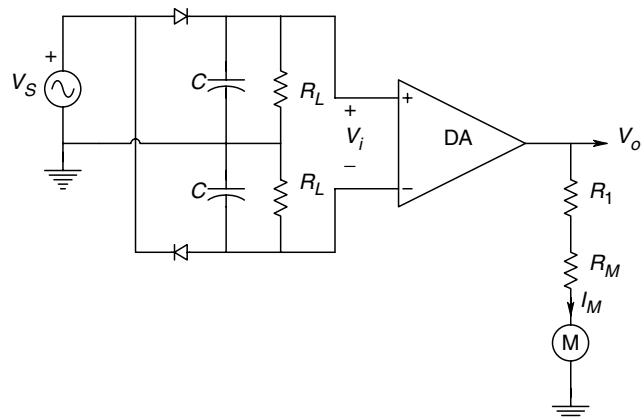
The circuit for this ac voltmeter is shown in Figure 8.33. In the heart of the circuit are two, matched vacuum thermocouples. The heater current of the input thermocouple is derived by ac amplification (or attenuation) of the ac input signal, V_s . The heater current

**FIGURE 8.31**

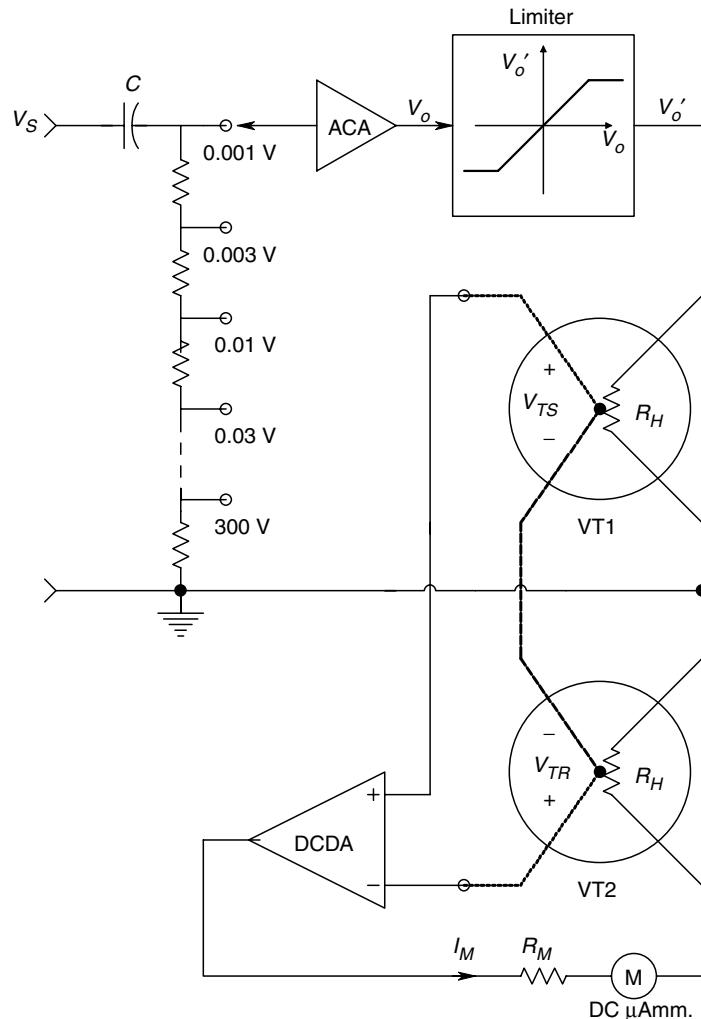
A. Schematic of a diode, peak detecting ac voltmeter. B. Steady state waveforms in the peak detector given a sinewave input. C. Steady state waveforms in the peak detector given a pulse train input. D. A second type of peak reading ac voltmeter.

of the feedback thermocouple comes from the dc current through the D'Arsonval meter movement. We will show below that the dc meter current is proportional to the true RMS value of $v_S(t)$ when certain conditions of the circuit parameters are met.

After attenuation and/or amplification by the linear ac amplifier, an ac signal ($v_S(t)$), proportional to v_{HS} , is applied to the input thermocouple heater, causing a rise in its temperature and a resultant dc output voltage, V_{TS} , from the thermocouple junction attached to the input heater. The input to the high gain dc differential amplifier is the difference between V_{TS} and V_{TR} , V_{TD} . V_{TD} is amplified and produces a dc output

**FIGURE 8.32**

Schematic of a peak-to-peak reading ac voltmeter using a DA.

**FIGURE 8.33**

Simplified schematic of a feedback type, true RMS AC voltmeter.

voltage, V_{OD} . The dc meter current is thus $I_M = V_{OD}/(R_o + R_M + R_H)$. From inspection of the circuit, we can write the following equations:

$$R = R_o + R_M + R_H \quad (8.77)$$

$$V_{TS} = K_T \overline{v_{HS}^2} \quad (8.78)$$

$$V_{TR} = K_T V_{HR}^2 = K_T (R_H I_M)^2 \quad (8.79)$$

In the equations above, R_o is the output resistance of the dc differential amplifier, K_D is its gain, R_H is the thermocouple heater resistance, R_M is the dc microammeter's resistance and K_T is the thermocouple's transfer gain in volts/mean squared volt. The dc meter current can be written from Ohm's law:

$$I_M = \frac{K_D(V_{TS} - V_{TR})}{R} = \frac{K_D K_T (\overline{v_{HS}^2} - V_{HR}^2)}{R} \quad (8.80)$$

If we substitute for V_{HR} , we obtain a quadratic equation in I_M :

$$IM^2 + I^M \frac{R}{K_D K_T R_H^2} - \frac{\overline{v_{HS}^2}}{R_H^2} = 0 \quad (8.81)$$

The solution to this equation is of the form:

$$I_M = -\frac{R}{2K_D K_T R_H^2} \pm \frac{1}{2} \sqrt{\frac{R^2}{(K_D K_T R_H^2)^2} + \frac{4\overline{v_{HS}^2}}{R_H^2}} \quad (8.82)$$

We extract the $4/R_H^2$ factor from equation 8.82 above and write:

$$I_M = -\frac{R}{2K_D K_T R_H^2} \pm \frac{1}{R_H} \sqrt{\frac{R_2}{(2K_D K_T R_H)^2} + \frac{\overline{v_{HS}^2}}{R_H^2}} \quad (8.83)$$

Now, if we make $V_{HS(RMS)} \gg \frac{R}{2K_D K_T R_H}$, I_M is approximately given by:

$$I_M \cong V_{HS(RMS)}/R_H \quad (8.84)$$

V_{OD} is given by:

$$V_{OD} = (V_{HS(RMS)})/R_H)(R_M + R_H) \text{ V dc} \quad (8.85)$$

In a typical vacuum thermocouple, $I_{M(MAX)}$ through the heater is 16 mA, and the heater resistance is 35 Ω. Thus, $V_{SH(RMS, MAX)}$ is 0.56 V. Now if we let $K_D=10^6$, $K_T=3 \times 10^{-2}$, $R_M=250 \Omega$ and $R_o=50 \Omega$, then $V_{HS(RMS, MIN)}$ would be around 1 mV. This design gives the meter around a 560:1 dynamic range over which meter deflection is proportional to the true RMS value of $v_S(t)$.

The Hewlett-Packard model 3400A true RMS voltmeter used a design similar to the one analyzed above.

8.5.2.4 True RMS AC Voltmeters Using the Direct Conversion Approach

In Section 2.6.4, we saw that there are several dedicated ICs available that will perform true RMS conversion on ac signals. The dc output of these ICs can be conditioned and used to drive a D'Arsonval dc microammeter. The maximum frequency response for a specified % error in conversion is generally obtained for maximum ac input. For example, The AD536 RMS converter has less than 1% error at frequencies up to about 140 kHz, when its input is a 7 V RMS sinewave, and less than 1% error at frequencies up to 6 kHz when the input is a 10 mV RMS sinewave. Thus, to make a true RMS voltmeter with ICs of this type, we need to adjust the input gain or attenuation such that the input to the IC is always as large as possible, without saturating the circuit. One way of insuring this accuracy and dynamic range is to always make measurements on the top one-third of the meter scale.

For some interesting TRMS voltmeter designs and applications of TRMS converter ICs, the reader should consult Analog Devices' *RMS to dc Conversion Application Guide* (1983).

Note that it is possible to build a TRMS analog voltmeter from a wide bandwidth analog multiplier, a low-pass filter, and a square root circuit. Such a circuit is capable of a wider bandwidth than is obtainable with a TRMS converter IC (e.g., if a four-quadrant analog multiplier such as an AD834 is used, an input signal bandwidth of 5 Hz to over 20 MHz for peak input voltages up to 10 V). The schematic of a squarer-averager-rooter is shown in Figure 8.34. Again, maximum performance in terms of low error and maximum bandwidth is obtained with the use of the full dynamic range of the analog multiplier and square rooter. The analog multiplier used in the square-rooter does not need outstanding bandwidth since it is operating on dc, so a high accuracy analog multiplier such as the AD534 can be used here.

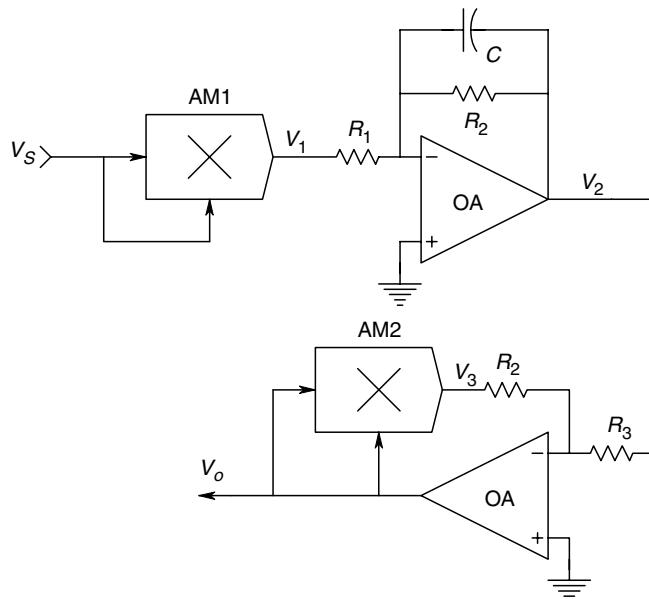


FIGURE 8.34

Schematic of a true RMS converter using an analog multiplier as a squarer, an op-amp low-pass filter averager and a square root circuit. A dc output is obtained.

8.5.3 Measurements of Amplifier Noise Voltages, Noise Factor and Figure

In making noise voltage measurements, one must be certain that the noise being measured is indeed uncontaminated by coherent interference or hum from sources external to the system under measurement. The best way to gain this assurance is to visualize the noise spectrum in question on a modern, sampling, FFT spectrum analyzer. Such instruments generally display a one-sided, root power spectrum whose units are RMS V/ $\sqrt{\text{Hz}}$. The presence of any spike-like peaks generally indicates the presence of contamination of the desired spectrum by coherent interference. Obviously, steps must be taken to remove extraneous coherent voltages from the desired spectrum in order to obtain valid noise measurements.

The quantitative measurement of noise and noise figure requires the use of true RMS ac voltmeters, such as we have described above. Noise measurements also require knowledge of amplifier gain and the equivalent Hertz noise bandwidth of bandpass filters that are conditioning the noise from the noise source(s). Noise measurements can be subdivided into two categories—the use of narrow bandpass filters to characterize the power density spectrums of noise sources, and broadband noise measurements over the entire bandwidth of a signal conditioning system. We are often interested in measuring the equivalent input noise sources of an amplifier, $e_{na}(f)$ and $i_{na}(f)$. $e_{na}(f)$ is the equivalent, short circuited input noise of the amplifier. It is a root power spectrum and its units are RMS V/ $\sqrt{\text{Hz}}$. $e_{na}(f)$ generally increases at low frequencies, is flat in a mid-range and rises again at high frequencies. Ideally, in order to obtain a maximum SNR at the amplifier's output, we would like to condition the input signal over a range of frequencies where $e_{na}(f)$ is a minimum.

Figure 8.35 illustrates a system which can be used to measure the flat, white noise portion of the $e_{na}(f)$ root spectrum. It is often necessary to follow amplifier A1 with a second, high gain, low noise amplifier, A2, which has a broader bandwidth than A1. That is, the bandpass characteristics of A1 dominate the system and determine B , the equivalent system Hz noise bandwidth. It can be shown [Northrop, 1990] that as long as the mid-band gain of A1 is greater than 5, any noise from A2 is negligible in the measurement process. If we assume $e_{na}(f)$ is flat over B , then we can write an expression

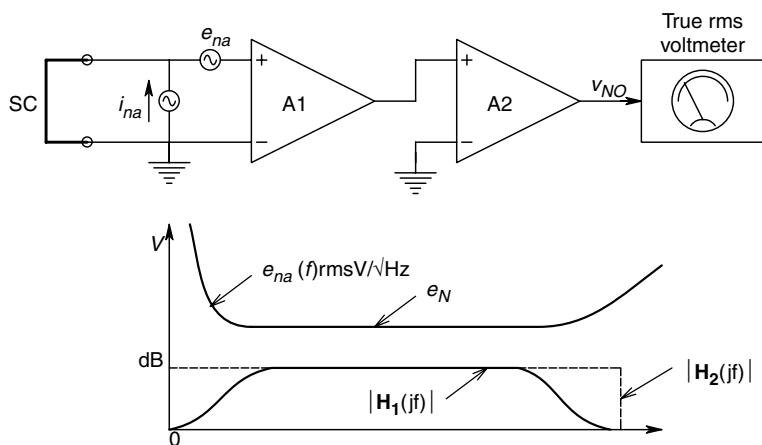


FIGURE 8.35

Circuit used to measure the white noise portion of e_{na} from A1. The noise passband excludes the $1/f$ region of the e_{na}^2 power density spectrum.

for the noise power density spectrum at the system output, $S_{NO}(f)$:

$$S_{NO}(f) = e_N^2 |H_{12}(j2\pi f)|^2 \quad (8.86)$$

The RMS meter must have a flat frequency response which is wider than that of A1. The RMS noise read by the meter is given by:

$$v_{NO(RMS)} = \sqrt{\int_0^\infty e_N^2 |H_{12}(j2\pi f)|^2 df} = e_N K_{V12} \sqrt{B} \text{ RMS V} \quad (8.87)$$

where

$$B = \int_0^\infty |H_{12}(j2\pi f)|^2 df / K_{V12}^2 \text{ Hz} \quad (8.88)$$

and K_{V12} is the mid-band gain of the cascaded amplifiers A1 and A2, and $H_{12}(j2\pi f)$ is the frequency response of the cascaded amplifiers. If $e_{na}(f)$ is not flat over B , then the RMS output noise voltage is given by:

$$v_{NO(RMS)} = \int_0^\infty e_{na}^2(f) |H_{12}(j2\pi f)|^2 df \quad (8.89)$$

To measure $e_{na}(f)$ in its low frequency, $1/f$ region, we note that the e_{na} power density spectrum can be described by:

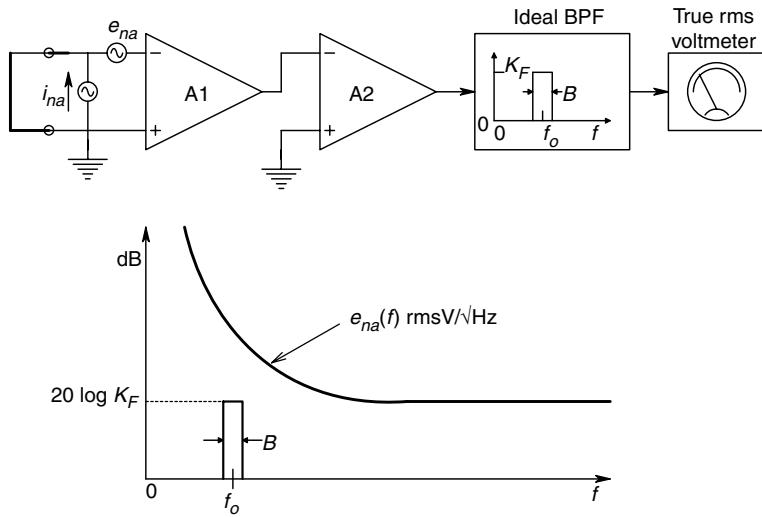
$$e_{na}^2(f) = b/f + e_N^2 \text{ MS V/Hz} \quad (8.90)$$

Now, we use a narrow bandpass filter to examine the spectral power of $e_{na}^2(f)$. We assume that the filter has a rectangular bandpass characteristic with peak gain K_F and noise bandwidth B Hz. Referring to Figure 8.36, we see that the mean squared output noise is given by:

$$\begin{aligned} v_{NO}^2 &= \int_{f_0 - B/2}^{f_0 + B/2} \left(\frac{b}{f}\right) K_F^2 K_{V12}^2 df + e_N^2 K_F^2 K_{V12}^2 B \\ v_{NO}^2 &= K_F^2 K_{V12}^2 \left[b \ln\left(\frac{2Q+1}{2Q-1}\right) + e_N^2 B \right] \end{aligned} \quad (8.91)$$

where $Q = f_0/B$. From equation 8.91 above, we can solve for the b coefficient:

$$b = \frac{[v_{ON}^2/(K_F^2 K_{V12}^2) - e_N^2 B]}{\ln\left(\frac{2Q+1}{2Q-1}\right)} \text{ Hz} \quad (8.92)$$

**FIGURE 8.36**

System used to measure the $1/f$ portion of the e_{na}^2 power density spectrum.

In certain modern, low noise op-amps, e_N is about $3 \text{ nV}/\sqrt{\text{Hz}}$ and b is about $27 \times 10^{-18} \text{ MSV}$.

Although in many applications, the amplifier's equivalent input noise current, $i_{na}(f)$, has negligible effect on the output SNR, there are instances where its value does have an effect. To measure the effect of $i_{na}(f)$, we replace the input short circuit with a resistor, R , at temperature T K. Now, there are three sources of noise voltage at A1's input: e_{nar} , the Johnson noise from R (assumed white), and the noise $i_{na}R$. Using the procedures from Chapter 3, we can write an expression for the mean squared output noise:

$$v_{NO}^2 = \int_0^\infty [e_{na}^2(f) + 4kTR + R^2i_{na}^2(f)] |H_{12}(j2\pi f)|^2 df \text{ MSV} \quad (8.93)$$

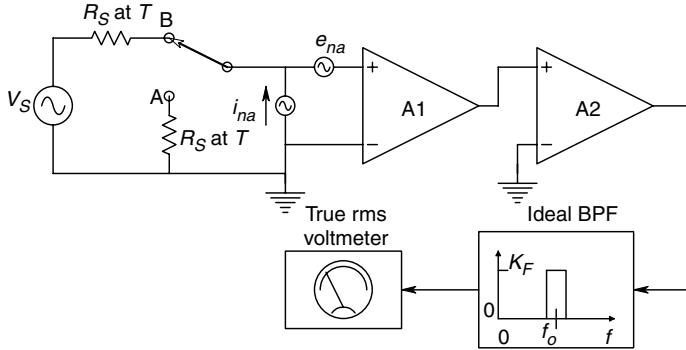
Now, if we assume both e_{na} and i_{na} are white over B , equation 8.93 can be simplified to:

$$\overline{v_{NO}^2} = [e_{na}^2 + 4kTR + R^2i_{na}^2] K_{V12}^2 B \text{ MSV} \quad (8.94)$$

Since we know that K_{V12} , B , e_{na} , R , T and B , it is easy to calculate i_{na} :

$$i_{na}^2 = \left(\frac{v_{NO}^2}{K_{V12}^2 B} - e_{na}^2 - 4kTR \right) / R^2 \text{ MS A/Hz} \quad (8.95)$$

Our final topic in this section will be to describe how we can measure an amplifier's *spot noise figure*, NF . Spot noise factor, F , is defined as the ratio of the mean squared signal to noise ratio at the amplifier's input to the mean squared signal-to-noise ratio at the amplifier's output. An ideal, noiseless amplifier would have an $F = 1$. An amplifier's noise figure is defined as $10 \log_{10}$ (noise factor). An amplifier's spot noise factor is a useful measure of its noise performance under various conditions of source resistance and frequency.

**FIGURE 8.37**

Circuit used to measure the spot noise figure of amplifier A1 over a wide range of frequencies. Amplifier A2 is assumed noiseless.

A circuit which can be used to measure spot noise figure of amplifier A1 is shown in Figure 8.37. A sinusoidal voltage source is used at the input to enable the measurement of F_{SPOT} . A narrow bandpass filter (NBPF) with peak gain, K_F , and equivalent Hz noise bandwidth, B , is used to select the range of noise frequencies for the measurement. A broadband, true RMS responding meter is used to read the total RMS voltage at the amplifier output. With the switch in position A, the meter reading squared can be written as:

$$\begin{aligned} v_{NOA}^2 &= \int_0^\infty K_{V12}^2 |H_F(j2\pi f)|^2 [e_{na}^2(f) + i_{na}^2 R_S^2 + 4kTR_S] df \\ v_{NOA}^2 &= \int_0^\infty K_{V12}^2 |H_F(j2\pi f)|^2 e_{na}^2(f) df + K_{V12}^2 K_F^2 B(i_{na}^2 R_S^2 + 4kTR_S) \text{ MS V} \end{aligned} \quad (8.96)$$

Note that the bandwidth of the meter and the amplifier are assumed to be greater than and overlap that of the NBPF, so that the filter's frequency response curve is dominant. Now, the switch is set to position B and the sinusoidal source is adjusted to the center frequency of the NBPF. Its amplitude is adjusted until:

$$\begin{aligned} v_{NOB}^2 &= 2v_{NOA}^2 = K_F^2 K_{V12}^2 (V_S^2 / 2) + \int_0^\infty K_{V12}^2 |H_F(j2\pi f)|^2 e_{na}^2(f) df \\ &\quad + K_{V12}^2 K_F^2 B(i_{na}^2 R_S^2 + 4kTR_S) \text{ MS V} \end{aligned} \quad (8.97)$$

Now, the MS SNR in bandwidth B at the amplifier's input is

$$\text{SNR}_{IN} = \frac{\overline{v_S^2}}{4kTR_S B} \quad (8.98)$$

The SNR_O is:

$$\text{SNR}_O = \overline{v_S^2} / \left[\int_0^\infty K_{V12}^2 |H_F(j2\pi f)|^2 e_{na}^2(f) df + K_{V12}^2 K_F^2 B(i_{na}^2 R_S^2 + 4kTR_S) \right] \quad (8.99)$$

Now, the spot noise factor,

$$F_{SPOT} \equiv \frac{SNR_{IN}}{SNR_O} \quad (8.100)$$

can be written, after some algebra as:

$$F_{SPOT} = \frac{\int_0^{\infty} |H_F(j2\pi f)|^2 e_{na}^2(f) df / K_F^2 + B(i_{na}^2 R_S^2 + 4kTR_S)}{4kTR_S B} \quad (8.101)$$

If we combine equation 8.96 and equation 8.97, it is easy to show that the numerator of equation 8.101 is equal to $V_s^2/2$, which is the mean squared sinusoidal input signal of frequency, f_o , required to make $v_{NOB}^2 = 2v_{NOA}^2$. Thus, we can finally write the expression for the spot noise factor as:

$$F_{SPOT} = \frac{V_s^2/2}{4kTR_S B} \quad (8.102)$$

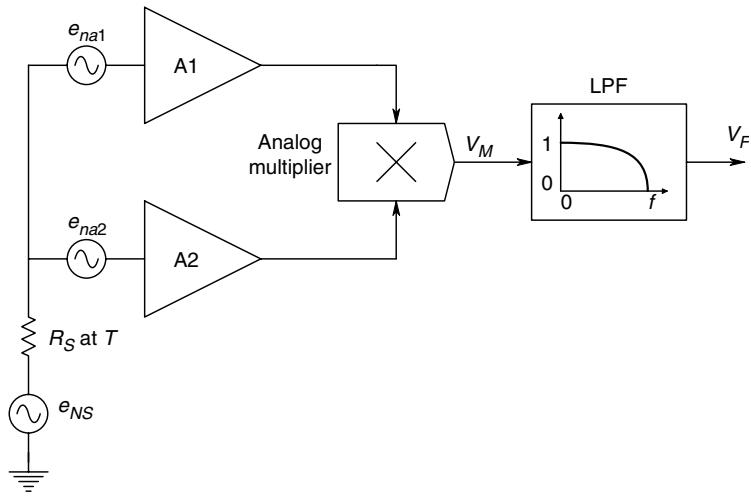
Note that V_s is the peak value of the sinusoid of frequency, f_o , required to make the true RMS meter reading increase by $\sqrt{2}$ when the switch is in the B position, B is the NBPF's equivalent Hz noise bandwidth and T is the temperature in K of R_S . F_{SPOT} is often presented as a plot of $10 \log_{10}(F_{SPOT})$ vs R_S and f_o values by manufacturers of low noise amplifiers and transistors.

An alternate means of measuring F_{SPOT} makes use of the same circuit as in Figure 8.37, except that the sinusoidal source is replaced with a calibrated, broadband, white noise source with a flat spectrum in the $1/f$ region of e_{na} . The same procedure is followed. First we measure v_{NOA}^2 and then adjust the noise source level to e_{NS}^2 MS V/Hz so that $v_{NOB}^2 = 2v_{NOA}^2$. It is left as an exercise for the reader to show that for the noise source method,

$$F_{SPOT} = \frac{e_{NS}^2}{4kTR_S} \quad (8.103)$$

Note that the center frequency of the NBPF and its noise bandwidth do not appear in this simple expression, but they should be specified in presenting the calculated values for F_{SPOT} . In general, a different value of e_{NS} will be found for each point in (f_o, R_S) space.

Often, we wish to measure the RMS voltage of a noise source in bandwidth B when the noise is of the same order of magnitude or smaller than the noise arising in our signal conditioning amplifiers. Figure 8.38 illustrates a means of measuring such low level noise voltages. The noise voltage under measurement, e_{NS} , appears in series with a Thevenin source resistance, R_S . This source is connected to two identical amplifier channels with gain K_V and noise bandwidth B Hz. Both amplifiers have short circuit input equivalent voltage noise sources which are statistically equal, have zero means, but otherwise, are independent and uncorrelated. The outputs of the two channels are multiplied together using a precision, four-quadrant, broadband, analog multiplier. The output of the multiplier is then passed through a low-pass filter to average it. For simplicity, let us assume that the noises have finite bandwidths less than B . Thus, the multiplier output

**FIGURE 8.38**

A correlation detector used to measure the mean squared voltage of the noise source, $e_{ns}(t)$ and of the Johnson noise from R_S over the bandwidth B . $e_{na(1)}$ and $e_{na(2)}$ have no effect on V_F .

can be written as:

$$\begin{aligned} v_M(t) &= (0.1) \{K_V[v_{NS}(t) + e_{na1}(t)]\} \{K_V[v_{NS}(t) + e_{na2}(t)]\} \\ &= (0.1)K_V^2[v_{NS}^2(t) + v_{NS}(t)e_{na2}(t) + e_{na1}(t)v_{NS}(t) + e_{na1}(t)e_{na2}(t)] \end{aligned} \quad (8.104)$$

Low-pass filtering $v_M(t)$ effectively computes $\overline{v_M}$. Since the three voltages have zero means and are statistically independent and uncorrelated with one another, the average of each cross-term product is the product of the averages, which by definition, are zero. Thus, the output of the filter is a dc voltage, V_F given by:

$$V_F = (0.1)K_V^2 \overline{v_{NS}^2} \quad (8.105)$$

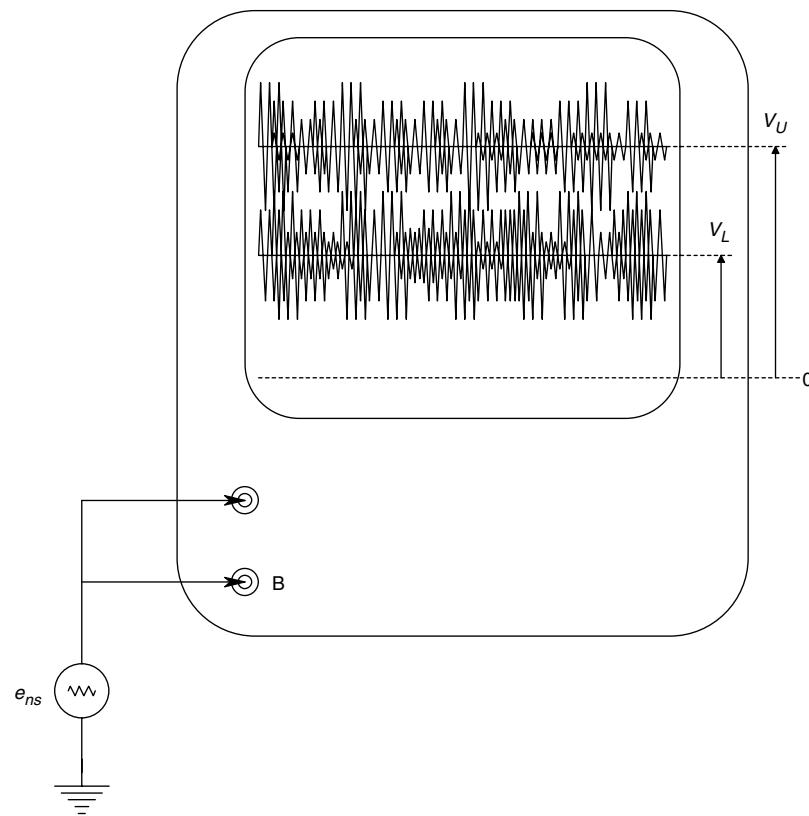
V_F is thus proportional to the mean square of the noise voltage under measurement, taken over bandwidth B . The effects of the amplifier voltage noise drops out; we have assumed the current noise to be negligible (i.e., $i_{na} R_S \ll e_{na}$). Also, because of R_S :

$$\overline{v_{NS}^2} = (4kTR_S + e_{NS}^2)B \quad (8.106)$$

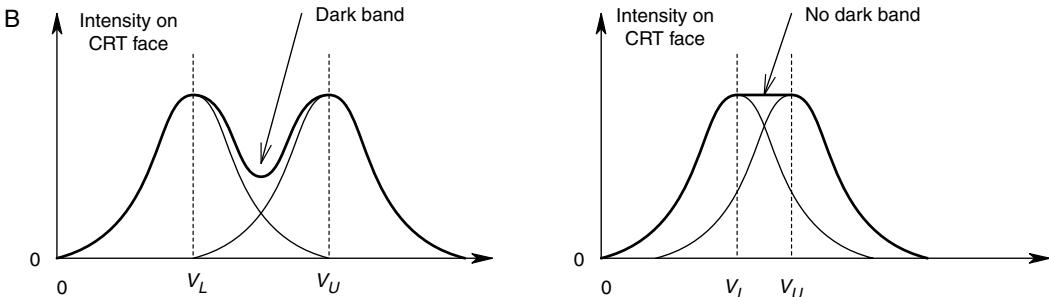
An approximate means of estimating the RMS value of a broadband, Gaussian noise source was described by Hewlett-Packard. The *tangential noise measurement method* is extremely simple. It makes use of a dual beam oscilloscope. The noise voltage is applied to both channels with the same gain. At wide trace separation, we see two, 'grassy' noise traces separated by a dark band. The vertical position control of one trace is then adjusted slowly to bring the grassy traces together to a critical separation, so that there is no dark band between the traces (i.e., there is a uniform illumination of the CRT screen between the traces). It can be shown that this condition corresponds to a trace separation of two standard deviations of the Gaussian noise. Referring to Figure 8.39,

A

Dual-beam oscilloscope



B

**FIGURE 8.39**

Noise measurement by the tangential method using a dual trace oscilloscope. A. Oscilloscope display showing two noise traces separated by a dc voltage, ($V_U - V_L$). The dc trace separation is adjusted until the noise appears as a solid band. B. Average CRT face intensity as a function of vertical deflection voltage.

we see that:

$$\sigma_x = \frac{V_U - V_L}{2} \quad (8.107)$$

If the measurement is repeated five times and the results averaged, the $\bar{\sigma}_x$ is claimed to be accurate to about ± 0.5 dB, or $\pm 6\%$.

8.6 AC Current Measurements

AC current measurements can be made over a wide range of frequencies and amplitudes. At audio and power frequencies, currents can be measured from nanoamperes to hundreds of kiloamperes. Different ranges require different techniques, which we shall describe below. At radio frequencies, the techniques available are considerably fewer. Regardless of the meter mechanism employed, ac ammeters are generally calibrated in terms of RMS of a sine wave.

8.6.1 Electromechanical (Analog) AC Ammeters

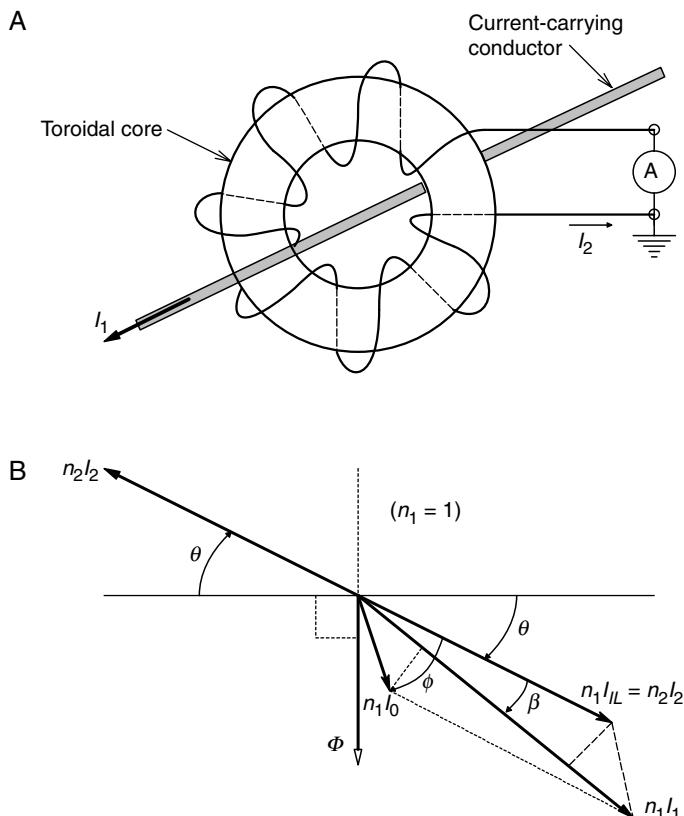
Most of the electromechanical meter movements we have already discussed in considering ac voltmeters can be used for ac ammeters, with attention to their range and input current frequencies. To measure currents above the full scale sensitivity of the meter movement, a shunt must be used. For currents in excess of about 5 A, a 50 mV external shunt is generally used. Recall that a 50 mV shunt has a resistance such that, when the full scale current enters it, 50 mV is developed across the parallel combination of the shunt and the 50 mV voltmeter which reads the current. Shunts are typically used for high current measurements at audio or power line frequencies using thermocouple, electrodynamometer, or iron vane millivoltmeters. They are not used with electrostatic or D'Arsonval/rectifier type ac meters. The thermocouple ammeter can be used at radio frequencies up to 50 MHz or so. The other meter movements, using coils, are only useful at power line frequencies.

Another means used at power line frequencies to measure high currents on conductors at high potentials is the *current transformer*. One or more turns of the alternating current carrying conductor are passed through the center of a high permeability, toroidal iron core, inducing a sinusoidally changing magnetic flux. Also wound on the toroidal core are a number of turns of wire which are connected to an ac ammeter, generally having a dynamometer or iron vane movement. Figure 8.40A illustrates a typical current transformer circuit. In a Midwest Electric Products, Inc. current transformer, the secondary (meter) winding was found to have a resistance of 0.04Ω , an inductance of 3.66 mH, and a turns ratio of 75/5. This transformer was rated for 2 VA, and works over a 25–400 Hz range. Stout (1950) gave a current transfer function for the current transformer as:

$$\frac{I_1}{I_2} \approx \frac{n_2}{n_1} \left[1 + \frac{I_o}{I_1} \sin(\theta + \phi) \right] \quad (8.108)$$

where n_2/n_1 is the ratio of secondary (output) winding turns to primary (input) turns, I_o is the excitation or magnetizing component of the primary current, I_1 , θ is the power factor angle between V_1 and I_1 and includes both the effect of the load (burden) and the current transformer. I_2 is the ac current in the secondary ammeter, and ϕ is the angle between the net magnetic flux in the current transformer's core, Φ_0 and the equivalent sine wave of I_o . A phasor diagram showing currents and flux in a typical current transformer is shown in Figure 8.40B. Current transformers should never be operated with their secondaries open circuited since dangerous high voltages can exist there.

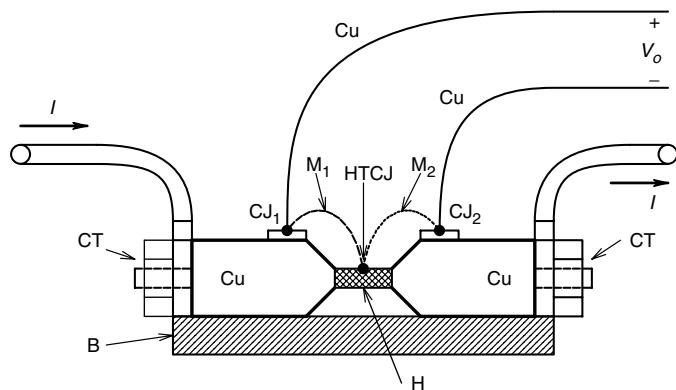
The *clamp-on ammeter* is a portable version of the current transformer in which the secondary load impedance is high, so sufficient secondary EMF is developed to operate an ac millivoltmeter, or a D'Arsonval/rectifier type ac voltmeter. Another meter

**FIGURE 8.40**

A. A current transformer used at power line frequencies. B. Phasor diagram showing the current and flux vectors in a typical current transformer.

movement used with a clamp-on ammeter is the rotating iron vane type used in the Columbia Electric Mfg. Co. 'Tong Test Ammeter'. In the Columbia meter movement, magnetic flux induced in the iron jaws by enclosing a current carrying conductor, acts to draw an iron vane into an airgap at the base of the jaws. This torque acts against a torsion spring, producing a pointer deflection which is a nonlinear function of primary current. Mechanical damping is produced by eddy currents induced in an aluminum damping pan. Ranges are changed in the Columbia clamp-on ac ammeter by plugging in different meter assemblies into the air gap in the clamp-on jaws. Full scale currents of 20–1000 A RMS are available. The bottom 20% of the Columbia meters' scales is not useful, because of deflection nonlinearity.

Yet another means of measuring ac current makes use of a *compensated thermocouple*, illustrated in Figure 8.41. Two, large copper blocks serve as attachment points for the input current line and as thermal reference points. The heater element is attached between the blocks, and the hot junction of the thermocouple is spot welded to the middle of the heater. The ends of the thermocouple wires are attached to copper strips, which are in intimate thermal contact with the copper blocks, but electrically insulated from them by mica sheets. The leads to the D'Arsonval microammeter are attached to the copper strips. The ac (or dc) input current causes a temperature rise in the heating element, which is sensed by the thermocouple. The voltage drop across the thermocouple's heater is designed to be 0.15 V at full scale current. Hence, this type of ammeter absorbs 150 mW for each ampere of full scale range. This type of ammeter operates at dc, at ac frequencies up to 50 MHz and for full scale currents of 5–50 A RMS [Stout, 1950].

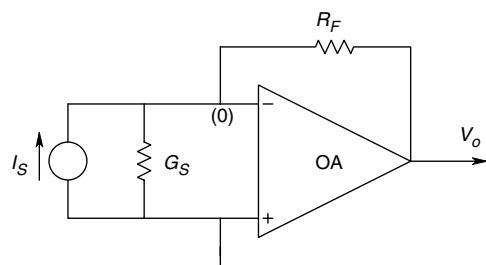
**FIGURE 8.41**

Section through a compensated thermoelement device used to measure the mean squared values of dc or ac currents. Note: B=bakelite base, Cu=copper blocks, CT=current input and output terminals, H=heater element, HTCJ=hot thermocouple junction, CJ1 and CJ2=compensating (cold) thermocouple junctions.

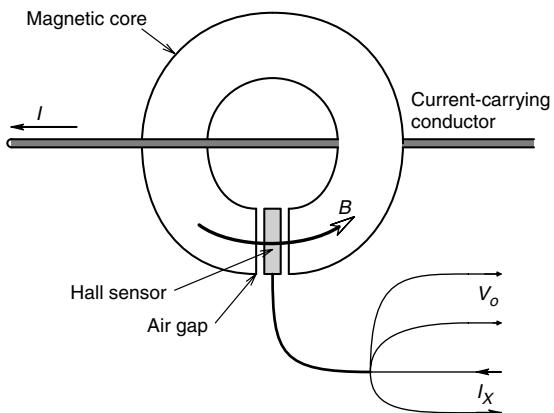
8.6.2 Electronic and Magneto-Optical AC Ammeters

To measure ac currents smaller than the full scale currents required to operate the inefficient electrodynamicmeter and iron vane meter movements (about 50 mA RMS) or the full scale current for vacuum thermocouple ac ammeters (about 3 mA RMS), it is necessary to amplify the current, or a voltage proportional to the current. The operational transresistor op-amp circuit offers a means of measuring very small ac currents flowing into ground, ranging from picoamperes to about 10 mA. In this case, the op-amp's summing junction appears as a virtual ground, as shown in Figure 8.42. The ac voltage at the output of the operational transresistor is simply $V_o = -R_F I_s$ and can now be displayed by any of the conventional ac, analog indicating meters we have discussed above, or digitally.

Another means of measuring low level ac currents is to use a current transformer followed by an ac amplifier, followed by an ac analog indicating meter, or digital readout. Pearson Electronics, Inc. makes a variety of wide bandwidth current sensors which have a voltage output. For example, the Pearson model 2877 current sensor has a 50 Ω output impedance, an open circuit transresistance of 1 V/A, a maximum peak current of 100 A, maximum RMS current of 2.5 A, a 2 ns rise time, a 0.2%/ μ s droop and -3 dB frequencies of 300 Hz and 200 MHz. At the other end of the current measurement range, the Pearson model 1423 has an open circuit transresistance of 0.001 V/A, a maximum peak current of 500,000 A, a maximum RMS current of 2500 A, a 300 ns rise time, a 0.7%/ms droop and

**FIGURE 8.42**

The operational transresistor circuit is used to convert a small current to a voltages, V_o .

**FIGURE 8.43**

A Hall sensor used to measure the ac current in a conductor. The Hall output voltage is proportional to the product of B_y and I_H , hence I .

-3 dB frequencies of 1 Hz – 1.2 MHz . Current transformers, such as the Pearson sensors, find application in measuring current transients in insulation breakdown studies, and in investigations of high power pulse circuits using SCRs, hydrogen thyratrons, strobe tubes, klystrons and magnetrons. Of course, steady state sinusoidal currents can also be measured. Since the maximum current sensitivity of the Pearson model 2877 sensor is 1 V/A , it is possible to resolve ac currents in the μA range using appropriate low noise amplifiers and bandpass filters.

Another means of measuring ac current in a conductor is to *measure the magnetic field around the conductor*. A sensitive Hall effect sensor probe (Section 6.4.6.2) can be put next to the conductor to intercept the solenoidal B field. The ac output voltage of the Hall sensor is amplified and then used to drive an appropriate meter. For more sensitive current sensing, a flux concentrating, iron 'C' core can be put around the conductor, as shown in Figure 8.43 and the Hall sensor put in its air gap. The ferromagnetic material and the length of the air gap must be chosen so that the magnetic material does not go into saturation at the desired, full scale, ac current to be measured.

The final ac current measuring system we will consider is the *Faraday magneto-optic sensor*, discussed in some detail in Section 6.7.1 of this text. A good, physical discussion of the Faraday magneto-optic effect can be found in Chapter 7 of the optics text by K. D. Möller (1988). Some of the earliest work on fiber optic Faraday effect current sensors was done by A.J. Rogers (1973). As shown in Figure 6.39 in this text, the Rogers system uses a ratio detector to make the system insensitive to amplitude variations in the laser source. Section 6.7.1 describes the operation of the ratio detector and shows that the output of the ratio detector is a sinusoidal voltage at power line frequency with peak amplitude proportional to the peak amplitude of the conductor by the Verdet constant. After amplification, this line frequency voltage can be displayed by one of the many ac responding meters we have described above. Other Faraday magneto-optic current sensor designs have been described by Cease and Johnston (1990) and Nicati and Robert (1988).

8.7 Magnetic Field Measurements

There are many applications in the design of rotating electric machines (alternators, dc generators and motors, brushless dc motors, stepping motors, induction motors, etc.)

where it is useful to be able to measure the magnetic field strength in air gaps in order to verify designs. There are also many applications in physics and chemistry, where it is necessary to measure magnetic fields (e.g., in mass spectrometers, magnetic resonance imaging systems and in particle accelerators). Recently, power line frequency magnetic fields have been implicated as possible contributing causes in health problems including birth defects, leukemia and cancer [EPRI Journal Jan/Feb 1990]. Thus, measurement of the spatial distribution of 60 Hz B fields in areas containing people, experimental animals or cell cultures is important for the clarification of the significance of 60 Hz B fields in health problems. The magnetic field vector, \mathbf{B} , is also called the *magnetic induction* or the *magnetic flux density*. Its units are newtons per coulomb divided by meters per second, or equivalently, volt seconds per square meter, or webers per square meter. The more common units for \mathbf{B} are gauss ($1 \text{ weber/m}^2 = 10^4 \text{ gauss}$), or teslas ($10^4 \text{ gauss} = 1 \text{ T}$). Magnetic flux is defined as the integral of \mathbf{B} over an area perpendicular to \mathbf{B} , or:

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{A} \quad (8.109)$$

The units of magnetic flux are webers (MKS), or maxwells, found by integrating one gauss over one square centimeter. The vector \mathbf{H} is called the *magnetic field*. It is related to \mathbf{B} by the well-known relation, $\mathbf{H} = \mathbf{B}/\mu$. The MKS units of \mathbf{H} are newton/weber or amperes/meter, and the CGS units of \mathbf{H} are oersteds ($1 \text{ A/m} = 4\pi \times 10^{-3} \text{ oersteds}$).

There are a variety of magnetic sensors which can be used to measure dc and ac \mathbf{B} fields in space. The first and most basic means is the solenoidal search coil. If the \mathbf{B} field is homogeneous over the area of the coil, a time varying \mathbf{B} field will induce an open circuit EMF given by the well-known relation:

$$E_M = NAB \dot{\theta} \cos(\theta) \quad (8.110)$$

where N is the number of turns in the coil, A is the area of the coil, θ is the angle \mathbf{B} makes with respect to a normal to the coil's area, and $\dot{\theta} = dB/dt$. In most power line \mathbf{B} fields, $B(t) = B_0 \sin(\omega t)$, so:

$$e_M(t) = N A \cos(\theta) \omega B_0 \cos(\omega t) \quad (8.111)$$

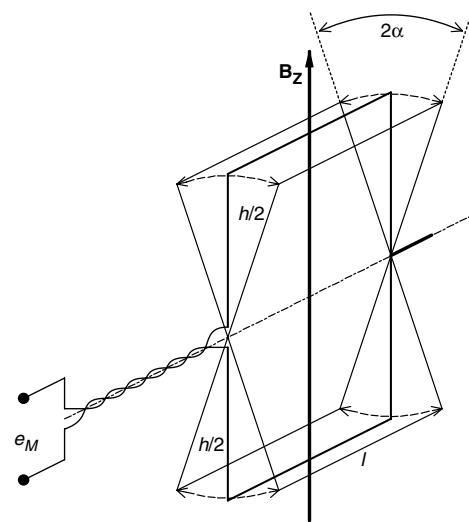
If the \mathbf{B} field is constant (e.g., the Earth's magnetic field), an EMF may be induced in the coil proportional to \mathbf{B} by rotating the coil at an angular velocity of ω_c r/s. In this case, the $\cos(\theta)$ term becomes $\cos(\omega_c t)$ and the induced EMF is given by:

$$e_M(t) = NAB \omega_c \sin(\omega_c t) \quad (8.112)$$

Several variations on the rotating coil magnetic field meter were described by Lion (1959). For example, a coil can be mechanically sinusoidally oscillated around a center position in which a constant \mathbf{B} vector is parallel with the plane of the coil, as shown in Figure 8.44. The EMF induced in the coil can be shown to be given by:

$$e_M(t) = NAB \alpha_o \omega_r \cos(\omega_r t) \quad (8.113)$$

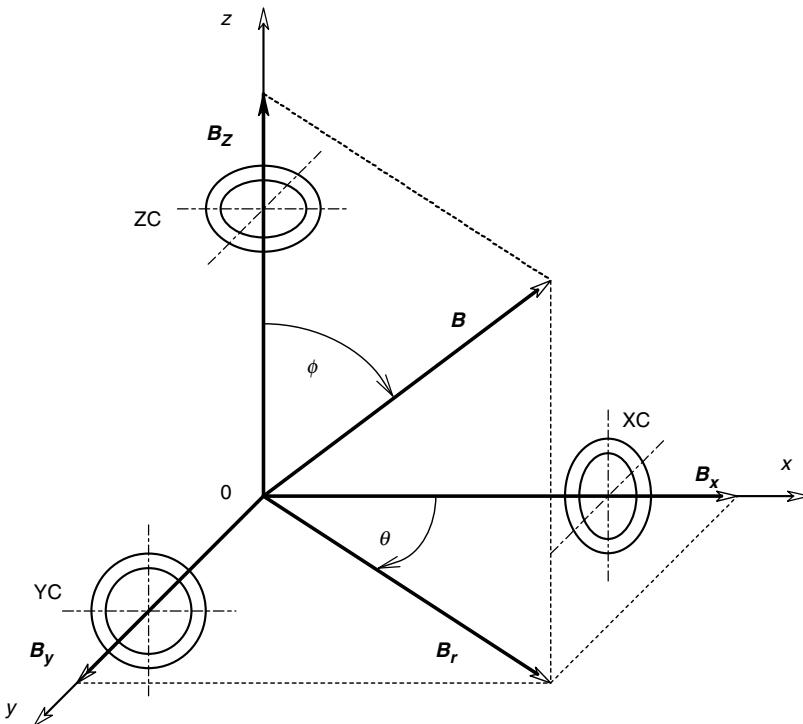
Here, α_o is the angle over which the coil is oscillated at mechanical radian frequency, ω_r , so $\alpha(t) = \alpha_o \sin(\omega_r t)$. We assume that $\alpha_o < 15^\circ$.

**FIGURE 8.44**

An oscillating search coil used to measure a dc magnetic field. (Lion, 1959)

Three orthogonal search coils can be used to determine the magnitude and direction of a power line frequency \mathbf{B} field. The \mathbf{B} vector can be resolved into three orthogonal vector components, B_x , B_y and B_z . Each component may be considered to produce an EMF in its respective search coil. Referring to Figure 8.45, we see that in spherical coordinates, the three components of \mathbf{B} can be written as scalars:

$$B_z = B \cos(\phi) \quad (8.114a)$$

**FIGURE 8.45**

Three, orthogonal search coils used to measure the x , y and z components of a time varying magnetic field.

$$B_X = B \sin(\phi) \cos(\theta) \quad (8.114b)$$

$$B_Y = B \sin(\phi) \sin(\theta) \quad (8.114c)$$

where ϕ is the co-latitude angle measured between the positive y axis and \mathbf{B} , and θ is the longitude angle, measured between the positive x axis and the projection of \mathbf{B} on the xy plane. This projection is \mathbf{B}_r and is equal to:

$$B_r = B \sin(\phi) \quad (8.115)$$

From equation 8.113 above, we can write expressions for the instantaneous EMFs induced in the coils:

$$e_{MX}(t) = NA \omega B \sin(\phi) \cos(\theta) \cos(\omega t) \quad (8.116a)$$

$$e_{MY}(t) = NA \omega B \sin(\phi) \sin(\theta) \cos(\omega t) \quad (8.116b)$$

$$e_{MZ}(t) = NA \omega B \cos(\phi) \cos(\omega t) \quad (8.116c)$$

Now, it is easy to show that if we square each of the coil EMFs with four-quadrant analog multipliers, average each multiplier output by passing it through a low-pass filter, add together the mean square voltages, then take the square root, we will obtain a dc voltage proportional to B , regardless of the orientation of the three coil probe in the \mathbf{B} field. This process is illustrated in Figure 8.46. Numerically, B can be shown to be:

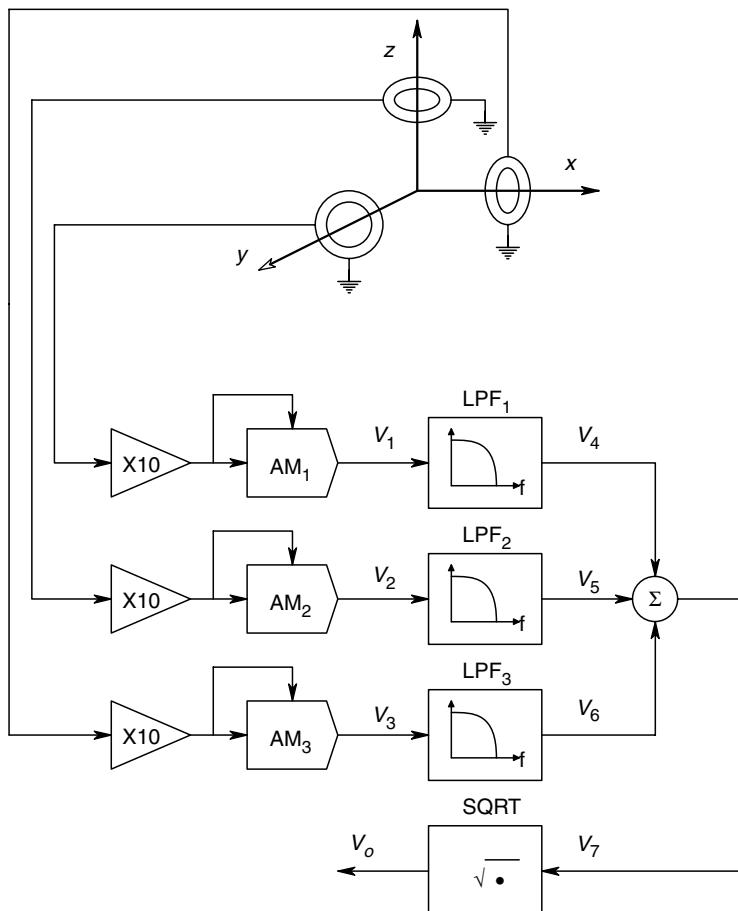
$$B = \frac{\sqrt{2} V_o}{NA \omega} \quad (8.117)$$

(We assume all coils to have identical N s and A s.) The angles ϕ and θ can also be found relative to the probe's x , y and z axes from trigonometric identities, although this is tedious. A better way to find the direction of \mathbf{B} is to experimentally orient the probe so that $e_{MX}(t)$ is maximum, and $e_{MY}(t)$ and $e_{MZ}(t)$ are minimum. Now the angle of \mathbf{B} is parallel with the probe's x axis.

Analog Hall effect sensors provide another major means of measuring either constant or time varying magnetic fields. The physics governing the behavior of Hall sensors were described in detail in Section 6.4.6.2. It was shown that the Hall sensor EMF is given by:

$$E_H(t) = \frac{B(t) \cos(\theta) I_x(t) R_H}{h} \quad (8.118)$$

where θ is the angle \mathbf{B} makes with the xz plane, $I_x(t)$ is the current flowing in the $+x$ direction (electrons in the n -semiconductor drift in the $-x$ direction) and R_H is the Hall constant, where $R_H = -1/qd$ for an n -semiconductor, and $R_H = 1/qd$ for p -semiconductor. q is the magnitude of the electron charge, d is the density of the carrier doping in the semiconductor and h is the thickness of the Hall slab in the y direction. In the operation of most analog Hall sensors, I_x is generally dc. However, in the measurement of a low level, constant \mathbf{B} , a sinusoidal or square wave, $I_x(t)$ can be used to permit noise reduction

**FIGURE 8.46**

Electronic system to compute the magnitude of the $\mathbf{B}(t)$ vector using the Pythagorean theorem.

by the amplification of $e_H(t)$ out of the $1/f$ noise region, and also, synchronous rectification can be used to further improve SNR.

Modern analog Hall sensors can be purchased with built-in IC dc amplifiers, or as the basic sensor chip. Unamplified Hall sensors have conversion gains ranging 10–55 mV/kG. Through the use of a 9 in. flux concentrating iron bar, the F. W. Bell model BH-850 Hall sensor has a gain of 18 mV/gauss. Normal operating current for the Bell BH-850 sensor is 200 mA. The resistance in its current path and its Thevenin output resistance for E_H are both 3.5Ω , giving good low noise performance.

Various companies make gaussmeters for various applications having either analog or digital meter readouts. For example, The Bell model 615 digital gaussmeter, a benchtop instrument, reads $10\text{--}10^6$ gauss in five ranges, has a dc to 2 kHz frequency range, is accurate to $\pm 0.5\%$ of FS and has a readout resolution of 1 part in 10^3 . The Bell model 4048 is a handheld, portable, digital gauss/tesla meter, with three full scale ranges—200, 2000 and 2×10^4 G. Accuracy is $\pm 2\%$ of reading for dc fields, $\pm 2.5\%$ for 45–100 Hz ac fields, $\pm 2\%$ for 100 Hz–3 kHz, and $\pm 6\%$ for 3–5 kHz fields. This instrument is available with transverse and axial Hall probes.

The practical lower limit of Hall sensor sensitivity is set by noise, both from the sensor's thermal and shot noise, as well as from the signal conditioning amplifier's e_{na} and i_{na} .

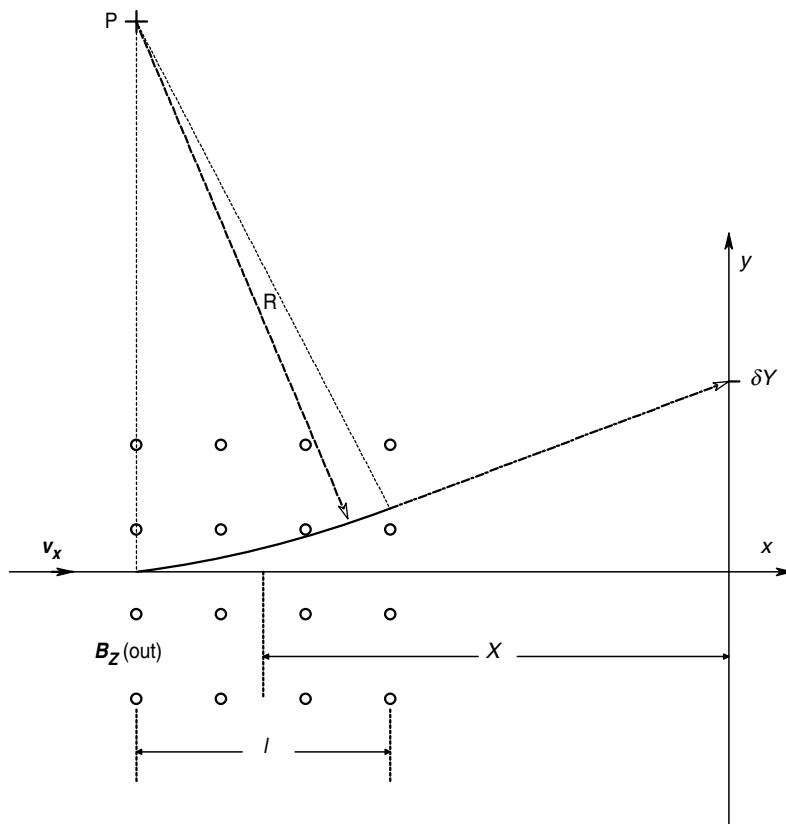
**FIGURE 8.47**

Diagram showing the behavior of a moving electron which enters a region of uniform magnetic field (\mathbf{B}_z out of page) of length *L*. The electron crosses the *y* axis with a net deflection from the center line of δY .

Milligauss resolution is practical with most modern Hall sensors. To achieve higher sensitivity in magnetic field measurements, it is necessary to use sensors which utilize different physical principles.

For example, at microwave frequencies, many materials exhibit a very strong Faraday magneto-optic effect (Section 6.6.1). The Verdet constant of certain ferromagnetic ferrites at 9 GHz is on the order of 0.1°/(oersted cm). By the use of such a ferrite, magnetic flux density variations on the order of 10^{-5} gauss have been detected [Lion, 1959].

The deflection of a collimated, low energy electron beam by a \mathbf{B} field is also an effective means of measuring low values of \mathbf{B} , ranging from 2×10^{-6} to 2×10^{-4} gauss [Lion, 1959]. Figure 8.47 illustrates the geometry of the CRT used for such a measurement. Note that \mathbf{B} must be perpendicular to the electron beam axis and the deflection axis. The electron beam is accelerated through a potential of *V* volts, giving the electrons in the beam an average velocity of:

$$\mathbf{v}_x = \sqrt{2V(q/m)} \mathbf{i}_x \text{ m/s} \quad (8.119)$$

where q/m is the ratio of electron charge to electron mass, and \mathbf{i}_x is a unit vector in the $+x$ direction. The beam passes through *L* meters of a uniform magnetic field of strength \mathbf{B}_z .

A Lorentz force of:

$$\mathbf{F}_Y = -q(\mathbf{v} \times \mathbf{B}) = -(q v_X B_Z) \mathbf{i}_Y \text{ N} \quad (8.120)$$

acts on each electron, causing the beam to deflect along a circular path with a radius R . R may be shown to be equal to:

$$R = (m v_X) / (q B_Z) \text{ m} \quad (8.121)$$

Krause (1953) shows that the beam deflection, δY , can be used to measure B_Z . Referring to Figure 8.47, B_Z can be expressed as:

$$B_Z = \frac{\delta Y}{L X} \sqrt{2 V(m/q)} \text{ Wb/m}^2 \quad (8.122)$$

Lion (1959) illustrates a means of detecting beam deflections, δY , too small to visualize on the phosphor screen of the electron beam magnetometer. The collimated beam strikes a split anode, shown in Figure 8.48. At zero B_Z , the beam is positioned so that exactly half its electrons strike each anode plate.

The anode currents are converted to two voltages which are subtracted by a dc differential amplifier. In the presence of a $B_Z > 0$, the beam deflects and more current strikes one plate than the other, causing an output signal. If this output signal is integrated and then conditioned to make a current to drive a deflection coil, the effective flux density of the deflection coil will be equal to $-B_Z$ in the steady state, giving a null output.

The most sensitive magnetic field sensors are the SQUID magnetometers. SQUID stands for *Superconducting QUantum Interference Device*. SQUIDs find application in biophysical measurements in neurophysiology. The nerve action potentials are accompanied by transient flows of electric currents, generally carried by ions. These transient

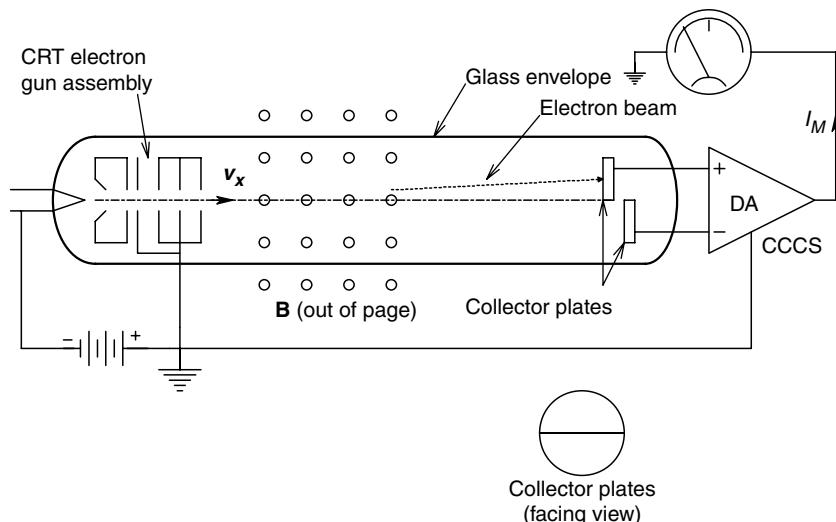


FIGURE 8.48

Use of a cathode ray tube to measure B_Z . The Lorentz force on the electrons passing through the orthogonal B field causes a deflection, Δy .

currents generate minute, time varying, magnetic fields which can be measured and located spatially in the brain by arrays of SQUID sensors [Ribeiro, Williamson and Kaufman, 1988; Northrop, 2002]. Besides studies in neuromagnetism, SQUIDs can be used in magnetocardiography, detection of changes in the Earth's magnetic field in geological studies, etc.

SQUIDs are made from various superconductor materials and are operated at the boiling point of liquid He at atmospheric pressure (4.2 K). There is a possibility that SQUIDs can be made from special, high temperature semiconductors which can operate at temperatures up to 125 K [Clarke and Koch, 1988].

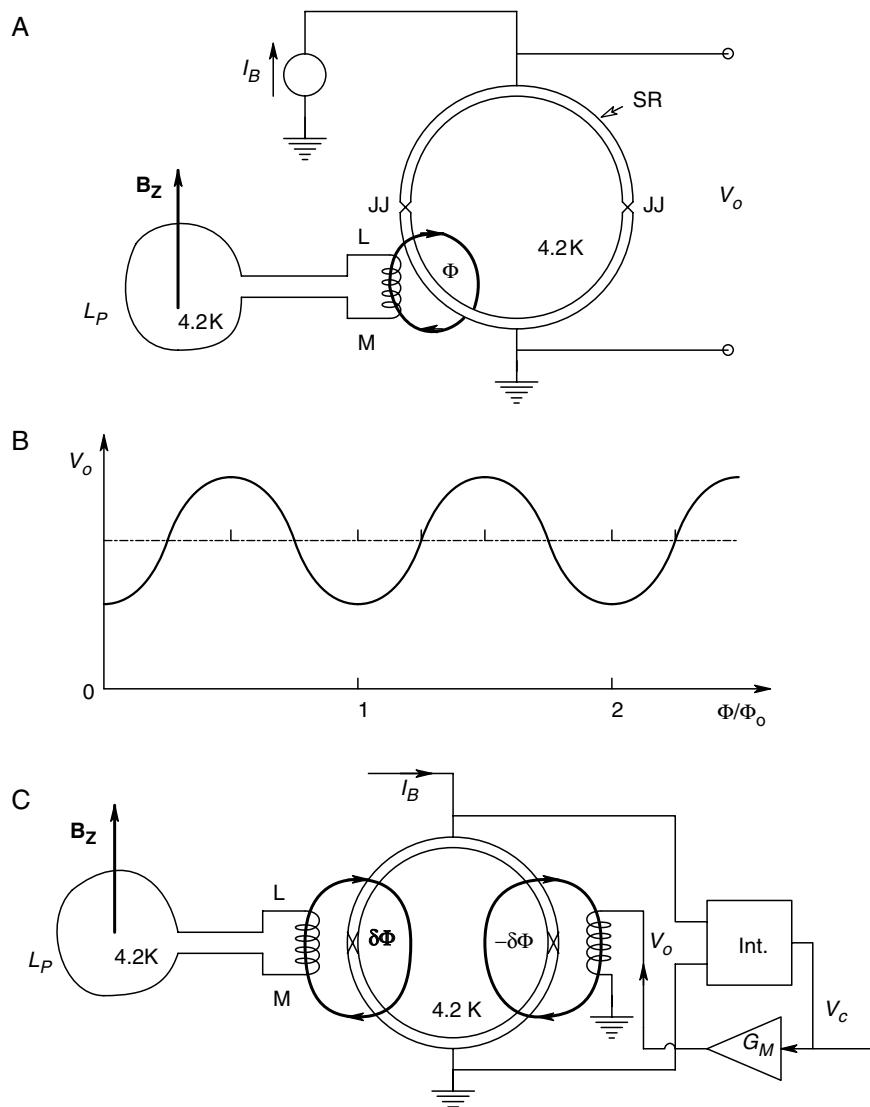
Conventional dc SQUIDs are the most sensitive magnetometers available. They are made from a superconducting ring of niobium or lead, and operated in liquid helium. The ring has an inductance, L , in the picohenry range and is joined with two Josephson junctions, as shown in Figure 8.49A. The SQUID is really a four-terminal device—two terminals are used to input a dc bias current, I_B and the same two terminals to monitor the output voltage, V_o . The voltage, V_o , remains zero until the bias current reaches a critical value, I_c . Then the output voltage increases with current and is also a function of the magnetic flux linking the SQUID ring. The dc bias current, I_B , is made greater than I_c . The superconducting SQUID ring circuit undergoes the phenomenon of *fluxoid quantization* in which the magnetic flux linking the SQUID is given by $n\Phi_0$, n is an integer and Φ_0 is the flux quantum, equal to $h/2q = 2 \times 10^{-15}$ Wb. If we apply an additional flux, Φ_i , through the SQUID ring, a supercurrent, $I_S = -\Phi_i/L$ is set up in the ring to create a flux which cancels Φ_i . In other words, $LI_S = -\Phi_i$. From Figure 8.49B, we see that at a constant bias current, the SQUID output voltage varies periodically as a function of Φ/Φ_0 . This phenomenon is a lot like the generation of intensity variations due to interference rings in an optical system, such as a Sagnac gyro. The SQUID sensor is often operated as a feedback device, as shown in Figure 8.49C. A bias flux is adjusted so that V_o is linearly proportional to small changes in input flux, $\delta\Phi$. V_o is then conditioned and integrated and used to control the current in a feedback coil such that a flux $-\delta\Phi$ is generated to cancel the input flux. This system operates the SQUID as a null-flux detector and is called a flux-locked SQUID. This closed loop operation permits a large dynamic range of input flux to be measured. As shown in Figure 8.49A, a superconducting input coil must be used to couple the flux to be measured into the SQUID toroid. It appears that optimum SQUID sensitivity occurs when the inductance of the pickup loop, L_p , is equal to that of the coupling loop, L_C . Assuming noiseless flux coupling, Clarke and Koch (1988) showed that the equivalent flux density noise of the SQUID can be given by:

$$B_N(f) = \frac{2\sqrt{2L_p}}{\pi r_p^2} \sqrt{\frac{\varepsilon(f)}{\alpha^2}} \text{ W/m}^2/\sqrt{\text{Hz}} \quad (8.123)$$

where α is the coupling coefficient between the squid and the input coil, r_p is the radius of the input coil, L_p is the inductance of the pickup coil and $\varepsilon(f)$ is the white noise energy/Hz of the SQUID:

$$\varepsilon(f) = 9kT L/R \text{ J/Hz} \quad (8.124)$$

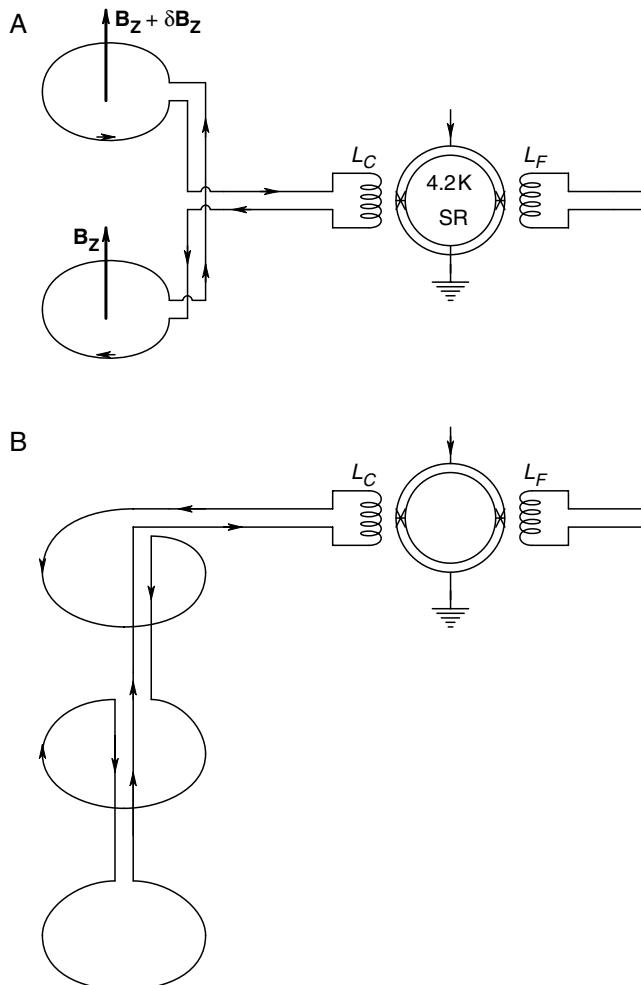
where k is Boltzmann's constant, T is in K, L is the SQUID's inductance and R is the resistance shunting each Josephson junction in the SQUID ring. For a typical Nb SQUID with $100 < L < 500$ pH, the noise energy is about 1 to 5×10^{-32} J/Hz [Clarke and Koch, 1988]. With such low noise, it is possible to resolve magnetic flux densities on the

**FIGURE 8.49**

A. Schematic of a dc SQUID device. Note: SR = superconducting ring of lead or niobium, JJ = Josephson junctions, I_B = dc bias current, L = inductance of the coupling coil, M = mutual inductance between coupling coil and the SR. B. Variation of SQUID output voltage at constant bias current as a function of the ratio of the applied flux to the flux quantum. C. A feedback or flux locked SQUID.

order of $10^{-14} \text{ T}/\sqrt{\text{Hz}}$ (1 tesla = 10^4 gauss). A very low white noise of $5 \text{ fT}/\sqrt{\text{Hz}}$ has been achieved using thin film technology. Of course, external magnetic noise pickup can confound sensitive measurements, so SQUID sensors are normally well shielded with layers of Mu metal, aluminum and even superconducting materials.

To further reduce the pickup of unwanted magnetic field noise, use is made of a *gradiometer coil*, shown in Figure 8.50A. This coil responds to spatial non-uniformities in the field under measurement (i.e., $\partial B/\partial z$). First order gradients from a magnetic dipole fall off as $1/r^4$, so the gradiometer coil discriminates against distant sources of interference in favour of local dipoles. Some workers have even used second derivative

**FIGURE 8.50**

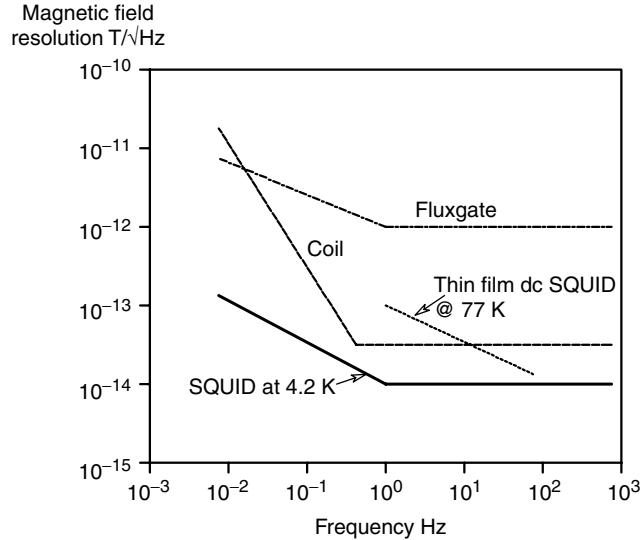
A. A SQUID with a gradiometer flux pickup. B. A SQUID with a second derivative gradiometer flux pickup.

gradiometer coils which respond to $\partial^2 B / \partial z^2$ to obtain even better resolution of local sources [Ribeiro, Williamson and Kaufman, 1988]. A second derivative coil is shown in Figure 8.50B.

A comparison of the magnetic flux density resolution of common magnetometers is shown in Figure 8.51, adapted from Clarke and Koch (1988). Note that the non-superconducting coil gives very high resolution at frequencies above 1 Hz and is generally robust and simple. The extra sensitivity of SQUID sensors is needed, however, for biomedical applications such as neuromagnetism and magnetocardiography.

8.8 Phase Measurements

It is often necessary to measure the phase difference between two periodic signals of the same frequency. In power distribution systems, the power factor angle between the ac supply voltage and the load current is of great economic and practical importance to the power generator and the end user. We have seen that several magneto-optical electric

**FIGURE 8.51**

A comparison of the resolution and bandwidths of various sensitive magnetic field sensors. (Clarke and Koch, 1988)

current sensing systems have ac outputs in which the phase difference is proportional to an optical polarization angle change, induced by the magnetic field produced by the current under measurement. Also, phase information is important in the description of feedback system performance and filter behavior.

Phase measurements can be classified as being done by analog or digital systems. To illustrate the concept of phase difference, let us define two periodic functions having the same frequency, but not coincident in time origin:

$$v_1(t) = v_1(t + T) = v_1(t + kT) \quad (8.125a)$$

$$v_2(t) = v_2(t + T + \Delta T) = v_2(t + kT + \Delta T) \quad (8.125b)$$

where ΔT is the time difference between v_1 and v_2 , T is the period of v_1 and v_2 , and k is an integer. Figure 8.52A illustrates two sine waves having a phase difference. v_2 is said to lead v_1 by a phase difference of φ , where:

$$\varphi = 2\pi(\Delta T/T) \text{ radians} \quad (8.126a)$$

or

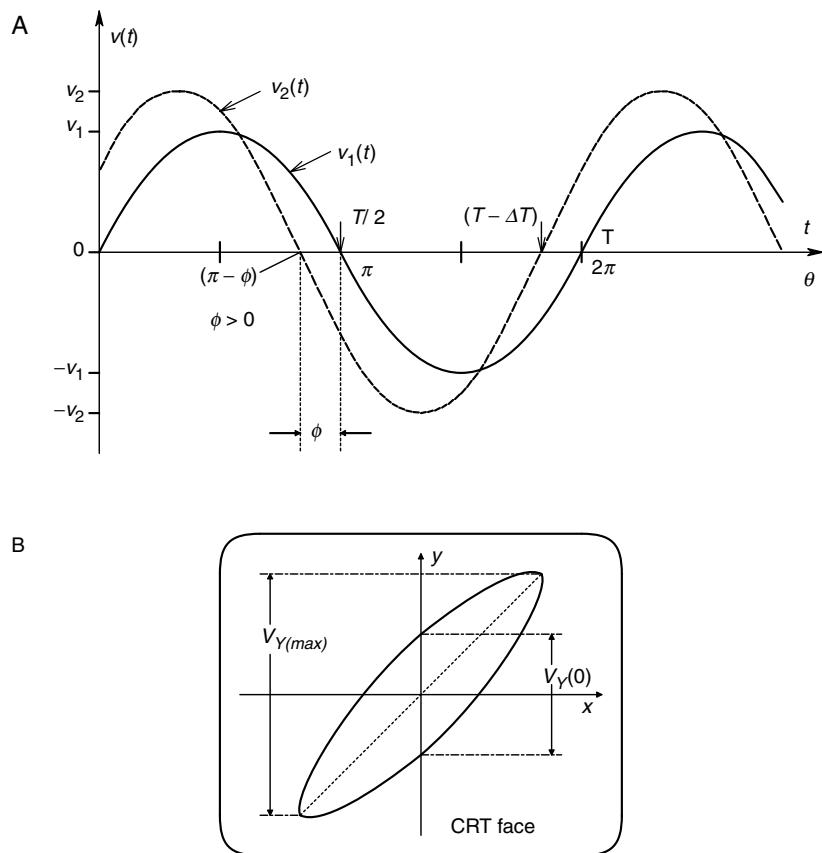
$$\varphi = 360(\Delta T/T) \text{ degrees} \quad (8.126b)$$

if

$$v_1(t) = V_1 \sin(\omega t) \quad (8.127a)$$

and

$$v_2(t) = V_2 \sin(\omega t + \varphi) \quad (8.127b)$$

**FIGURE 8.52**

A. Figure illustrating the phase difference between two sine waves having the same frequency and different amplitudes. B. A Lissajous' figure on a dual beam oscilloscope face.

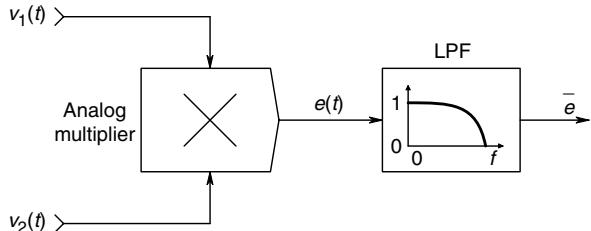
8.8.1 Analog Phase Measurements

Probably the simplest and easiest direct measurement of phase angle can be made on an oscilloscope using the *Lissajous' figure* formed when two sine waves, given by equation 8.127 are input to the x (horizontal deflection) axis and the y (vertical deflection) axis, respectively. In general, we see an ellipse on the CRT screen, as shown in Figure 8.52B. The ellipse must be centered at the origin (center) of the CRT screen. At $t=0$, $v_1=0$ and $v_2 = V_2 \sin(\varphi)$. Thus, it is easy to show that the phase angle is given by:

$$\varphi = \sin^{-1}[V_Y(0)/V_Y(\max)] \quad (8.128)$$

From equation 8.128, when the Lissajous' figure is a straight line, $\varphi=0^\circ$ and when it is a circle, $\varphi=90^\circ$. Accuracy of the Lissajous figure method of phase measurement is poor near $\pm 90^\circ$. At phase angles near integer multiples of 180° , accuracy is better and is limited by the thickness of the oscilloscope trace, as also one's ability to estimate the distances to its intersections with the vertical axis of the CRT's graticule.

Somewhat more accurate analog phase measurement techniques have been devised based on waveform averaging. Most of these techniques have been used in phase lock loops to implement the phase detector element [Northrop, 1990].

**FIGURE 8.53**

The analog multiplier (quadrature) phase detector.

The first analog phase detector we will describe is the analog multiplier/low-pass filter. In this system, shown in Figure 8.53, the input signals must be in quadrature for proper operation. That is,

$$v_1(t) = V_1 \sin(\omega_1 t) \quad (8.129a)$$

and

$$v_2(t) = V_2 \cos(\omega_1 t + \varphi) \quad (8.129b)$$

The output of the analog multiplier is:

$$e(t) = (V_1 V_2 / 20) [\sin(2\omega_1 t + \varphi) - \sin(\varphi)] \quad (8.130)$$

After passing through the low-pass filter, we have the dc component of $e(t)$:

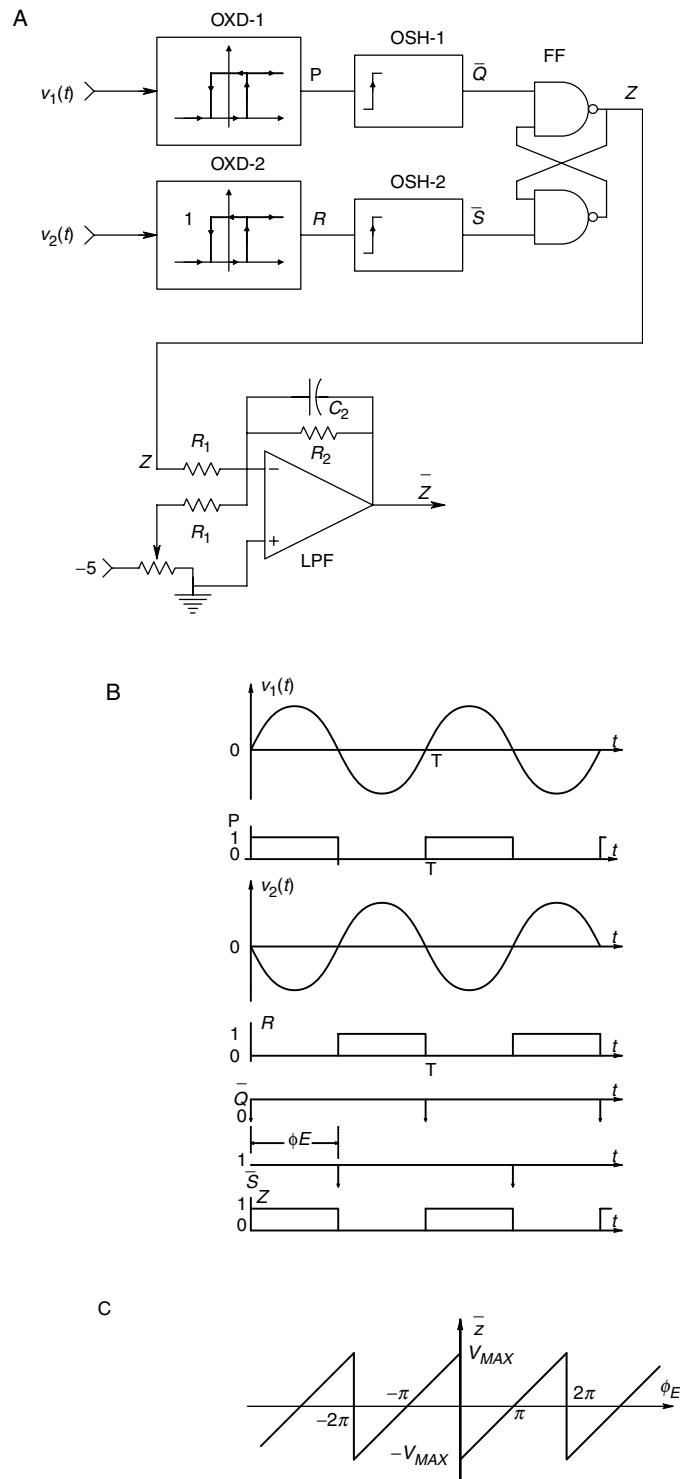
$$\bar{e} = -(V_1 V_2 / 20) \sin(\varphi) \cong -(V_1 V_2 / 20) \varphi \quad (\varphi \text{ in radians for } \varphi < 0.25R) \quad (8.131)$$

The sign change is the result of trigonometrical algebra. Linearity in this phase detector is preserved for $\varphi < 12^\circ$, otherwise an inverse sine nonlinearity must be used to recover a voltage proportional to φ .

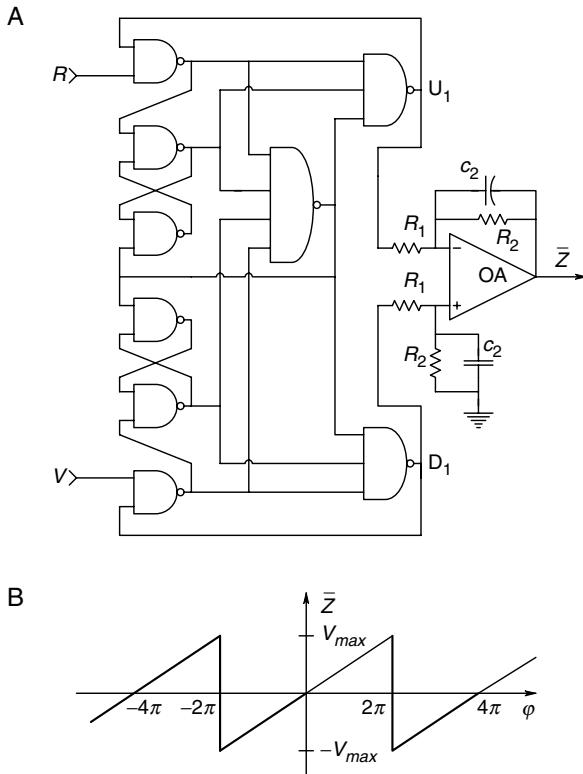
A *NAND gate flip-flop (FF) phase detector*, which works with TTL signals is shown in Figure 8.54A. The input analog sine waves are converted to TTL square waves by amplitude comparators connected as zero crossing detectors with hysteresis for noise rejection. The TTL square wave outputs from the comparators are further conditioned by one-shot multivibrators such as 74LS123s, which have narrow complimentary outputs with dwell times generally much less than $T/100$. A simple NAND gate R-S flip-flop produces an output whose duty cycle depends on the phase difference between the input signals. As seen from the waveforms of Figure 8.54B, the FF output has a 50% duty cycle when φ is 180° . The FF output, $Z(t)$, is inverted and averaged by the op-amp low-pass filter and the dc level is adjusted so that, when $\varphi = 180^\circ$, the op-amp output voltage, \bar{Z} , is zero. A plot of \bar{Z} vs φ is shown in Figure 8.54C; note that it is periodic.

An *exclusive NOR gate* can also be used for an analog phase detector. The input signals are converted to TTL square waves for input to the ENOR gate; the gate's output is a double frequency TTL pulse train whose duty cycle varies linearly with φ . As in the case of the RSFF phase detector, the ENOR's output, $W(t)$, is averaged by low-pass filtering and the dc level is adjusted. The dc op-amp output, W , is also a periodic function of φ .

A third, logic based phase detector IC utilizes NAND gates. Figure 8.55A illustrates the simplified innards of the *Motorola MC4044* phase detector and shows how its two outputs can be connected to an op-amp circuit which does low-pass filtering and dc

**FIGURE 8.54**

A. Schematic of a NAND gate, RS flip-flop phase detector. Note: OXD = zero crossing detector, OSMV = one-shot multivibrator (e.g., 1/2 74123), FF = RS flip-flop, LPF = low-pass filter. B. Waveforms in the NAND gate flip-flop phase detector. C. Note the periodic form of this phase detector's transfer curve which has zeros at odd multiples of 180° phase difference.

**FIGURE 8.55**

A. Combinational logic used in the Motorola MC4044 digital phase detector. B. I/O characteristic of the MC4044 phase detector.

adjustment. The I/O characteristics of the 4044 phase detector are shown in Figure 8.55B. Note that unlike the analog phase detectors described above, the output of the 4044 phase detector system is zero when $\varphi=0$, and the linear range extends to a full $\pm 360^\circ$.

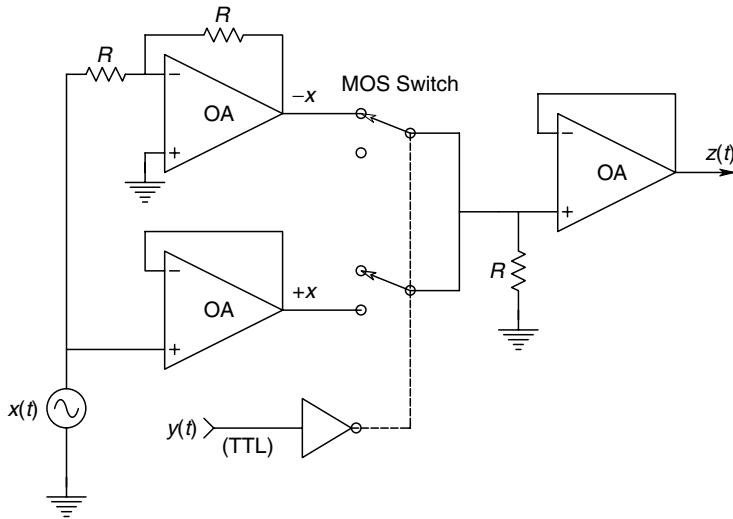
Other analog phase detectors are of the *switched amplifier design*. Such switched detectors are also of the quadrature type. In effect, the signal $v_2(t) = V_2 \cos(\omega_1 t + \varphi)$ is multiplied by the function, $\text{SGN}[v_1(t)]$. $\text{SGN}[x]$ is the well-known signum function; $\text{SGN}[x < 0] = -1$, $\text{SGN}[x \geq 0] = +1$. Thus:

$$z(t) = K_A \text{SGN}[v_1(t)] V_2 \cos(\omega_1 t + \varphi) \quad (8.132)$$

From a consideration of the Fourier series for the $\text{SGN}[v_1]$ square wave, it is easy to show that for $\varphi < 12^\circ$, the average of $z(t)$, $\bar{z} = K_D \varphi$. An example of a switched amplifier phase detector is illustrated in Figure 8.56.

8.8.2 Digital Phase Detectors

We have seen above that analog phase detectors generate waveforms whose average values are proportional to the phase shift, φ . Digital phase meters, on the other hand, measure the time interval, ΔT , as set forth in equation 8.125b. The time interval is measured by counting clock pulses generated from a precision quartz crystal oscillator between successive, positive-going, zero crossings of v_1 and v_2 . For example, the Stanford Research Systems Model SR620 Universal Time Interval Counter uses a precision 90 MHz clock to count the intervals between the successive positive zero crossings of v_1 and v_2 .

**FIGURE 8.56**

A switched amplifier phase detector. This is a quadrature detector with zeros at odd multiples of 90° phase difference between the sine wave input at $x(t)$ and the TTL square wave at $y(t)$. The output, $z(t)$, must be low-pass filtered to recover V_o proportional to $\sin(\theta_E)$, where θ_E is the phase difference between a cosine wave applied at x and a 'sinusoidal' square wave at y .

It also measures the period, T_1 of v_1 . An internal microprocessor system then computes and displays

$$\varphi = 360^\circ \Delta T/T_1 \text{ degrees} \quad (8.133)$$

The SRS SR260 phase meter works in the range 0–100 MHz, has a total range of $\pm 180^\circ$, a resolution of $(25 \times 10^{-12} \times (1/T_1) \times 360 + 0.001)^\circ$ and an error of less than $\pm(10^{-9} \times (1/T_1) \times (360 + 0.001))^\circ$. The SR620 permits the averaging of a preset number of φ s to compute a mean φ for greater accuracy in the presence of phase noise. The SRS SR620 is an example of a state-of-the-art instrument having an IEEE488 instrument/computer interface, and internally generated display voltages so that, in the absence of a host computer, the instrument output and status can be displayed on any XY lab oscilloscope.

We next consider the millidegree phase meter of Du and Northrop. This system was described in detail in Section 7.4.2, and is illustrated in Figure 7.63. A phase locked loop is run from the reference input signal as a phase locked frequency multiplier. Originally designed to be used in an optical system in which the chopping frequency was 100 Hz, the Du and Northrop system generates a clock frequency which is 360,000 times the input frequency, or 36 MHz. Due to the 360,000:1 frequency multiplication of the PLL, there will always be 1 clock pulse per millidegree of input phase, regardless of any small drift in input frequency. The high frequency PLL clock pulses are gated to a cascade of three high speed, TTL counters (74F579) during the interval, ΔT , where $v_1 > 0$ and $v_2 \leq 0$. The total count is obviously equal to φ in millidegrees. To improve accuracy in the presence of phase noise, the cumulative counts for $N = 2^k$ intervals, ΔT_n , can be averaged ($k = 0, 1, \dots, 8$ and $n = 1, 2, \dots, N$). The upper bound on the input frequency of the Du and Northrop phase measuring system is set by the PLL VCO (50 MHz for an NE564) or the counters (115 MHz for 74F579s). Also, the range of frequency drift tolerated by the system is set by the lock range of the PLL. Another system limitation is set by counter overflow; a φ over 46.6° will cause the counters to exceed their maximum count of 2^{24} .

8.9 Frequency and Period (Time) Measurements

There are two basic definitions of frequency—average frequency, measured as events (e.g., zero crossings) per unit time, and instantaneous frequency, measured as the reciprocal of the time intervals between successive events. Most modern frequency meters are digital systems which measure average frequency by counting events (cycles) over an accurately determined gate time. However, there are several other less accurate means of measuring frequency which may be used. Some of these methods compare the unknown frequency with a frequency standard. These include the Lissajous figure viewed on an oscilloscope, the zero beat heterodyne method, the FFT spectrum analyzer and the tunable filter.

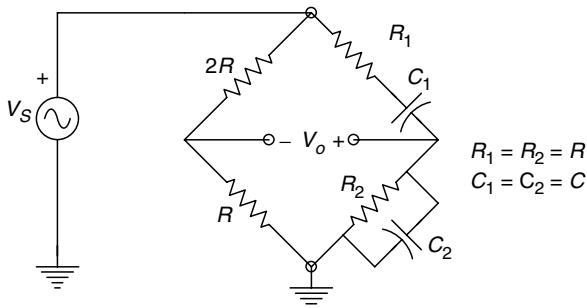
The oscilloscope can be used to measure frequency by comparing the unknown frequency to an accurate, variable frequency standard by a direct observation of the Lissajous' figure on its CRT. The Lissajous pattern is made by putting the waveform with the unknown frequency on the y (vertical) axis, and the known frequency waveform on the x (horizontal) axis. When the known frequency is adjusted so that the observed pattern is either a stationary straight line, ellipse or circle, the known frequency exactly equals the unknown frequency. The Lissajous' method works best at audio and low radio frequencies.

The zero beat method also makes use of a variable frequency, standard sinusoidal source which is mixed with the unknown frequency signal. Recall that mixing, in its purest form is multiplication. Mixing can also be accomplished by adding the two signals together, and then passing them through a nonlinear amplitude transfer function, such as $y = a + bx + cx^2 + dx^3$, etc. The output from the mixing process can be shown to contain sine waves having frequencies which are the sum and difference of the two input frequencies. Thus, as the variable known frequency is adjusted close to the unknown frequency, the frequency of the output difference frequency term approaches zero. This zero beat phenomenon can be detected by actually listening to the audio beat frequency resulting from mixing the standard frequency source with the unknown. The beat frequency method has an apparent dead zone, where the mixed difference frequency lies below the human range of hearing. Visual observation of the beat frequency signal on an oscilloscope or dc voltmeter can partially overcome the dead zone problem.

Another means of estimating the frequency of a coherent signal is to observe its RMS spectrum on the display of a fast Fourier transform (FFT) spectrum analyzer. In such instruments, the frequency resolution is determined by the sampling frequency and the number of samples. Resolution can be as high as one part in 1024 in a 1024 point transform and is obviously lower for fewer points.

The Wien bridge, which was discussed in Section 5.5.15 as a means of measuring capacitance. The Wien bridge is also unique in that its output exhibits a null at its tuned frequency, behaving like a passive notch filter. When the bridge is configured as shown in Figure 8.57, it is easy to show that the output null occurs when $f = 1/(2\pi RC)$ Hz. Again, the null can be detected at audio frequencies by headphones, or by an ac voltmeter. Typically, capacitors $C_1 = C_2 = C$ are varied together to obtain null. Alternately, the resistors $R_1 = R_2 = R$ can be varied together with the Cs remaining constant. The Wien bridge method's accuracy depends on the sensitivity of the null detector and the calibration of bridge components. It is best used at audio frequencies.

Fundamental to all frequency and period measurements is the requirement for a very accurate frequency source. In most bench top instruments, the frequency reference is

**FIGURE 8.57**

A Wien bridge used as a frequency sensitive circuit.

obtained from a quartz crystal oscillator operated under conditions of strict temperature control in an oven generally kept to $\pm 0.01^\circ\text{C}$ or better. For example, the Hewlett-Packard HP105B quartz frequency standard has buffered sinusoidal outputs at 5, 1 and 0.1 MHz, and a short term stability measured over at 5 MHz of 5 parts in 10^{12} (measured over a 1 s averaging time). The 'ageing rate' is $< 5 \times 10^{-10}$ per 24 hours for this oscillator.

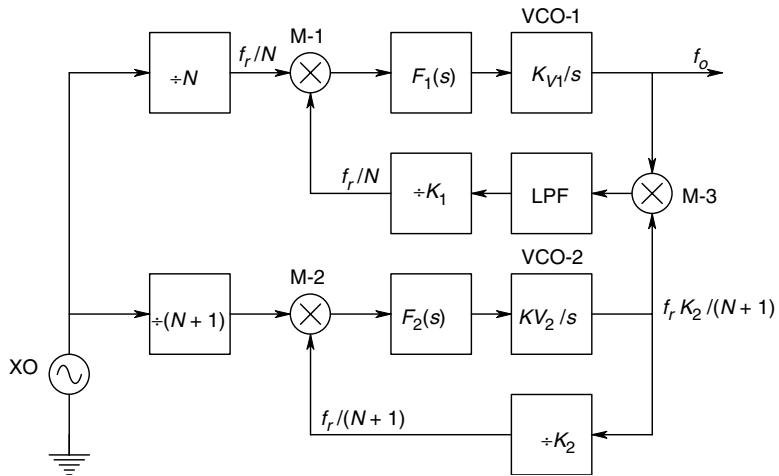
Accurate, variable, reference frequencies can be derived from accurate, fixed frequency oscillators by several means. Direct frequency multiplication uses the generation of harmonics of the reference frequency waveform. The reference waveform is passed through an amplitude comparator to generate a TTL or ECL logic waveform of the same frequency. This logic waveform is then passed through a one-shot multivibrator to create a train of narrow pulses having a Fourier spectrum containing many harmonics. The pulse train is then made the input to a high Q tuned circuit which has its center frequency equal to that of the desired harmonic. It is practical to recover harmonics up to order 9 with this technique. One can also use synchronous binary logic 'rate multipliers', such as the 7497, to reduce the oscillation frequency to f_R , given by the relation,

$$f_R = f_o(M/64) \quad (8.134)$$

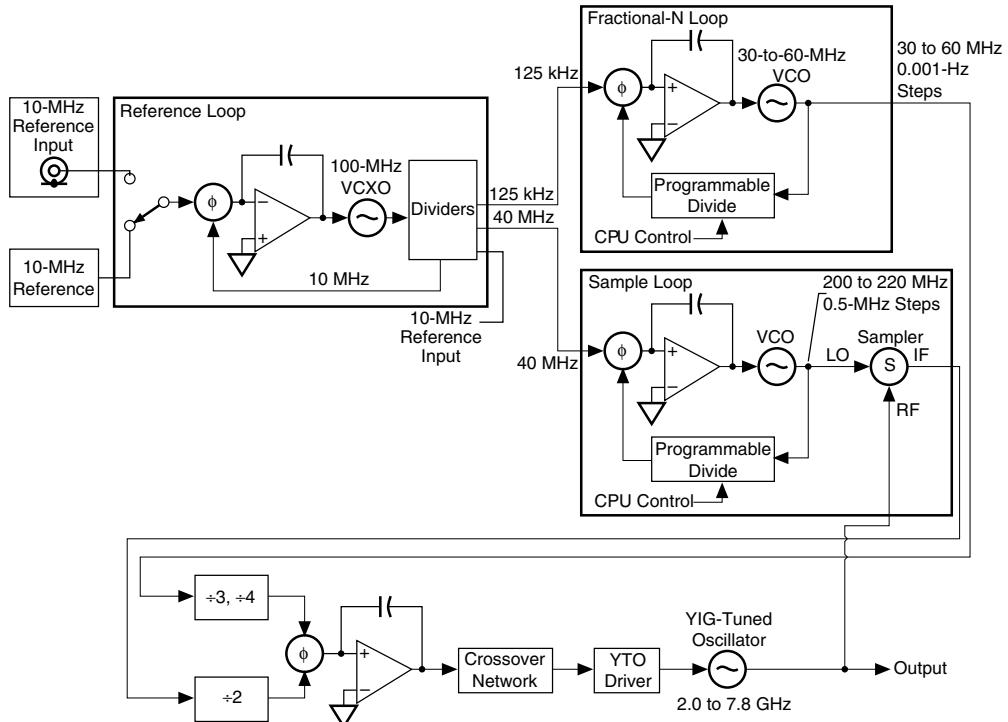
where f_o is the oscillator frequency, M is a six-bit binary number, e.g.,

$$M = D_0(2^0) + D_1(2^1) + D_2(2^2) + D_3(2^3) + D_4(2^4) + D_5(2^5) \quad (8.135)$$

and $D_k = 1$ or 0. More complex frequency synthesis techniques make use of two or more interconnected phase lock loops, mixers, rate multipliers and crystal oscillator sources. Figure 8.58 illustrates an example of a vernier loop synthesizer. A detailed treatment of PLL frequency synthesizers may be found in the texts by Egan (1981) and by Kinley (1980). The Hewlett-Packard Model 8360 series of microwave frequency synthesized sweep generators make use of a single, high accuracy 10 MHz quartz reference source, and uses three PLLs and a feedback scheme to obtain 1 Hz resolution over a range of 2 MHz–40 GHz (HP83642A). The HP 8360 synthesizer architecture is shown in Figure 8.59. The four subsystems in this synthesizer are a 2–7.8 GHz YIG-tuned oscillator (YTO), a reference PLL, a fractional- N PLL, and a sampler loop. A portion of the RF output from the YTO is sent to a sampler, where it is mixed with a high order harmonic of the 200–220 MHz PLL, forming a 20–40 MHz IF signal. A phase detector compares this IF signal to the output of the fractional- N PLL. This phase detector output is summed into the tuning control for the YIG oscillator. The sampler loop can be incremented in 500 kHz steps, and the fractional- N loop can be stepped in 0.001 Hz increments. This mode of operation permits this family of microwave frequency synthesizers to maintain 1 Hz resolution over their entire output range.

**FIGURE 8.58**

A phaselock, vernier loop, frequency synthesizer.

**FIGURE 8.59**

System diagram of a Hewlett-Packard HP8360 frequency synthesizer. (Source: With permission, Hewlett-Packard)

The primary time/frequency standard in current use is the cesium 133 atomic beam clock. A diagram of the cesium clock is shown in Figure 8.60. This standard was discussed in detail in Chapter 2. To summarize the cesium clock's properties, it oscillates at 9.192 631 770 GHz, having an effective Q of 2×10^8 . *Coordinated universal time* (UTC) whose basic unit is the second, is defined as 9,192,631,770 periods of the cesium 133 beam

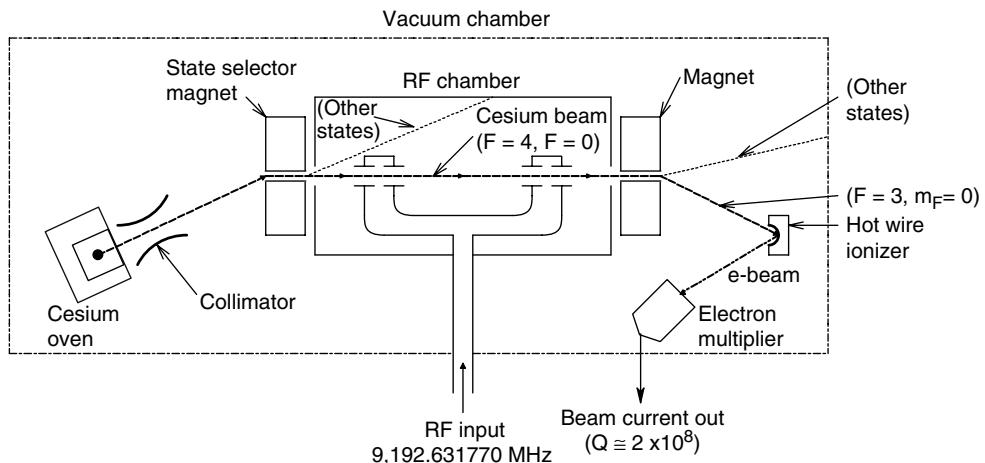
**FIGURE 8.60**

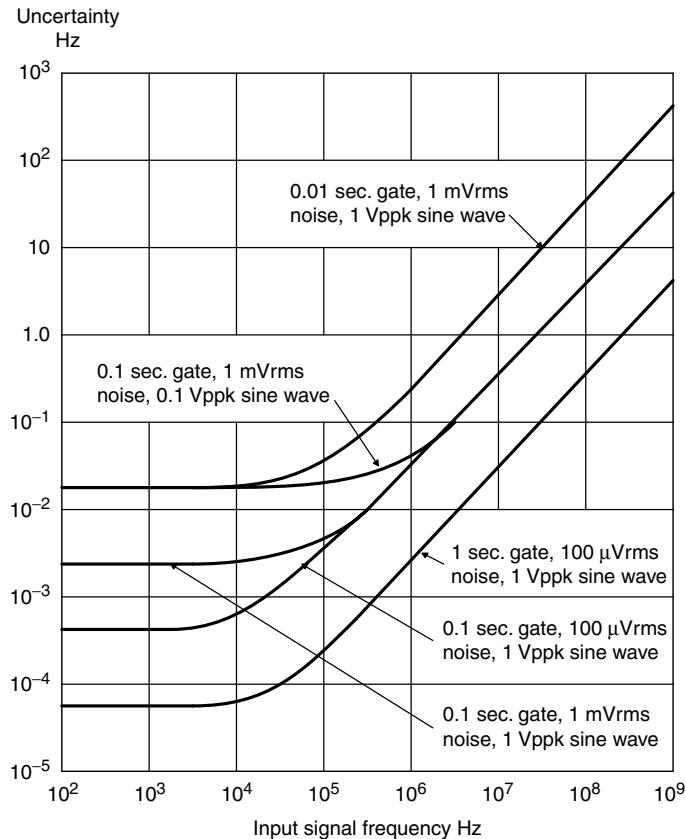
Diagram of a cesium beam frequency standard clock. (Adapted from Oliver and Cage, 1971.)

oscillator. This international standard was adopted in October, 1967. The Hewlett-Packard HP5061B(Opt 004) Cesium Beam Frequency Standard has a long term stability of $\pm 2 \times 10^{-12}$ over the life of the cesium beam tube. Accuracy is also $\pm 2 \times 10^{-12}$. The HP5051B cesium clock has sinusoidal outputs at 10, 5, 1 and 0.1 MHz. HP cesium clocks are used to calibrate and synchronize the SATNAV, Omega LORAN-C and GPS radio navigation systems for land vehicles, boats and aircraft. Cesium beam clocks are ordinarily used to adjust secondary rubidium and quartz oscillators used as secondary standards for frequency or period determination.

Rubidium frequency standards are second in the hierarchy of accuracy. Similar to the operation of a cesium beam clock, the atomic resonance of a rubidium vapor cell is used to synchronize a quartz crystal oscillator in a frequency lock loop. The long term stability of the rubidium vapor oscillator is $\pm 1 \times 10^{-11}/\text{month}$. It, too, has outputs at 5, 1 and 0.1 MHz.

The two principal modes of operation of universal counter timer instruments, such as the HP5345A, are frequency measurement by counting events, such as positive zero crossings of the signal with unknown frequency over the duration of an accurately determined gate window. The duration of the gate window is determined by logic circuits which count a preset number of internal master clock cycles. The gate dwell can be set in the range 1000 s–100 ns in the HP5345A counter. The frequency range on channel A is 0.05 mHz–500 MHz, and periods can be measured from 2 ns to 2×10^4 s. The master quartz clock in the HP5345A counter timer runs at 10 MHz and has a long term stability of less than $\pm 5 \times 10^{-10}/\text{day}$, and a 1 s stability of less than $\pm 1 \times 10^{-11}$. Figure 8.61 illustrates the Hz uncertainty of the HP5345A as a function of applied frequency, input noise and gate time. Note that the measured frequency uncertainty increases with input noise and input frequency, and decreases with gate time. Counter timer instruments can also be operated to measure the time, δT , between positive (or negative) zero crossings of two periodic inputs, thus providing phase information. They can also be run as accurate timers to measure the time between input trigger pulses.

Instantaneous frequency measurements are used to describe the unevenness or frequency noise in a periodic signal. An example of such a noisy frequency source is the human electrocardiogram [Northrop, Wu and Horowitz, 1967]. Many physiological factors control and modulate the rate of the heartbeat. These include blood pCO_2 , emotions, exercise, etc. Other applications of instantaneous frequency measurements have been to

**FIGURE 8.61**

Frequency resolution error of the HP5345A counter-timer instrument. Noise accompanying the input signal and internal uncertainties affect the accuracy of frequency and period measurements. (Source: With permission, Hewlett-Packard)

characterize nerve impulse sequences [Northrop and Horowitz, 1966] and the variations in the rotational speed of a turbine.

The *instantaneous pulse frequency demodulator* (IPFD) was described in Section 7.2.3, and two circuits for IPFDs are shown in Figures 7.17 and 7.18. The j th element of instantaneous frequency, r_j , is defined as the reciprocal of the j th period or inter-event interval. That is:

$$r_i \equiv 1/(t_i - t_{i-1}), j = 2, 3, \dots \quad (8.136)$$

where t_j is the time of occurrence of the j th event and t_{j-1} is the time of occurrence of the next previous event. Any periodic waveform can thus be characterized in the time domain by a number sequence, $\{r_i\}$, which describes the variation of frequency of the waveform on an interval by interval basis. As seen in equation 7.90, the instantaneous frequency number sequence can be converted to a stepwise analog voltage output, given by:

$$q(t) = k \sum_{i=2}^{i=\infty} [1/(t_i - t_{i-1})] [U(t - t_i) - U(t - t_{i+1})] \quad (8.137)$$

where $U(t-a)$ is the unit step function, defined as 0 for $t < a$ and 1 for $t \geq a$.

IPFDs can be of either analog or digital design and generally have found applications in characterizing the irregularities of frequency in audio frequency waveforms.

In summary, we see that there are a variety of methods which can be used to measure or estimate a signal's frequency, which can range from 10^{-3} Hz (mHz) to the hundreds of GHz ($> 10^{11}$ Hz). Frequency measurements can be made by direct counting of events per unit time, or by comparison with an accurate, calibrated oscillator. At very low frequencies, it is more effective to measure the period of the unknown signal, rather than to count its cycles over some fixed interval. Averaging of either period or frequency measurements improves accuracy.

8.10 Resistance, Capacitance and Inductance Measurement

In this section, we will examine and review the various means of measuring resistance, capacitance and inductance. We have already examined null methods of measuring R , C and L at audio frequencies using various bridge circuits in Chapters 4 and 5. These methods will be referred to, but not repeated here. Rather, we will examine various active and passive ohmmeter circuits, means of characterizing linear and nonlinear (voltage variable) capacitances with dc and high frequency ac, and the means of measuring the properties of inductances at high frequencies.

8.10.1 Resistance Measurements

Techniques have been developed to measure resistances from 10^{-7} to over $10^{14}\Omega$. Needless to say, specialized instruments must be used at the extreme ends of this range. We have already seen in Chapter 4 that very accurate resistance measurements are commonly made using dc Wheatstone or Kelvin bridges and a dc null detector such as an electronic nanovoltmeter. The values of the resistances used in the arms of these bridges must, of course, be known very accurately. We shall discuss other dc means of measuring resistance below.

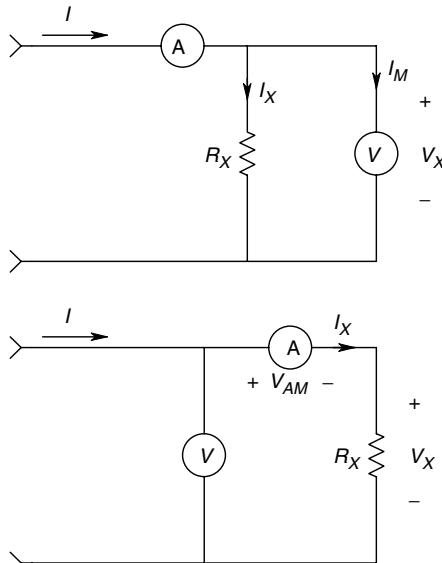
The voltmeter-ammeter method is probably the most basic means of measuring resistance. It makes use of Ohm's law and the assumption that the resistance is linear. As shown in Figure 8.62, there are two basic configurations for this means of measurement—the ammeter being before R_X , which is in parallel with the voltmeter, and the ammeter being in series with R_X after the voltmeter. In the first case, the ammeter measures the current in the voltmeter as well as R_X , while in the second case, the voltmeter measures the voltage drop across the ammeter plus that across R_X . It is easy to show that in the first case, R_X is given by:

$$R_X = \frac{V_X}{I - V_X/R_{VM}} \quad (8.138)$$

where V_X is the voltmeter reading, I is the ammeter reading and R_{VM} is the resistance of the voltmeter (ideally, infinite). In general, $R_{VM} = \eta V_{FS}$, where $\eta = \Omega/V$ sensitivity of the voltmeter and V_{FS} is its full scale voltage range.

In the second case, we find:

$$R_X = V/I_X - R_{AM} \quad (8.139)$$

**FIGURE 8.62**

Two circuits that can be used to measure resistance using a voltmeter and ammeter.

where V is the voltmeter reading, I_X is the ammeter reading (current through R_X) and R_{AM} is the resistance of the ammeter. The first method is best for measuring low values of R_X where $R_{VM} \gg R_X$ and the second method has less error for $R_{AM} \ll R_X$. Obviously, R_X must be found by calculation and the accuracy of the result depends on the accuracy of the meters.

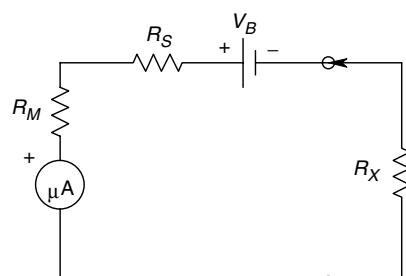
The voltmeter-ammeter method of measuring resistance can be used to measure very low resistances, of the order of $\mu\Omega$. An electronic current source is used to pass a known amount of dc current through the unknown, low resistance. An electronic, dc nanovoltmeter is connected across R_X . Obviously, V_M is directly proportional to R_X .

Common, passive, benchtop volt-ohm-milliammeters (multimeters) use a simple series ohmmeter circuit, shown in its simplest form in Figure 8.63, have a nonlinear (hyperbolic) scale on the D'Arsonval microammeter with zero ohms full scale and infinite ohms at zero deflection. A typical benchtop VOM series ohmmeter will have five R_H ranges (R_H is the indicated R_X at half-scale meter deflection), typically ranging 12Ω - $120\text{k}\Omega$. Inspection of the simple circuit shows that the dc meter current is given by:

$$I_M = \frac{V_B}{(R_M + R_S) + R_X} \quad (8.140)$$

By Ohm's law, the full scale meter current is:

$$I_{M(FS)} = V_B / (R_M + R_S) \quad (8.141)$$

**FIGURE 8.63**

A simple, series ohmmeter circuit.

We note that $(R_M + R_S)$ is the Thevenin resistance that the resistor R_X 'sees'. It is convenient to define $(R_M + R_S)$ as R_H , the half-deflection resistance. That is, when $R_X = R_H$, $I_M = I_{M(FS)}/2$. Further analysis of the simple series ohmmeter is made simpler by defining the ohmmeter's fractional meter deflection, F , as:

$$F = \frac{\theta}{\theta_{FS}} = \frac{I_M}{I_{M(FS)}} = \frac{R_M + R_S}{R_M + R_S + R_X} = \frac{1}{1 + R_X/R_H} \quad (8.142)$$

Here, θ is the meter deflection angle, θ_{FS} is the full scale deflection angle and the R_S s are as shown in Figure 8.63.

To find the most sensitive part of the series ohmmeter's scale, we define the ohmmeter's sensitivity, S as dF/dR_X . This derivative is easily found from equation 8.142 above as:

$$S = \frac{dF}{dR_X} = \frac{R_H}{(R_H + R_X)^2} \quad (8.143)$$

S clearly has a maximum with respect to R_H . If we find dS/dR_H and set it equal to zero, the peak in the sensitivity is seen to occur when $R_H = R_X$. That is, the center of the series ohmmeter's scale is the most sensitive (and most accurate) part.

Since the voltage, V_B , of the series ohmmeter's battery drops as the battery is used and ages, a variable resistor, R_V , is added in parallel with the microammeter to compensate for the drop in V_B . This practical series ohmmeter circuit is shown in Figure 8.64. R_V is adjusted with the meter leads shorted ($R_X = 0$), so $I_M = I_{MFS}$. In this circuit, R_H is still the Thevenin resistance that R_X sees, which is:

$$R_H = R_S + \frac{R_M R_V}{R_M + R_V} \quad (8.144)$$

An increase of R_V to compensate for a drop in V_B will thus change R_H from its design value and cause a small meter calibration error.

Shunt ohmmeters are less frequently encountered than are series ohmmeters. Shunt ohmmeters are used to measure low resistances and have R_{HS} s ranging 0.5–50 Ω . They are primarily used to measure resistances associated with coils, such as dc motor armature windings. The circuit of a shunt ohmmeter is shown in Figure 8.65. In this circuit, the battery must supply a substantial current, often in the ampere range. Consequently, a robust battery, such as a lead acid motorcycle type battery, is used. As in the case of the series ohmmeter, we can define the half deflection resistance, R_H , as the Thevenin resistance seen by R_X :

$$R_H = \frac{R_M R_S}{R_M + R_S} \quad (8.145)$$

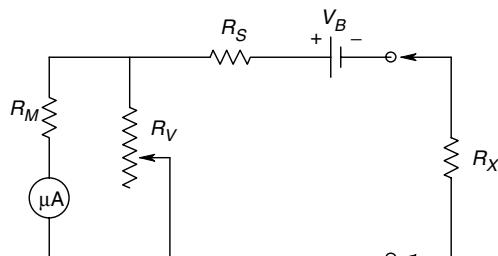
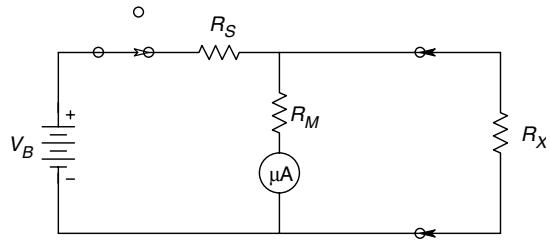


FIGURE 8.64

A practical series ohmmeter circuit; R_V is used to set the ohmmeter zero in compensation for changes in the value of V_B .

**FIGURE 8.65**

Circuit of a simple, shunt ohmmeter.

Full scale meter current occurs when $R_X = \infty$ and is simply:

$$I_{M(FS)} = V_B / (R_M + R_S) \quad (8.146)$$

In general, the meter current is given by:

$$I_M = \frac{V_B R_X / (R_X + R_S)}{R_M + R_X R_S / (R_X + R_S)} = \frac{V_B R_X / (R_M + R_S)}{R_X + R_M R_S / (R_M + R_S)} \quad (8.147)$$

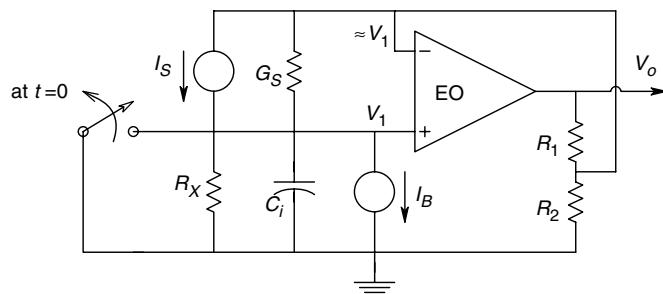
Zero meter deflection occurs for $R_X = 0$ (short circuited input). The fractional deflection factor of the shunt ohmmeter is found to be:

$$F = \frac{\theta}{\theta_{FS}} = \frac{I_M}{I_{M(FS)}} = \frac{R_X}{R_X + R_H} \quad (8.148)$$

It can also be shown that the center of the shunt ohmmeter's scale is also the point of greatest sensitivity.

Both series and shunt ohmmeters have relatively low accuracy, generally 2–5% at center scale, which is adequate for most non-critical applications, such as measuring a motor coil's resistance, checking circuit continuity, or verifying resistor values (for those persons who do not know the resistor color code).

We next examine several *electronic ohmmeter circuits*. The readout of such circuits can be either analog or digital. In general, their accuracy is an order of magnitude better than that obtained with a series or shunt ohmmeter. Electronic ohmmeters can measure resistances ranging $0.1 \mu\Omega$ – $10^{18} \Omega$. Obviously, special techniques and instruments must be used to measure resistances at the extreme ends of the range cited above. The first electronic circuit we shall consider is the *Normal Mode ohmmeter*, shown in Figure 8.66. The op-amp can be an IC, electrometer type, having a bias current on the order of 40 fA. A calibrated current source, easily made from a battery and a large resistor, injects a dc

**FIGURE 8.66**

Schematic of a 'normal mode' electronic ohmmeter. I_S and G_S represent the Norton equivalent of a floating current source.

current, I_S , into the unknown resistance, R_X . The voltage at the op-amp's non-inverting input is given by KCL:

$$V_1(G_X + sC_i) + I_B = I_S \quad (8.149)$$

or

$$V_1 = \frac{(I_S - I_B)R_X}{(1 + sC_iR_X)} \quad (8.150)$$

It is easy to see that the op-amp's output, $V_o = V_1 (1 + R_1/R_2)$, so:

$$V_o = (1 + R_1/R_2) \frac{(I_S - I_B)R_X}{(1 + sC_iR_X)} \quad (8.151)$$

If the switch is opened at $t=0$, the input may be considered to be $(I_S - I_B)/s$, and $v_o(t)$ will rise exponentially with time constant C_iR_X to a steady state value given by:

$$v_{o(SS)} = (1 + R_1/R_2)(I_S - I_B)R_X \text{ V} \quad (8.152)$$

Thus, the op-amp output voltage is directly proportional to R_X and a linear ohms scale can be used on an analog output meter. Two problems arise with this circuit when measuring $R_X > 10^{10} \Omega$. The bias current must be kept much less than I_S and if C_i is of the order of several hundred pF, the ohmmeter's response time constant can become significantly long (tens or hundreds of seconds), requiring excessive time to obtain a steady state reading. If we assume that full scale deflection of the normal mode electronic ohmmeter is $v_{o(SS)} = 1 \text{ V}$, then full scale resistances from $1-10^{10} \Omega$ can be measured with I_S values ranging $10 \text{ mA}-1 \text{ pA}$, respectively, with $(1 + R_1/R_2) = 100$.

To measure resistors in excess of $10^{10} \Omega$, the *Fast Mode electronic ohmmeter configuration*, shown in Figure 8.67, is generally used. Now the unknown resistor is placed in the feedback loop of the op-amp. One end of R_X is at virtual ground at the summing junction and the other end is connected to the V_2 node. The voltage V_2 is clearly equal to $-(I_S + I_B)R_X$. By KCL, we may write:

$$V_2(G_2 + G_1) - V_O G_1 - I_S = 0 \quad (8.153a)$$

$$-(I_S + I_B)R_X(G_2 + G_1) - V_o G_1 - I_S = 0 \quad (8.153b)$$

$$-I_S[R_X(G_2 + G_1 + 1)] - I_B R_X(G_2 + G_1) = V_o G_1, \quad (8.153c)$$

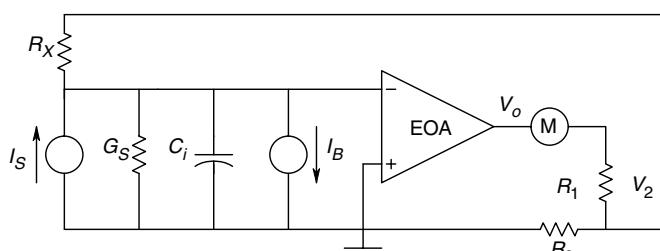


FIGURE 8.67

Schematic of a 'fast mode', electronic ohmmeter. R_X is placed in the op-amp's feedback loop. The summing junction is at virtual ground.

Assuming $R_X \gg R_1, R_2$, we can solve for the output voltage:

$$V_o = -R_X I_S (1 + R_1/R_2) - R_X I_B (1 + R_1/R_2) \quad (8.154)$$

The second term is negligible as long as $I_S \gg I_B$. If the output current is to be read by the microammeter, M, we note that:

$$I_o = (V_o - V_2)/R_1 = -I_o R_X (G_2 + G_1) + I_S R_X G_1 = -I_S R_X G_2 \quad (8.155)$$

For example, if $R_X = 10^{14} \Omega$, $I_S = 1 \text{ pA}$ and $R_2 = 10^4 \Omega$, then $I_o = -10 \text{ mA}$.

High megohm resistors ($1 \text{ G}\Omega$ or higher) are central to the operation of many electrometer instruments, including feedback picoammeters, as described in Section 8.4.2 of this text. Such resistors are typically made from metal oxides and are enclosed in sealed, glass tubes. The *voltage coefficient* of this type of resistor, defined as $(\Delta R/R)/V$, is typically less than 5 ppm up to 100 V applied voltage, and the tempcos of high megohm resistors are of the order of 100 ppm/ $^\circ\text{C}$ for $10^8 \Omega$, and 1000 ppm/ $^\circ\text{C}$ for $10^{11} \Omega$. High megohm resistors are delicate and must be protected from mechanical shock and moisture. Their glass envelopes must be kept free of dirt and fingerprint oils and salt by cleaning with pure alcohol after handling.

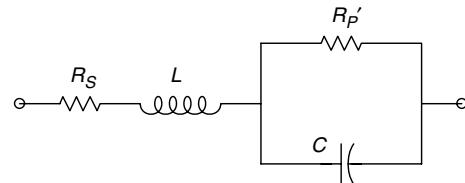
Measurement of high megaohm resistors with fast mode electronic ohmmeters is not as simple as just connecting R_X to the meter and taking the reading. The reading can be influenced by static and slowly changing electrostatic fields. Even on humid days, most persons have a dc electric field associated with their bodies, even when cotton clothing is worn. Such fields can influence the measurement of R_X , destroying accuracy. To uncouple R_X from dc electrostatic influences and to prevent pickup of power line **E** and **B** fields by the measurement apparatus, R_X should be thoroughly shielded, both electrostatically and magnetically. (The 60 Hz fields can induce voltages that may saturate the amplifier, giving incorrect dc readings of resistance.)

8.10.2 Capacitance Measurements

We have seen that accurate capacitance measurements can be made at audio frequencies (generally 1 kHz) using a variety of bridge circuits, described in Section 5.5.1 of this text. It is often important to characterize capacitors at radio frequencies and at ultra low frequencies, the latter range being important in the characterization of capacitor dielectrics, as well as dielectrics used as electrical insulation. There are also voltage variable capacitances associated with semiconductor *pn* junctions, which it important to measure. These include the capacitance of varactor diodes (reverse biased *pn* junction diodes), the gate-source capacitance (C_{gs}), gate-to drain capacitance (C_{gd}), drain-source capacitance (C_{ds}), C_{iss} and C_{rss} of FETs, and the C_π and C_μ of BJTs. We will describe below some of the non-bridge means to measure capacitors and to characterize their equivalent circuits.

8.10.2.1 The Use of Q-Meter for Capacitance Measurement

At radio frequencies, the equivalent circuit for a capacitor will not only include a resistance to account for dielectric power losses, but also losses due to skin effect in the capacitor's leads. Equivalent inductances may also appear at VHF, UHF and SHFs due to lead geometry and physical layout of the plates and dielectric. Figure 8.68 illustrates one such equivalent circuit. A practical instrument for measuring the net capacitance of a

**FIGURE 8.68**

High frequency, lumped parameter equivalent circuit of a practical capacitor.

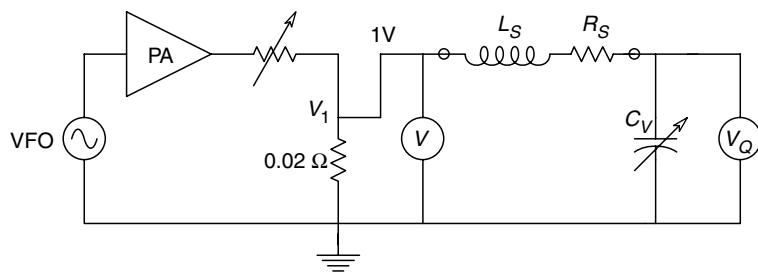
capacitor at a radio frequency is the *Q-meter*. (The Q-meter can also be used to measure inductor *Q* and inductance, as described in the next section.) There are several commercial Q-meters:

1. The Boonton type 260A covers the frequency range 50 kHz–50 MHz.
2. The Boonton type 190A Q-meter covers the range 20–260 MHz
3. The Marconi type TF-1245 is used with separate voltage sources covering the range 40 kHz–50 MHz (type TF-1246), and 20–300 MHz (type TF-1247)
4. The Hewlett-Packard Type HP 4342A Q-meter operates in the range 22 kHz–70 MHz

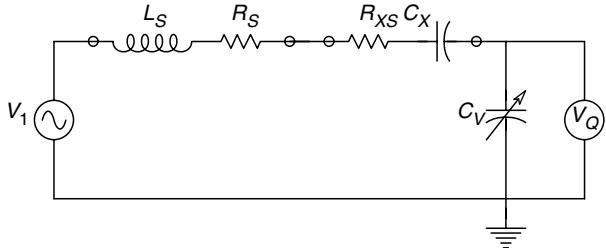
The Q-meter is basically a very simple circuit, consisting of a low impedance RF voltage source, a variable capacitor and an RF voltmeter. The Q-meter is used with standard coils and capacitors in many of its measurement modes. A schematic of a Q-meter is shown in Figure 8.69. A variable frequency RF current is passed through a $0.02\ \Omega$, non-inductive resistor. The RF voltage across this resistor appears to be from a nearly ideal RF voltage source, V_1 .

Measurement of an unknown capacitance, C_X , can be done in several ways. In the first method, C_X is placed across the internal capacitance, C_V , and a standard series inductor of appropriate value is placed in series. The circuit is resonated at f_o and the capacitance of the variable capacitor is recorded as C_1 . Next, C_X is removed and the variable capacitor is again tuned to C_2 to resonate the circuit. It is easy to see that $C_X = C_2 - C_1$. This technique works as long as $C_X < \text{max}C_V$. If C_X is larger than $\text{max}C_V$, then a known, standard capacitor, C_S , can be placed in series with C_X to make the series combination less than $\text{max}C_V$. Once the capacitance of the series combination of C_X and C_S is known, C_X can be calculated.

Cooper (1978) gave a more general method for measuring C_X and its equivalent series resistance, R_{XS} , on a Q-meter. Refer to Figure 8.70, in which the unknown capacitor is

**FIGURE 8.69**

Circuit of basic Q-meter. Note: VFO = variable radio frequency (RF) oscillator, PA = power amplifier. V and V_Q = high frequency ac voltmeters.

**FIGURE 8.70**

Series Q-meter circuit used to measure an unknown capacitance at R_{fs} .

placed in series with the standard inductor having inductance, L_S , and series resistance, R_S . As in the parallel technique described above, two measurements must be made. First, the unknown capacitor is shorted out and the Q-meter is resonated with C_V at f_0 . The Q and C_V for this case are recorded as Q_1 and C_1 . Next, the short is removed and resonance is again observed, giving Q_2 and C_2 . Now, in the first case, $X_C = X_L$, or:

$$\frac{1}{2\pi f_0 C_1} = 2\pi f_0 L_S \quad (8.156)$$

and

$$Q_1 = \frac{2\pi f_0 L_S}{R_S} = \frac{1}{2\pi f_0 C_1 R_S} \quad (8.157)$$

For the second measurement, it can be shown that the reactance of the unknown capacitor is given by:

$$X_X = \frac{1}{2\pi f_0 C_X} = \frac{C_2 - C_1}{2\pi f_0 C_2 C_1} \quad (8.158)$$

Hence

$$C_X = \frac{C_1 C_2}{C_2 - C_1} \quad (8.159)$$

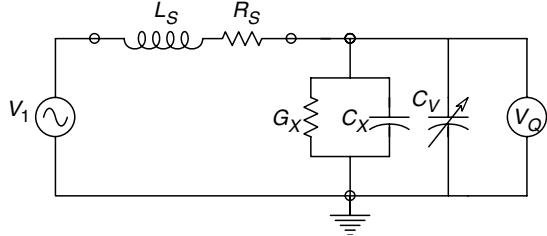
The series resistive component of the unknown capacitance can be shown to be given by [Cooper, 1978]:

$$R_{XS} = \frac{C_1 Q_1 - C_2 Q_2}{2\pi f_0 C_1 C_2 Q_1 Q_2} \quad (8.160)$$

The series dissipation factor of the unknown capacitor, $D_{SX} \equiv \omega C_X R_{XS}$, can be written from the above relations:

$$D_{SX} = \frac{C_1 Q_1 - C_2 Q_2}{Q_1 Q_2 (C_2 - C_1)} \quad (8.161)$$

It is also possible to place the unknown capacitor (impedance) in parallel with C_V , as shown in Figure 8.71. This configuration is used when $X_X > X_{C1}$. As in the series

**FIGURE 8.71**

Parallel connection of the Q-meter used to measure an unknown capacitance.

case treated above, we first remove the unknown capacitor and resonate the Q-meter by adjusting C_V . At resonance, $X_L = X_{C1}$, thus:

$$C_1 = 1/(\omega^2 L_S) \quad (8.162)$$

and

$$Q_1 = \omega L_S / R_S = 1/(\omega C_1 R_S) \quad (8.163)$$

Next, the unknown capacitor is placed in parallel with C_V which is adjusted to again obtain resonance. At resonance,

$$X_L = \omega L_S = \frac{X_{C2} X_X}{X_{C2} - X_X} = \frac{1}{\omega C_1} \quad (8.164)$$

Equation 8.164 can be manipulated to yield:

$$X_X = \frac{1}{\omega(C_1 - C_2)} \quad (8.165)$$

from which it is easy to write

$$C_X = C_1 - C_2 \quad (8.166)$$

In addition, Cooper (1978) shows that the equivalent shunt conductance of the unknown capacitance can be written as:

$$G_X = \omega C_1 \frac{Q_1 - Q_2}{Q_1 Q_2} S \quad (8.167)$$

and the dissipation factor of the unknown capacitor can be calculated from

$$D_{XP} = \frac{1}{\omega C_X R_X} = \frac{C_1(Q_1 - Q_2)}{Q_1 Q_2(C_1 - C_2)} = \frac{C_1}{Q_1/Q_2} \frac{\Delta Q}{\Delta C} \quad (8.168)$$

The Q-meter is seen to be a versatile instrument for measuring the properties of unknown capacitors at radio frequencies. However, the accuracy is modest (1–2%) and calculations are required to obtain the equivalent circuit parameters.

8.10.2.2 Capacitance Measurement by Q/V

Capacitance is defined by the basic relation, $C \equiv Q/V$. Hence if one applies a dc voltage, V_C , to an initially uncharged capacitor and then measures the total charge, Q , accumulated in the steady state, the ratio of Q to V_C is, by definition, the capacitance in Farads. This basic definition is used in the design of several commercial instruments used to measure capacitors. The Keithley Model 595 Quasistatic CV Digital Capacitance Meter uses a novel, stepwise programming of the voltage applied to the capacitor. A sequence of N steps of voltage, each step of height ΔV , where ΔV is selectable to be 0.01, 0.02, 0.05 or 0.10 V, can be applied to C_X . The maximum range of voltage is ± 20 V. As each step of voltage is applied to the unknown capacitor, an electrometer charge amplifier circuit integrates the current flowing into the capacitor as a result of each voltage step. The Model 595 meter samples the voltage proportional to the charge at the integrator output just before each voltage step is applied, and it samples the charge amplifier's output twice at an interval ΔT after its output reaches steady state from charging C_X . This sampling is shown in Figure 8.72. The op-amp's dc bias current may thus be estimated from:

$$I_B = (Q_3 - Q_2)/\Delta T \quad (8.169)$$

This bias current must be subtracted from the current charging the capacitor under measurement. The charge amplifier's output can be shown to be:

$$v_o(t) = -\frac{I_B}{C_F} - \frac{\Delta V C_X}{C_F} \{1 - \exp[-(t - t_1)/RC_X]\} \quad (8.170)$$

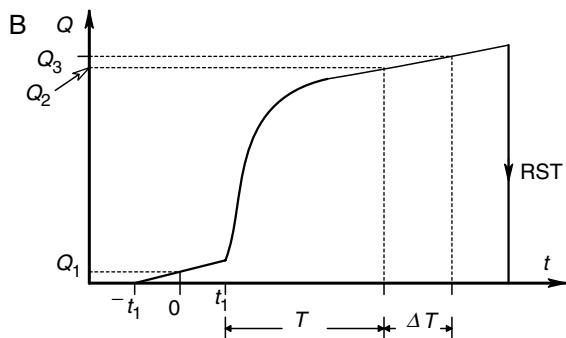
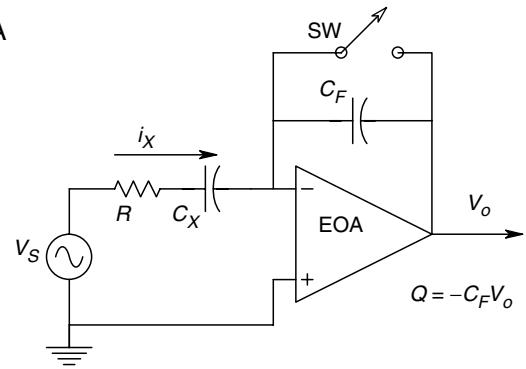


FIGURE 8.72

A. Electrometer op-amp charge amplifier circuit used to measure voltage variable capacitances by the $\Delta Q/\Delta V$ method. B. Charge amplifier output waveform in response to one input voltage step, showing sampling points used in the Keithley model 595 CV capacitance meter.

Referring again to Figure 8.72 and equation 8.170, we see that in the steady state, $C_X(V)$ is given by:

$$C_X(V) = \frac{(Q_3 - Q_1) - I_B(T + t_1)}{\Delta V} \quad (8.171)$$

C_X can also be expressed in terms of the charge amplifier's steady state output voltage:

$$C_X(V) = -\frac{C_F}{\Delta V} v_{o(ss)}(T + t_1) - \frac{I_B(T + t_1)}{\Delta V} \quad (8.172)$$

The V reported by the Keithley 595 CV meter is $(V_{LAST} + 0.5 \Delta V)$. If the capacitor is linear, then $C_X(V)$ is constant over the range of V . If the capacitance is a function of voltage, such as the capacitance of a reverse biased pn junction, then $C_X(V)$ can be plotted vs V with the Keithley 595 instrument in order to characterize the capacitor's voltage dependency. The model 595 CV meter can also be used to measure the junction capacitances of BJTs and FETs, as well as MOS chip capacitors. In spite of its digital display, sampling, microprocessor control and IEEE-488 bus, the 595 is basically an analog instrument and has 1% accuracy.

Another application of the $C=Q/V$ method to characterize capacitor (and other) dielectrics has been described by Mopsik (1984). The procedure has been called *time domain spectroscopy*, or TDS. The basic circuit for the *Mopsik system* is shown in Figure 8.73. An electrometer charge amplifier (integrator) is connected to a node between the capacitor and dielectric under investigation and a low loss, reference capacitor, C_R . C_R is variable and has a dry air dielectric. The input to C_X is a positive step voltage, V_C , from an op-amp. A voltage of $-V_C$ is applied simultaneously from another op-amp to C_R . The purpose of C_R is to permit a charging current equal and opposite to that of C_X to be summed at the charge amplifier's summing junction.

In analyzing the Mopsik system, we will consider two cases. The first, where $C_X = C_R$ (i.e., there is perfect compensation) and the second where $C_X \neq C_R$. The second case is more realistic, as perfect matching of C_X and C_R will not ordinarily be obtained in practice.

In the first case, there will be three components to the system output. One is due to the dc 'leakage current' through the equivalent conductance, G_X , shunting C_X . G_X describes the behavior of the C_X 's dielectric; G_X may have a purely ohmic component, as well

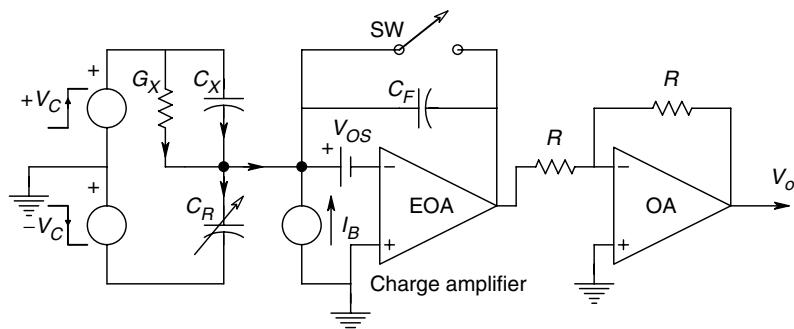
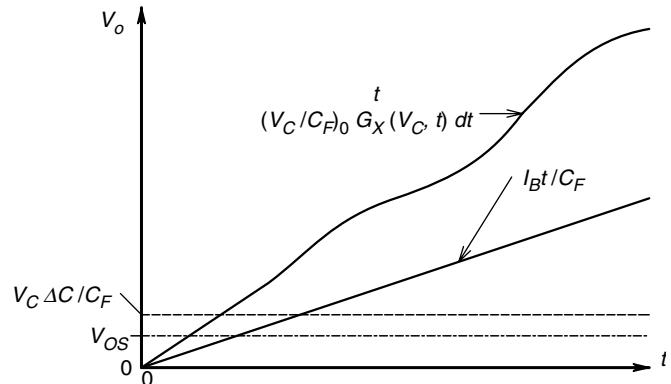


FIGURE 8.73

The Mopsik (1984) circuit for characterization of capacitances in the frequency domain.

**FIGURE 8.74**

Plot of the electrometer charge amplifier's output in the Mopsik system. The three components of equation 8.173 are shown.

as time and voltage dependent nonlinear components. Such behavior may arise from the presence of 'free' electrons and holes in the dielectric, as well as mobile positive and negative ions, and impurity centers capable of 'trapping' charged particles and ions until sufficient energy releases them. The other two components at the output are due to the op-amp's bias current, I_B , and offset voltage, V_{OS} . If we assume that C_F has zero charge at $t=0$ when the steps of $\pm V_C$ are applied to the C_X and C_R , then we may write in the time domain as:

$$v_o(t) = \int_0^t \frac{V_C}{C_F} G_X(V_C, t) dt + \frac{I_B}{C_F} t + V_{OS} U(t) \quad (8.173)$$

A plot of these components is shown in Figure 8.74. Note that in the case of a good low conductivity dielectric, I_B may exceed $V_C G_X$ in magnitude, making the bias current error a significant component at the output. However, V_{OS} can generally be nulled out at a particular operating temperature.

In the second case, $\Delta C = C_X - C_R \neq 0$, so there will also be a capacitive charging component to the output of the charge amplifier. If we assume that the capacitors charge instantaneously, then the charge amplifier output will be:

$$v_o(t) = \int_0^t \frac{V_C}{C_F} G_X(V_C, t) dt + \frac{I_B}{C_F} t + \left(V_{OS} + \frac{V_C \Delta C}{C_F} \right) U(t) \quad (8.174)$$

Note that the ΔC term can be positive or negative.

In the steady state, for a linear capacitor, we have the well-known relation, $C = Q/V_C$. Here, Q is the total charge accumulated in an ideal capacitor in the steady state. Mopsik noted that the charge, Q , does not really reach a steady state, but continues to increase after the pure capacitance component, C_X , has charged to the applied dc voltage, V_C . The continuing increase in Q is due to dielectric phenomena. Mopsik defined a 'time variable capacitance' as:

$$C(t) = Q(t)/V_C = v_o(t)C_F/V_C \quad (8.175)$$

In order to characterize the so-called, time variable capacitor given by equation 8.175 in the frequency domain, Mopsik takes the Fourier transform of the $v_o(t)$ transient, given by equation 8.174, times C_F/V_C . This Fourier transform can easily be shown to be:

$$C^*(j\omega) = \frac{G_X(j\omega, V_C)}{j\omega} + \frac{I_B}{V_C(j\omega)^2} + \left[\frac{V_{OS}C_F}{V_C} + \Delta C \right] \frac{1}{j\omega} \quad (8.176)$$

Note that both $C^*(j\omega)$ and $G_X(j\omega, V_C)$ are complex (i.e., they have real and imaginary components).

Mopsik defined a $C^*(j\omega)$ by equation 8.177:

$$C^*(j\omega) = C'(\omega) - jC''(\omega) \quad (8.177)$$

At a given frequency, we can show that the $C^*(j\omega)$ model for the unknown capacitor is entirely equivalent to the standard, parallel R-C model. Assume an ac voltage, V_C , is impressed across C^* , and the parallel model. Thus the current is:

$$I^* = V_C j\omega C^*(j\omega) = V_C [j\omega C'(\omega) + \omega C(\omega)] = I_P = V_C [G_P + j\omega C_P] \quad (8.178)$$

By comparing terms, we find that:

$$C_P = C'(\omega) \quad (8.179a)$$

$$G_P = \omega C''(\omega) \quad (8.179b)$$

Referring to Figure 8.75, we see that the loss angle tangent, $\tan(\delta)$, at frequency ω must be the same for both capacitor models.

$$\tan(\delta^*) = \frac{C''(\omega)}{C'(\omega)} = \tan(\delta) = \frac{1}{\omega R_P C_P} \quad (8.180)$$

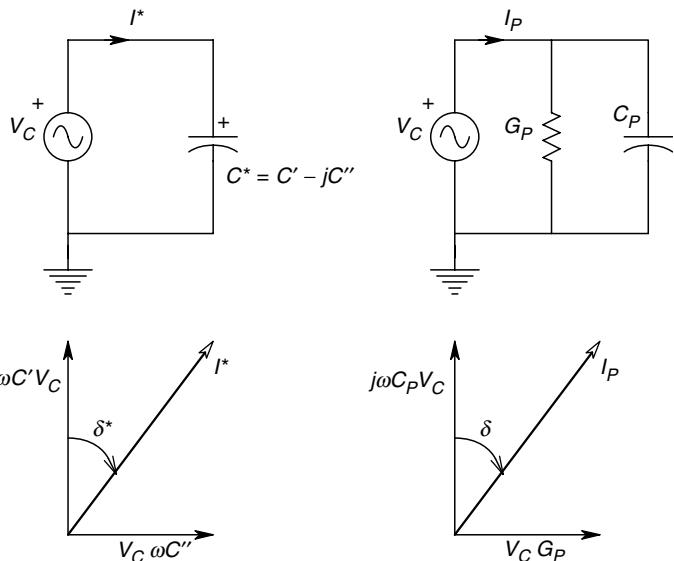


FIGURE 8.75

Equivalence of the loss tangent for the familiar, parallel R-C capacitor model, and the complex capacitor model of Mopsik ($I^* = I_P$).

$\tan(\delta)$ can also be expressed in terms of complex permittivity, ε^* . Here, we assume that:

$$C^*(j\omega) = \frac{A\varepsilon^*(\omega)}{d} = \frac{A}{d}[\varepsilon(\omega) - j\varepsilon''(\omega)] \quad (8.181)$$

Thus,

$$\tan(\delta) = \frac{\varepsilon''(\omega)}{\varepsilon'(\omega)} \quad (8.182)$$

The Mopsik system calculates $\tan[\delta(\omega)]$ for the unknown capacitor and dielectric from the ratio of $C''(\omega)$ to $C'(\omega)$. $C'(\omega)$ is the real part of $C^*(j\omega)$, which is found by taking the Fourier transform of $v_o(t)C_F/V_C$ as written in equation 8.181 above. $C'(\omega)$ can be expressed as:

$$C'(\omega) = RE\{C^*(j\omega)\} = [IM\{G_X(\omega, V_C)\} - I_B/(V_C\omega)]/\omega \quad (8.183)$$

The imaginary part of $C^*(j\omega)$ is:

$$C''(\omega) = IM\{C^*(j\omega)\} = [RE\{G_X(j\omega, V_C)\} + V_{OS} C_F/V_C + \Delta C]/\omega \quad (8.184)$$

Equations 8.183 and 8.184 above illustrate the source of artifacts in the Mopsik system. Clearly, I_B can have a major effect on the $\tan[\delta(\omega)]$ result, as can the offset voltage and a non-zero difference between C_R and C_X . When we can assume that $C_R = C_X$ and V_{OS} and $I_B = 0$, then the loss tangent is:

$$\tan[\delta(\omega)] = \frac{RE\{G_X(j\omega, V_C)\}}{IM\{G_X(j\omega, V_C)\}} \quad (8.185)$$

Clearly, the Mopsik approach to characterizing a capacitor and its dielectric is unique. Since the unknown capacitance is assumed to be proportional to the charge, $q(t)$, it too, will be a function of time and can be Fourier transformed. Hence, we have a complex capacitance which is a function of frequency and which will have a real and an imaginary part. It is doubtful in practice that the $\tan[\delta(\omega)]$ calculated by the Mopsik method is independent of the electrometer's I_B . Such independence would be possible only if the current, $V_C G_X \gg I_B$. A typical I_B for a solid state, electrometer op-amp is about 30 fA. Also, the initial step transient artifacts due to V_{OS} and ΔC will contribute error to the high frequency portion of $C^*(j\omega)$.

8.10.3 Inductance Measurements

Measurement of inductances can be made from subsonic to ultra high frequencies by a variety of means. Meaningful measurements can be made over the range of nano-henries to kilohenries, but not at the same frequency. Circuits used to characterize inductances were introduced in Section 5.2, and audio frequency inductance measuring bridges were described in Section 5.5.2. Audio frequency bridges are the most accurate means of measuring inductance at audio frequencies. In this section, we will describe two other means of measuring inductance—the ac voltmeter method, and the Q-meter.

Inductances can also be measured with vector impedance meters, described in Section 8.10 below.

8.10.3.1 Voltmeter Method of Estimating Inductance

At power line frequencies, it is possible to find the inductance of a coil by first measuring the coil's dc resistance, R'_L , with an ohmmeter. The actual real part of the series impedance of the coil at the measurement frequency will generally be slightly higher than R'_L due to magnetic core losses and skin effect at radio frequencies. The effective, ac, series coil resistance is defined to be R_L . A simple series circuit with the coil and a known resistor, R , is used (Figure 8.76A). The only instrument needed is a high input impedance ac voltmeter. The excitation voltage (V_S), the voltage across the coil (V_L) and the voltage across the resistor (V_R) are measured. By Kirchoff's voltage law, we know the vector sum of V_L and V_R must equal V_S . Also, we know that the phase angle of the current in a series inductive circuit lags the phase of the voltage across the circuit. We can thus draw the voltage vector (phasor) diagram as shown in Figure 8.76B. The three known voltages form the three sides of a triangle. The angle of V_R , $-\theta$, is also the angle of the current in the circuit, I . The angle of V_S is assumed to be zero. By Ohm's law, the magnitude of the current is:

$$I = V_R/R \quad (8.186)$$

The magnitude of the voltage across the coil can be written as:

$$V_L = |IZ_L| = \frac{V_R}{R} \sqrt{\omega^2 L^2 + R_L^2} \quad (8.187)$$

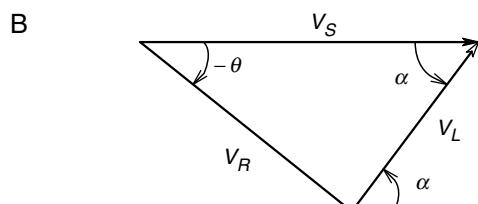
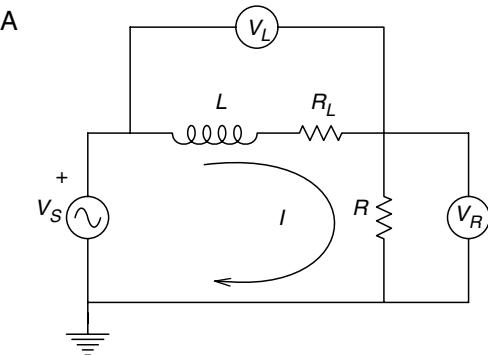


FIGURE 8.76

A. Circuit of voltmeter method of estimating inductance at audio and power line frequencies. B. Vector (phasor) voltages in the circuit of Figure 8.76A.

Equation 8.187 can be solved for L , since we know ω , R_L , R , V_R and V_L :

$$L = (1/\omega) \sqrt{R^2 V_L^2 / V_R^2 - R_L^2} \quad (8.188)$$

L can also be solved for in terms of the current angle, $-\theta$, ω , R and R_L , and also in terms of V_S , V_R , ω , R and R_L . The derivation of these expressions are left as exercises at the end of the chapter.

8.10.3.2 The Use of Q-meter for Inductance and Q Measurement

We have already seen in Section 8.10.2.1 how a Q-meter can be used to measure an unknown capacitor at radio frequencies. Measurement of an unknown series inductance and its Q are quite simple. Referring to the basic Q-meter schematic shown in Figure 8.69, we replace the standard inductor with a short circuit—the unknown inductor goes in the Z_S position, and Z_P is made infinite. The frequency of V_1 is set to the desired value and V_1 is adjusted to be 1.0 V. Next, C_V is varied to obtain series resonance in the circuit. At resonance:

$$\frac{V_Q}{V_1} = \frac{V_C}{V_1} = \frac{IX_C}{IR_L} = \frac{IX_L}{IR_L} = \frac{\omega L}{R_L} = Q_L \quad (8.189)$$

Hence, V_Q in volts equals the inductor's Q . This simple analysis assumes that the range of C_V is adequate to resonate with L (i.e., $f=1/(2\pi\sqrt{LC_V})$). If resonance is not possible within the working range of C_V , then, depending on whether we are dealing with an inductor at the lower end of the range (90 nH) or one at the upper end of the range (130 mH), we can add a known, standard series inductance in series with the unknown inductor, or a known, standard capacitor in parallel with C_V , respectively. The Boonton Model 260A Q-meter allows measurement of inductances $>5\ \mu\text{H}$ to $\pm 3\%$, and Q_s to $\pm 5\%$.

It should be stressed that the Q-meter measures the effective inductance and the Q is determined as the ratio of the effective inductive reactance to the effective series real part of the loop impedance. At HF and VHF, the effective parallel capacitance of the coil acts to reduce the effective Q (equation 5.3).

8.11 Vector Impedance Meters

A linear, two terminal, passive, electrical component may be characterized at a given frequency by its vector impedance, or admittance. Impedance, \mathbf{Z} , and admittance, \mathbf{Y} , are defined in rectangular vector form by:

$$\frac{\mathbf{V}}{\mathbf{I}} = \mathbf{Z} = R + jX = \frac{1}{\mathbf{Y}} = \frac{1}{G + jB} \quad (8.190)$$

It is understood that \mathbf{V} is the sinusoidal voltage across the two port, and \mathbf{I} is the vector current through it. Impedance can be measured at a given frequency by computing the vector ratio:

$$\mathbf{Z} = Z \angle \theta = \frac{V \angle \theta_1}{I \angle \theta_2} = Z \angle (\theta_1 - \theta_2) \quad (8.191)$$

If the magnitude of \mathbf{I} is held constant, then measurement of \mathbf{V} and the phase difference between \mathbf{V} and \mathbf{I} will characterize \mathbf{Z} at a given frequency. Figure 8.77 shows the block

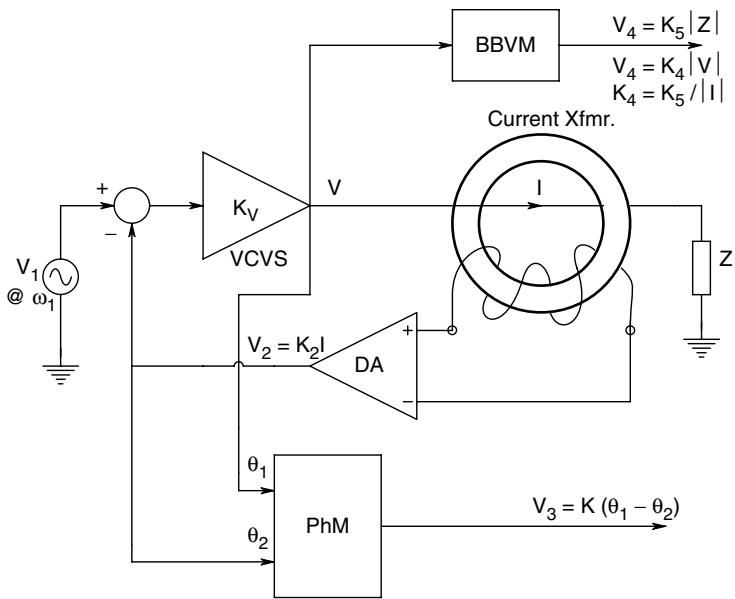
**FIGURE 8.77**

Diagram of a vector impedance meter. Note: PhM = phase meter, BBVM = broadband (RF) voltmeter (e.g., a thermocouple type), OSC = oscillator, VCVS = voltage controlled voltage source.

diagram of a basic, audio and video frequency vector impedance meter (VIM). The VIM is operated in the constant current magnitude mode, although some VIMs can operate in the constant voltage mode. A wide band current transformer is used to sense the current flowing through the unknown impedance to ground. The voltage output of the current transformer is conditioned and used to parametrically adjust the applied voltage to maintain constant current. The microcomputer in the VIM calculates the impedance magnitude from the magnitudes of V and I , and also samples and presents the output of the phase meter. In the phase meter, the analog V and I signals are converted to TTL signals by high speed comparators and these signals are inputs to a NAND gate digital phase comparator, such as shown in Figure 8.55. The admittance, Y , can be easily found by changing the sign of the phase angle of Z , and taking the reciprocal of the magnitude of Z . VIMs can also give the user impedance in rectangular vector form (i.e., R and X), by multiplying $|Z|$ by $\cos(\theta)$ and $\sin(\theta)$ respectively. Once X is known and the impedance is known to be inductive, the equivalent L can be found by dividing X by the known $2\pi f$. A similar means can be used to obtain a numerical value of C when the impedance is known to be from a capacitor.

Hewlett-Packard offers a number of VIMs covering a wide range of frequencies, useful for a variety of applications, including the screening of critical components in manufacturing processes. We will discuss the specifications of one representative HP VIM to give the reader a feeling for the state-of-the-art in this type of instrument. The HP model 4192A LF Impedance Analyzer (VIM) operates in the range 5 Hz–13 MHz. This instrument can measure 11 impedance related parameters over its frequency range, namely $|Z|$, $|Y|$, θ , R , X , G , B , L , C , Q and D . Outputs are read on two, four and one-half digit numerical displays, and can be accessed through the HP-IB (IEEE-488) bus. The range of Z , X and R is $0.1 \text{ m}\Omega$ – $1.3 \text{ M}\Omega$, and the range of Y , G and B is 1 nS – 13 S . Frequency accuracy is $\pm 50 \text{ ppm}$, and frequency synthesizer steps are 0.001 Hz for 5 Hz–10 kHz, 0.01 Hz for 10–100 kHz, 0.1 Hz for 100 kHz–1 MHz and 1 Hz

for 1–13 MHz. Series or parallel equivalent circuit models may be selected according to user requirements for impedance and admittance measurements. The complete range of inductance which can be measured is from $10\text{ }\mu\text{H}$ (at high frequencies) to 1 kH (at low frequencies) with a basic accuracy of $\pm 0.27\%$. Capacitance can be measured in the range 0.1 fF – 199 mF , with a basic accuracy of $\pm 0.15\%$. $D = 1/Q$ can be measured in the range 10^{-4} – 20 with an accuracy of 10^{-3} for C measurements and 0.003 for L measurements. Maximum % accuracy in measuring the 11 impedance parameters is a function of frequency and the value of the parameter being measured. The best accuracy generally occurs between 100 Hz and 1 MHz, and it varies inversely with parameter magnitude (e.g., best phase accuracy is obtained when measuring low values of Y , G or B , and when measuring low values of $|Z|$, R or X ; likewise, best accuracy in measuring Z , R or X occur for low ohm values of these parameters, etc.).

The HP 4191A RF impedance analyzer is another VIM intended to work in the 1–1000 MHz range. In addition to measuring the 11 impedance parameters mentioned above, this meter also measures the reflection coefficient, $\Gamma(\omega)$ of a standard transmission line with characteristic impedance, Z_o , terminated with an unknown impedance, Z_L . $\Gamma(\omega)$ is defined as:

$$\Gamma(\omega) = \frac{V_r}{V_i} = \frac{Z_L - Z_o}{Z_L + Z_o} = \frac{VSWR - 1}{VSWR + 1} = \Gamma \angle \theta \quad (8.192)$$

The HP 4191A VIM uses two microwave directional couplers to isolate the input wave, V_i , and the reflected wave, V_r . Hence, $\Gamma(\omega)$ can be calculated from the vector quotient and knowing Z_o , we can calculate Z_L . In addition to presenting Γ in polar form, the HP 4191A VIM also can give Γ_X and Γ_Y over a range of 10^{-4} – 1 with a resolution of 10^{-4} . Typical accuracy in measuring $\Gamma(\omega)$ is $\pm 0.2\%$ in the range 1–100 MHz, rising to about $\pm 0.5\%$ at 1 GHz.

In summary, vector impedance meters are seen to be versatile, modern, microprocessor based instruments, capable of measuring a wide variety of impedance related parameters, including inductance, capacitance, Q and D over a wide range of frequencies and values. Modern VIMs have IEEE 488 bus I/O, which allows them to be directly controlled by, and report their measurements to a computer system. They also are relatively expensive.

8.12 Chapter Summary

Chapter 8 dealt with descriptions of the common means of measuring electrical parameters including dc voltage, dc current, electric field strength, ac voltage, ac current, phase, frequency, resistance, capacitance, inductance, vector impedance and vector admittance. The most accurate means of measurement were seen to be null methods, in which precise parametric standards are used for comparison. Null methods were treated in detail in Chapters 4 and 5.

The tremendous range of component values used in electrical engineering systems design generally means that no one instrument type, design or frequency is suitable to make accurate measurements over the entire practical range of a parameter's values. Due to the presence of parasitic capacitance in inductors, and parasitic inductance in capacitors, high frequency measurement of inductance and capacitance were seen to be more realistically treated as complex impedance or admittance measurements.

Finally, we described the organization of modern, microprocessor controlled, vector impedance meters which are ubiquitous instruments for the characterization of two-terminal device impedances over wide ranges of frequencies.

Problems

- 8.1 An $\eta=20,000 \Omega/V$ dc voltmeter is designed to have V_{FS} full scale volts. It is used to measure the dc voltage between two nodes in a dc circuit (Figure P8.1). Looking into the nodes, we see a Thevenin equivalent circuit with V_{OC} and R_{TH} as shown. Since the voltmeter draws current, the dc voltage read by the meter, V_m , will be less than V_{OC} by an amount ΔV . That is, $V_m = V_{OC} - \Delta V$.

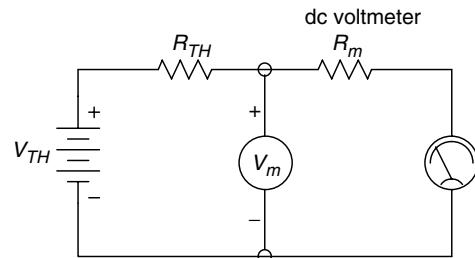


FIGURE P8.1

- (A) Give an expression for the total input resistance of the voltmeter.
 (B) Derive an expression for the fractional error in the reading of V_{OC} , $\Delta V/V_{OC}$, in terms of V_{FS} , η and R_{TH} .
- 8.2 A series ohmmeter having $R_H = 10^4 \Omega$ and $V_{OC} = 9.0 \text{ V}$ is connected to a silicon diode, as shown in Figure P8.2. The diode has $v_T = 0.025 \text{ V}$ and $I_{RS} = 10 \mu\text{A}$. What resistance does the ohmmeter read?

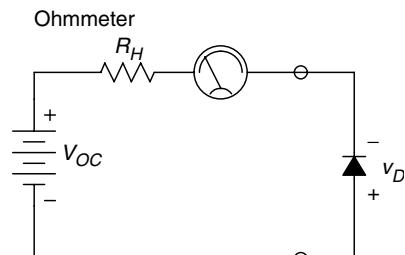


FIGURE P8.2

- 8.3 Two series ohmmeters are connected as shown in Figure P8.3 to a $1\text{k}\Omega$ resistor. What resistance does each ohmmeter read? Note that $S = I_x/I_{FS} = I_x/(V_{OC}/R_H) = R_H/(R_H + R_x)$.
- 8.4 A charge amplifier, shown in Figure P8.4, is used to measure the charge on a helicopter which has just landed. The helicopter's body capacitance on the ground is $C_H = 1 \text{ nF}$. The steady state dc voltage at the op-amps' output, neglecting V_{OS} and I_B , is -4.0 V .
- (A) Find the charge in coulombs on helicopter, Q_H . Is Q_H positive or negative?

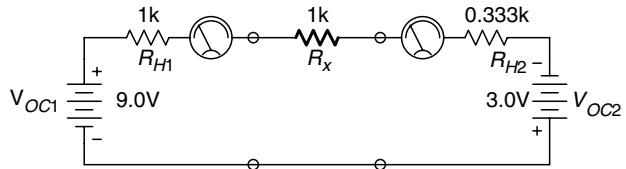


FIGURE P8.3

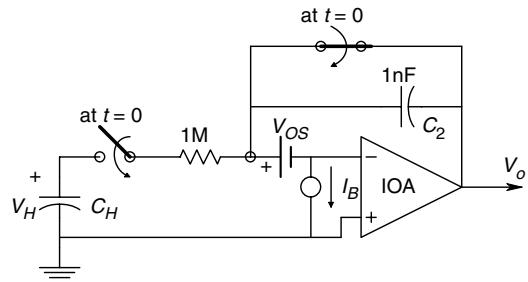


FIGURE P8.4

- (B) Using superposition, find the error voltage at the op-amp's output after 10 seconds due to $I_B = 50 \text{ fA}$, $V_{OS} = 1.0 \text{ mV}$.

8.5 A Faraday magneto-optical system is shown in Figure P8.5. A conductor carrying I amperes is wound as an N -turn solenoid around a polarization preserving fiber optic cable (PPFOC). Current in this solenoid induces $\alpha + \psi$ rotation of linearly polarized light (LPL) in the PPFOC segment AB. Now N' turns of the PPFOC are wound around the current carrying conductor. N' is chosen so that a net rotation of the LPL of $-\alpha$ degrees occurs.

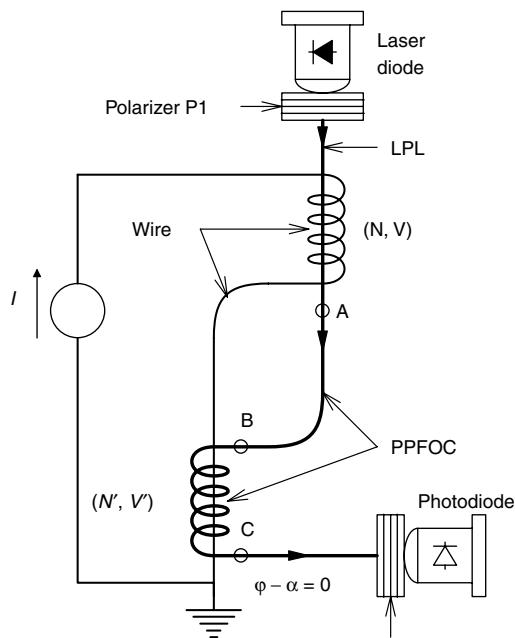


FIGURE P8.5

- (A) Find N'/N required so that $\alpha = \psi$, and the net polarization angle rotation in segment CD of the PPFOC is zero. Assume the Verdet constant, V , of the

solenoidal rotation is $3800^\circ/\text{(tesla meter)}$; V' of the FO system is $2.68 \times 10^{-4}^\circ/\text{(amp turn)}$. The average axial \mathbf{B} in the solenoid is $(3/4)\mu_o I/L_c$ tesla.

- (B) Find ψ for a current of 100 A. Let $N=100$ turns.

- 8.6 The electro-optical feedback system of Figure P8.6 uses a Gilham polarimeter to measure the large dc current, I , in the insulated conductor using the Faraday magneto-optical effect. The induced polarization rotation in the fiber optic cable is $\psi_i = V'N'I$ degrees. The feedback rotation is $\psi_F = VBI_F$. At null, $\psi_i = \psi_F$. V and V' are Verdet constants.

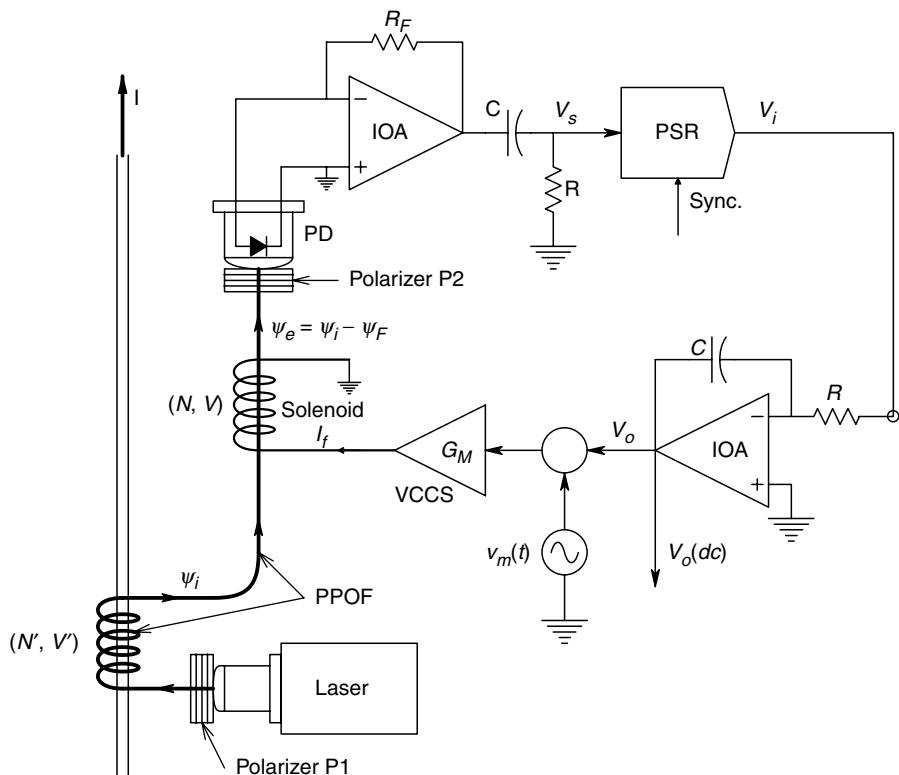


FIGURE P8.6

- (A) Derive an expression for V_o in terms of systems parameters and I in the cable, assuming a steady state null.
 (B) Make a block diagram for the system. Assume $V_i = -\psi_e I_4 K_s K_p \theta_m$ volts, where I_4 = the laser beam intensity in watts, K_s = detector gain in V/W, K_d = the PSR gain in V/V and θ_m = peak polarization angle modulation by the current, $G_M v_m(t)$.
 (C) Give the transfer function for the system, $\frac{V_o}{I}(s)$. Give the system's corner frequency in r/s.

- 8.7 A Hall sensor is used to measure a large dc current on the order of 100s of amperes. The dc current through the conductor acts like a 1-turn coil to induce a dc \mathbf{B}_y in the ferrite 'C' core (Figure P8.7). The measurement system uses dc feedback to generate

a \mathbf{B}_f that nulls the net B in the core. That is, $\mathbf{B}_e = \mathbf{B}_y - \mathbf{B}_f$. In order to gain sensitivity, the Hall output voltage is modulated at 1 kHz by passing an ac, I_x , into the Hall device. That is, $i_x(t) = I_{xo} \sin(2\pi 10^3 t)$. Thus, $v_H(t) = KB_e \sin(2\pi 10^3 t)$. \mathbf{B}_e in turn is proportional to $(I_{dc} - NI_F)$. N is the number of turns of the feedback coil on the core and I_F is the dc feedback current. At null, $v_H(t)$ and the dc, $V_2 \rightarrow 0$, and $I_{dc} = NI_F$. $I_F = G_M V_o$.

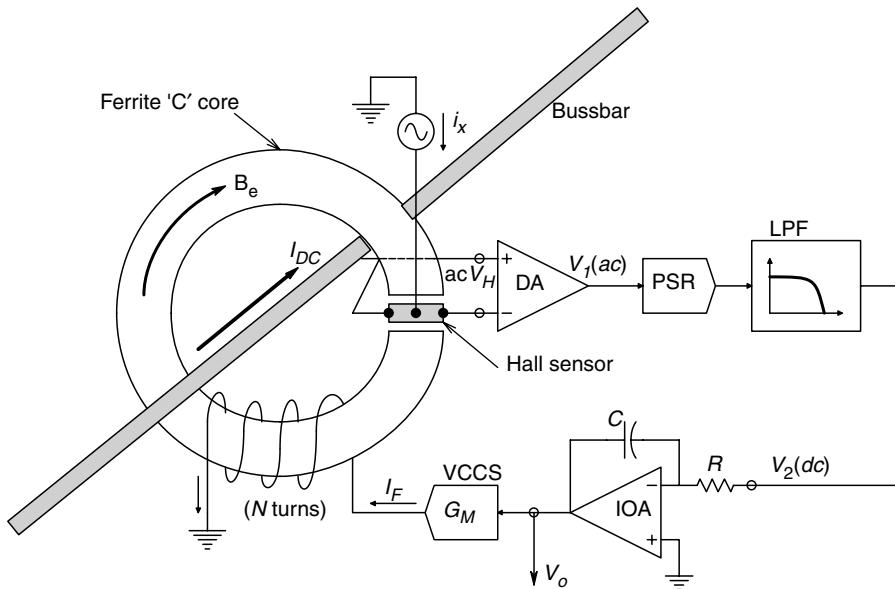


FIGURE P8.7

Derive an expression for I_{dc} in terms of the integrator dc output voltage, V_o .

- 8.8 A hall sensor is used to make an ac milliammeter. A solenoid coil makes an axial flux density of $B_y = \mu_0 N i_s(t)/(2b)$ Wb/m², where $N = 100$ is the number of turns, $b = 0.01$ m is the length of the coil, and $\mu_0 = 4\pi \times 10^{-7}$ Ns²/C². The Hall sensor is connected as shown in Figure P8.8. $I_X = 0.01$ A(dc), $R_H = 62.5$ (Hall constant) and $h = 0.0001$ m (sensor thickness). The unknown current is sinusoidal, $i_s(t) = I_S \sin(\omega t)$.

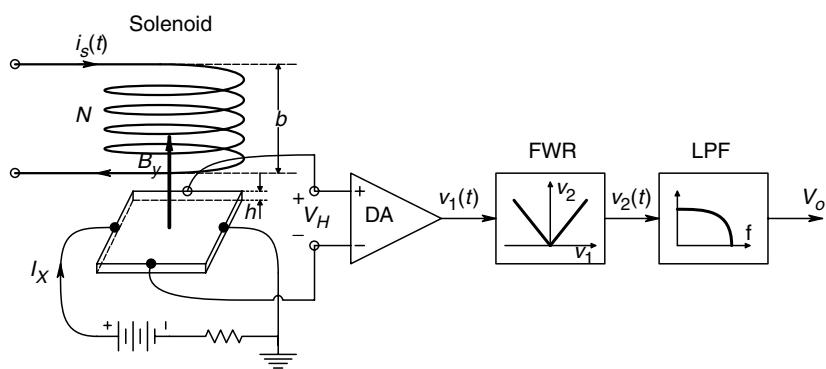


FIGURE P8.8

- (A) Find an expression for the Hall voltage, $v_H(t)$, in terms of system parameters and I_S .
- (B) $v_H(t)$ is amplified by a DA with gain $K_V = 100$. The ac voltage, $v_1(t)$, is full wave rectified and low-pass filtered to form a dc output, V_o , proportional to I_S . Both the LPF and the rectifier have unity gain. Find I_S to make $V_o = 1.0$ V.
- 8.9 A Hall sensor is used to make a dc wattmeter (Figure P8.9). In order to gain sensitivity, $i_x(t)$ is made an ac current proportional to V_L by using an analog multiplier, $v_s(t) = V_S \sin(\omega t)$, $i_x(t) = (G_M V_L V_S / 10) \sin(\omega t)$. Thus the Hall voltage, V_H , is an ac signal whose peak value is proportional to $P_{dc} = V_L I_L$. The magnetic flux density is given by $B_y = I_L \mu_0 N / 2b$ tesla.

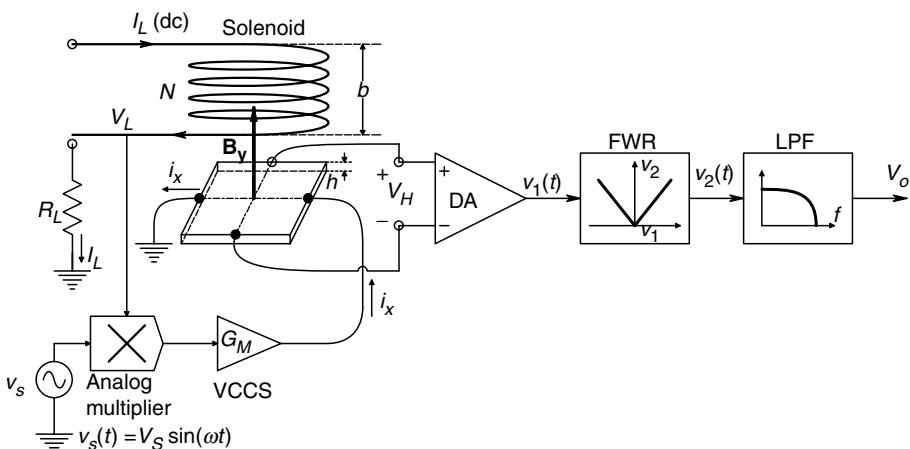


FIGURE P8.9

- (A) Find an expression for the Hall voltage, $v_H(t)$.
- (B) Plot and dimension the output of the full wave rectifier.
- 8.10 Figure P8.10 illustrates an ac Wattmeter that measures the real, average power dissipated in the load, Z_L . Note that $v_L(t) = V_L \sin(\omega t)$, $i_L(t) = I_L \sin(\omega t + \phi)$. Refer to Problem 8.9 for relations governing the Hall sensor's operation.
- (A) Find an expression for $v_H(t)$.
- (B) Find an expression for V_o .
- 8.11 Figure P8.11 shows the circuit for a true RMS voltmeter based on two matched, vacuum thermocouples (VTCs). It uses a Type 1 feedback loop in which \bar{V}_2 in the steady state and is equal to zero. The dc integrator output causes a dc current, I_F , to flow in the reference thermocouple heater, such that its temperature equals the temperature in the input thermocouple heater. Thus, for $V_2 = 0$, the TC voltages, $V_M = V_R$.
- (A) Derive an expression for the dc output, V_M , of the input vacuum TC.
- (B) Derive an expression for the RMS v_s , assuming steady state conditions.

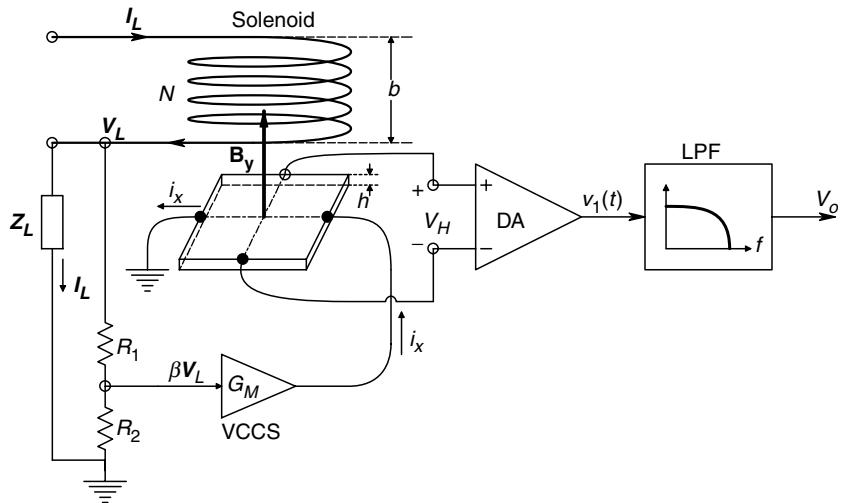


FIGURE P8.10

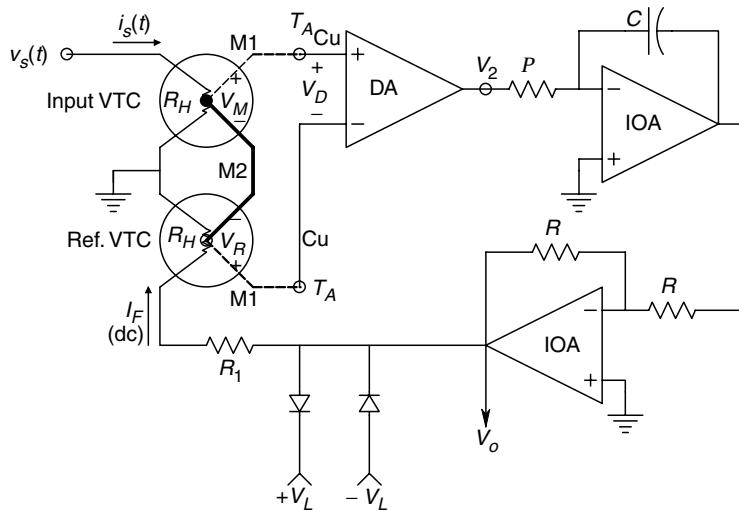


FIGURE P8.11

8.12 A manufacturing process for sheet Teflon™ winds the finished product at high speed, under tension, on a roll. Although steps are taken to control surface static charges, some charge may accumulate on the roll, which is a dielectric. The manufacturer wishes to sense the magnitude and sign of the surface charge density in order to adjust processing steps to null it. Assume the sheet winds on the roll at a constant linear velocity, $v = 10 \text{ m/s}$. The top surface is assumed to have an average charge density of $\rho \text{ C/m}^2$. A thin, soft wire brush of width $W = 0.5 \text{ m}$ makes contact with the top of the moving plastic sheet and sweeps the surface it touches clear of loose charges. Two circuits are proposed to measure ρ . In A, the switches periodically switch to the measurement (M) position for T_m seconds, then to the G position for T_g seconds in order to ensure that V_o and the charge on C_F are zero. V_o is read at the end of each T_m second interval by a sample-and-hold amplifier (not shown). In circuit B, the current from the brush is converted continuously

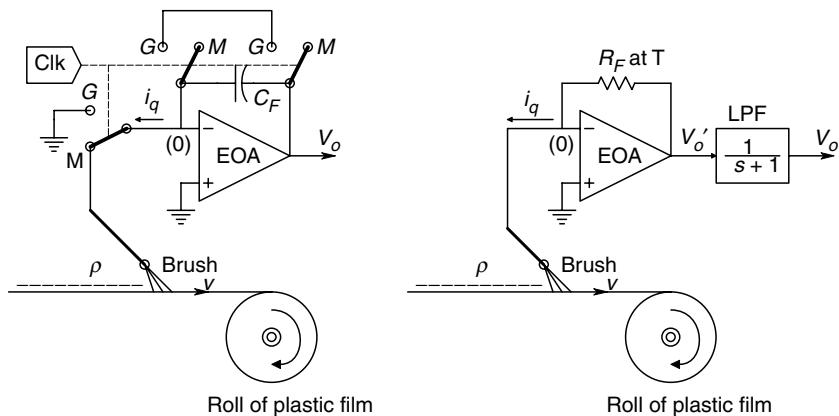


FIGURE P8.12

to output voltage V'_o . V'_o is low-pass filtered by a real pole filter with a 1 s time constant. The op-amp's $e_{na} = 10 \text{ nV}/\sqrt{\text{Hz}}$, $i_{na} = 1 \text{ fA}/\sqrt{\text{Hz}}$, $R_F = 10^6 \Omega$ at 300 K. All noises are white.

- For the A circuit, derive an expression for ρ from V_o at $t = T_m$.
- For the B circuit, derive an expression for ρ in terms of V_o and system parameters.
- Find a numerical value for ρ_{min} that will give a mean squared output SNR at V_o of 10.

9

Digital Interfaces in Measurement Systems

9.1 Introduction

As we have seen in Figure 1.1, a modern instrumentation system generally includes a computer which is used to supervise, coordinate and control the measurements, and is often used to store data (data logging), condition data display it in a meaningful, summary form on a monitor. In this chapter, we describe the hardware associated with the conversion of analog information to digital formats, and noise and resolution problems associated with the analog-to-digital conversion (ADC) process.

We also examine various means of digital communication with instruments, including the IEEE-488.2 bus, various serial data protocols and other parallel bus architectures used with PCs. Commercially available interface cards are described and problems associated with sending both analog and digital data over long distances at high data rates on cables are analyzed.

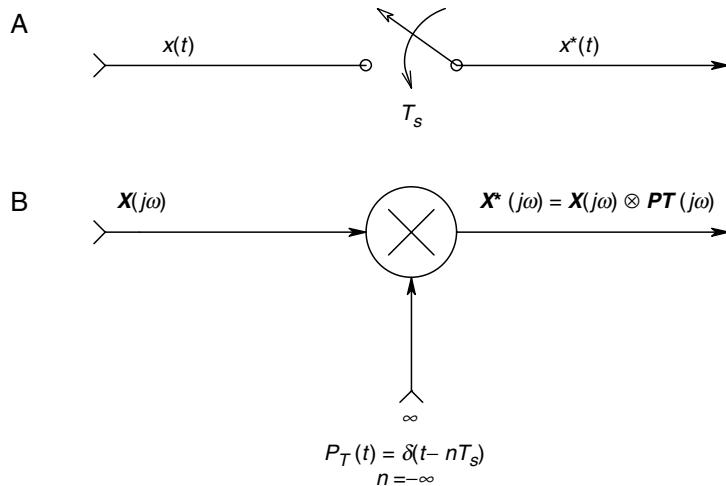
In describing data conversion interfaces, it is expedient to first consider *digital-to-analog converters* (DACs) because these systems are used in several designs for *analog-to-digital converters* (ADCs). Data conversion from analog to digital or digital to analog forms is often done continuously and periodically. Continuous, periodic data conversion has a significant effect on the information content of the converted data, as we will show in the next section. Theoretically, a noise free analog signal sample has infinite resolution. Once the analog signal sample has been converted to digital form, it is represented by a digital (binary) number of a finite number of bits (e.g. 12), which limits the resolution of the sample (one part in 4096 for 12 bits). This obligatory rounding off of the digital sample is called *quantization* and the resulting errors can be thought of as being caused by a *quantization noise*. It is shown in Section 9.4 that adding broadband noise (dithering) to the analog signal being digitized mitigates the effect of quantization noise on signal resolution.

Integrated circuit ADCs and DACs are available in a wide spectrum of specifications. They are available with quantization levels ranging from 6 to over 20 bits and with conversion rates ranging from low audio frequencies to less than 1 ns per sample. Most ADCs and DACs have parallel digital outputs or inputs, respectively, although a few are designed to use serial data I/O protocols.

Before discussing data conversion hardware, we will first discuss the linear dynamics of data conversion, quantization noise and dithering.

9.2 The Sampling Theorem

The sampling theorem is important because it establishes a criterion for the minimum sampling rate which must be used to digitize a signal with a given, low-pass, power

**FIGURE 9.1**

A. Symbol for an ideal sampling process (analog-to-digital conversion). T_s = the period between successive samples. B. Impulse modulation equivalent to ideal sampling. $X^*(j\omega) = \text{Fourier transform of the sampled signal}$. $X^*(j\omega)$ can be expressed as the complex convolution of $X(j\omega)$ and $P_T(j\omega)$. (Source: Northrop, 1990)

density spectrum. The relation between the highest significant frequency component recoverable in the signal's power density spectrum and the sampling frequency is called the *Nyquist criterion*. The implications of the Nyquist criterion and *aliasing* are described below.

Since analog-to-digital data conversion is generally a periodic process, we will first analyze what happens when an analog signal, $x(t)$, is periodically and ideally sampled. The ideal sampling process generates a data sequence from $x(t)$ only and exactly at the sampling instants, when $t = nT_s$, where n is an integer ranging from $-\infty$ to $+\infty$ and T_s is the *sampling period*. It is easy to show that an ideal sampling process is mathematically equivalent to *impulse modulation*, as shown in Figure 9.1. Here the continuous analog signal, $x(t)$, is multiplied by an infinite train of unit impulses or delta functions which occur only at the sampling instants. This multiplication process produces a periodic number sequence, $x^*(t)$, having infinite resolution at the sampler output. In the Frequency Domain, $X^*(j\omega)$ is given by the complex convolution of $X(j\omega)$ with the Fourier transform of the pulse train, $P_T(j\omega)$. In the Time Domain, the pulse train can be written as:

$$P_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (9.1)$$

Here, $\delta(t - a)$ is a unit impulse (delta function) which is defined only at the point $t = a$. That is, it has infinite height, zero width and unity area at $t = a$, and is zero for non-zero argument. The periodic function, P_T , can also be represented in the time domain by a Fourier Series in complex form:

$$P_T(t) = \sum_{n=-\infty}^{\infty} C_n \exp(-j\omega_s t) \quad (9.2)$$

where

$$\omega_s = \frac{2\pi}{T_s} \text{ r/s} \quad (9.3)$$

The complex form Fourier series coefficients are given by:

$$C_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} P_T(t) \exp(+j\omega_s t) dt = \frac{1}{T_s} \quad (9.4)$$

Thus the complex Fourier series for the pulse train is found to be:

$$P_T(t) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \exp(-jn\omega_s t) \quad (9.5)$$

The sampler output is the time domain product of equation 9.5 and $x(t)$:

$$\begin{aligned} x^*(t) &= x(t) \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \exp(-jn\omega_s t) \\ &= \frac{1}{T_s} \sum_{n=-\infty}^{\infty} x(t) \exp(-jn\omega_s t) \end{aligned} \quad (9.6)$$

The Fourier theorem for complex exponentiation is:

$$\mathbf{F}\{y(t) \exp(-jat)\} \equiv \mathbf{Y}(j\omega - ja) \quad (9.7)$$

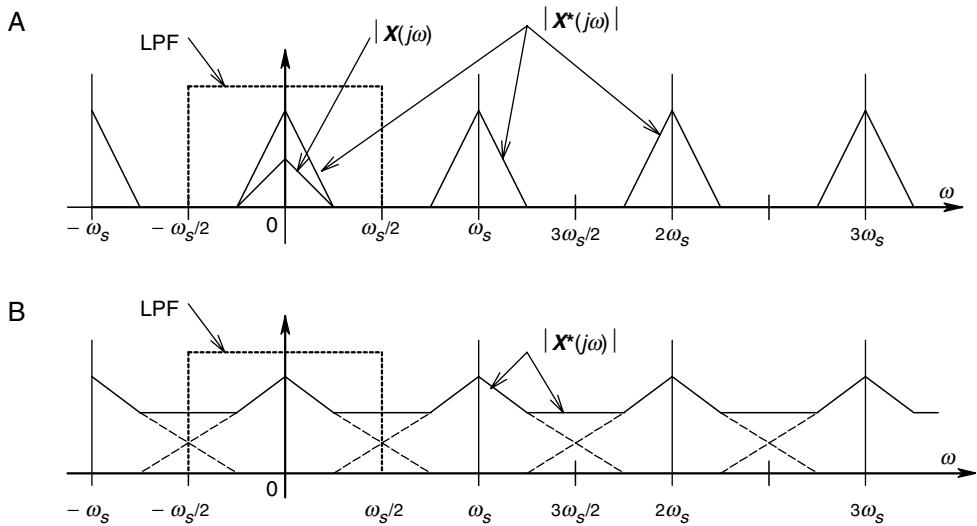
Using this theorem, we can write the Fourier transform for the sampler output as:

$$\mathbf{X}^*(j\omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \mathbf{X}(j\omega - jn\omega_s) \quad (9.8)$$

Equation 9.8 for $\mathbf{X}^*(j\omega)$ is the result of the complex convolution of $\mathbf{X}(j\omega)$ and $\mathbf{P}_T(j\omega)$. It is in the *Poisson sum form*, which helps us to visualize the effects of (ideal) sampling in the frequency domain and to understand the phenomenon of aliasing. In Figure 9.2A, we plot the magnitude of a typical $\mathbf{X}(j\omega)$. Note that $\mathbf{X}(j\omega)$ is assumed to have negligible power above the Nyquist frequency, $\omega_s/2$ r/s. When $x(t)$ is sampled, we see that $\mathbf{X}^*(j\omega)$ has a *periodic spectrum*.

Note that $x(t)$ can be recovered from the sampler output by passing $x^*(t)$ through an ideal low-pass filter, as shown in Figure 9.2A. In Figure 9.2B, we have assumed that the baseband spectrum of $x(t)$ extends beyond $\omega_s/2$ r/s. When such an $x(t)$ is sampled, the resultant $\mathbf{X}^*(j\omega)$ is also periodic, but the high frequency corners of the component spectra overlap. An ideal low-pass filter therefore, cannot uniquely recover the baseband spectrum and thus, $x(t)$. This condition of overlapping spectral components in $\mathbf{X}^*(j\omega)$ is called *aliasing* and it can lead to serious errors in digital signal processing.

Due to the problem of aliasing, all properly designed analog-to-digital conversion systems used in FFT spectrum analyzers and related equipment must operate on input signals that obey the Nyquist criterion, that is, the power density spectrum of $x(t)$, $S_{xx}(f)$, must have no significant power at frequencies above one-half the sampling frequency. One way to ensure that this criterion is met is to use properly designed analog low-pass, *anti-aliasing filters* immediately preceding the sampler. Anti-aliasing filters are generally high order, linear phase, LPFs that attenuate the input signal at least by 40 dB at the Nyquist frequency.

**FIGURE 9.2**

A. Magnitude of the Fourier transform of the sampler output. Note the spectrum is periodic at the sampling frequency. The input spectrum contains no power above the Nyquist frequency, $\omega_s/2$. B. Magnitude of the Fourier transform of the sampler output when the input signal's spectrum extends beyond the Nyquist frequency. This condition produces aliasing.

Many designs are possible for high order anti-aliasing filters. For example, Chebychev filters maximize the attenuation cut-off rate at the cost of some passband ripple. Chebychev filters can achieve a given attenuation cut-off slope with a lower order (fewer poles) than other filter designs. It should be noted that in the limit as passband ripple approaches zero, the Chebychev design approaches the Butterworth form of the same order, which has no ripple in the passband. Chebychev filters are designed in terms of their order, n , their cut-off frequency and the maximum allowable peak-to-peak ripple (in decibels) of their passband.

If ripples in the frequency response stopband of an anti-aliasing filter are permissible, then elliptical or Cauer filter designs may be considered. With stop band ripple allowed, even sharper attenuation in the transition band than obtainable with Chebychev filters of a given order can be obtained. Elliptic LPFs are specified in terms of their order, their cut-off frequency, the maximum peak-to-peak passband ripple, and their minimum stopband attenuation.

Bessel or Thomson filters are designed to have linear phase in the passband. They generally have a ‘cleaner’ transient response, that is, less ringing and overshoot at their outputs, given transient inputs.

As an example of anti-aliasing filter design, Franco (1988) shows a sixth-order Chebychev anti-aliasing filter made from three, concatenated, Sallen and Key (quadratic) low-pass modules. This filter was designed to have an attenuation of 40 dB at the system’s 20 kHz Nyquist frequency, a corner frequency of 12.8 kHz, a –3 dB frequency of 13.2 kHz, and a ± 1 dB passband ripple.” [Quoted from Northrop, 1990]

One problem in the design of analog anti-aliasing filters to be used with instruments having several different sampling rates is adjusting the filters to several different Nyquist frequencies. One way of handling this problem is to have a fixed filter for each separate sampling frequency. Another way is to make the filters easily tuneable, either by digital or analog voltage means. Some approaches to the tuneable filter problem are discussed in Section 7.3 in Northrop (2004), and Chapter 10 in Northrop (1990).

Anti-aliasing LPFs are generally not used at the inputs of digital oscilloscopes, because one can see directly if the sampled waveform is sampled too slowly for resolution. An optional analog LPF is often used to cut high frequency interference on digital oscilloscope inputs. This, however, is not an anti-aliasing filter.

9.3 Quantization Noise

When a band limited analog signal is periodically sampled and then converted by an N -bit ADC to digital form, an uncertainty in the digitized signal level exists which can be considered to be equivalent to a broadband, quantization noise added to the analog signal, which is digitized by an infinite bit ADC. To illustrate the properties of quantization noise, refer to the quantization error generating model of Figure 9.3. A Nyquist limited, analog signal, $x(t)$, is sampled and digitized by an N -bit ADC. The ADC's digital output is the input to an N -bit DAC. The quantization error is defined at sampling instants as the difference between DAC output and the analog input signal.

Figure 9.4 illustrates the nonlinear function relating sampler input, $x^*(t)$, to the DAC output, $y(t)$. In this example, $N=4$. The quantizer has (2^N) levels and (2^N-1) steps. Compared to the direct path, the error $e(nT_s)$ can range over $\pm q/2$, where q is the voltage step size of the ADC. It is easy to see that:

$$q = \frac{V_{MAX}}{(2^N - 1)} \text{ V} \quad (9.9)$$

where V_{MAX} is the maximum (peak-to-peak) range of the ADC/DAC system. For example, if a 10-bit ADC is used to convert a signal ranging -5 to $+5$ V, then by equation 9.9, $q=9.775$ mV. If the $x(t)$ has zero mean and its probability density function (PDF) has a standard deviation, σ_x , which is large compared to q , then it can be shown that the PDF of $e(nT_s)$ is well-modelled by a uniform (rectangular) density function over $|e| \leq q/2$. This rectangular PDF is shown in Figure 9.5. The mean squared error voltage is found from:

$$E\{e^2\} = e^2 = \int_{-q/2}^{q/2} e^2 (1/q) \, d\,e = \frac{(1/q)e^3}{3} \Big|_{-q/2}^{q/2} = \frac{q_2}{12} \text{ MSV} \quad (9.10)$$

Thus, it is possible to treat quantization error noise as a zero mean, broadband noise with standard deviation of $q/\sqrt{12}$ V, added to $x^*(t)$.

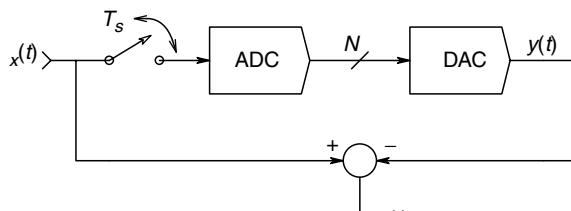
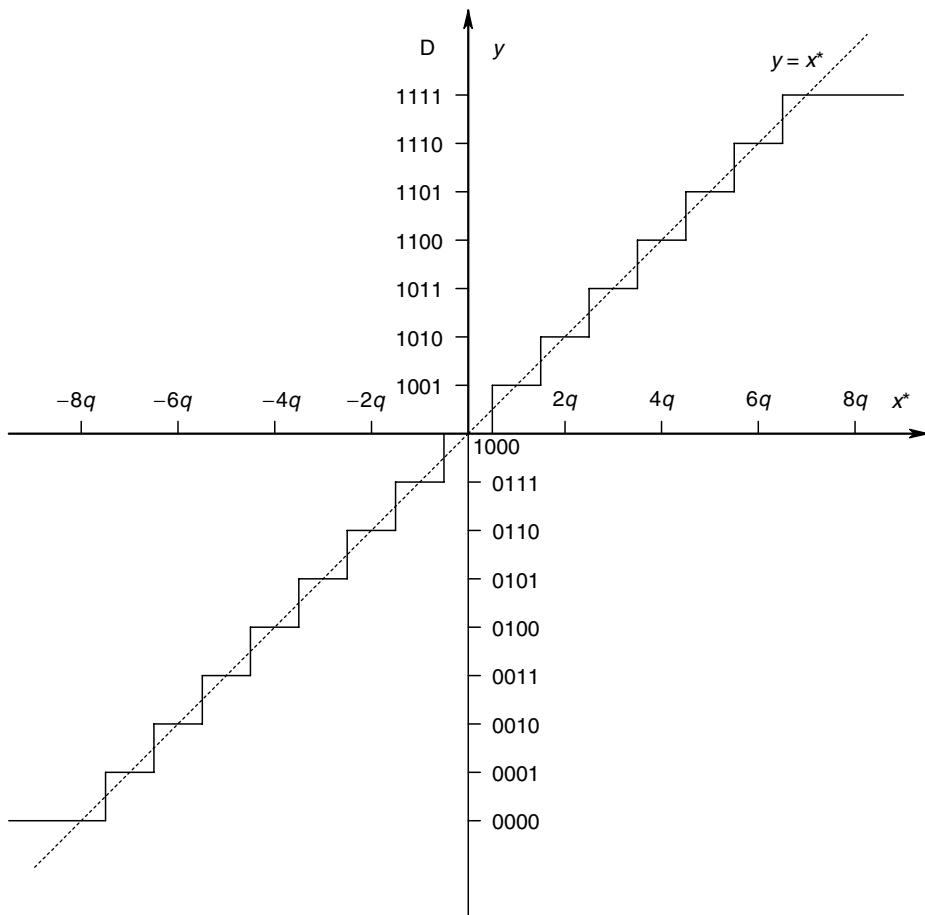
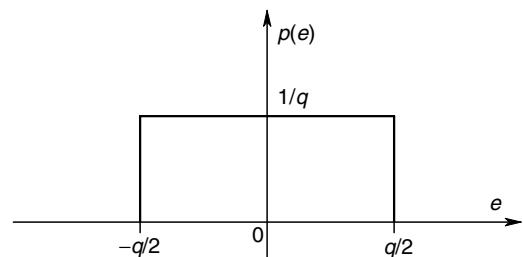


FIGURE 9.3

Block diagram of the system used to observe quantization error in the A/D process.

**FIGURE 9.4**

I/O diagram of a 4-bit quantizer (ADC/DAC in Figure 9.3). Note that there are $2^4=16$ levels and $(2^4-1)=15$ steps.

**FIGURE 9.5**

Probability density function of the quantization error of the system of Figure 9.3. See Figure 9.4 for the significance of \$q\$.

In order to minimize the effects of quantization noise for an \$N\$-bit ADC, it is important that the analog input signal, \$x(t)\$, uses nearly the full dynamic range of the ADC. In the case of a zero mean, time varying signal which is Nyquist band limited, gains and sensitivities should be chosen so that the expected peak \$x(t)\$ does not exceed the maximum voltage limits of the ADC. Also, if \$x(t)\$ has a Gaussian PDF, the dynamic range of the ADC should be about \$\pm 3\$ standard deviations of the signal. Under this condition, it is possible to derive an expression for the mean squared SNR of the ADC and its

TABLE 9.1

SNR values for an N -bit ADC treated as a quantizer

N	dB SNRq
6	31.2
8	43.4
10	55.4
12	67.5
14	79.5
16	91.6

Note: Total input range is assumed to be $6\sigma_x$ V.

quantization noise. Let the signal have an RMS value σ_x volts. From equation 9.9, we see that the quantization step size can be written as:

$$q \approx \frac{6\sigma_x}{(2^N - 1)} \text{ V} \quad (9.11)$$

This relation for q can be substituted into equation 9.10 for the variance of the quantization noise. Thus the output noise is:

$$N_o = \frac{q^2}{12} = \frac{36\sigma_x^2}{12(2^N - 1)} \text{ MSV} \quad (9.12)$$

Thus, the mean squared SNR of the quantizer is:

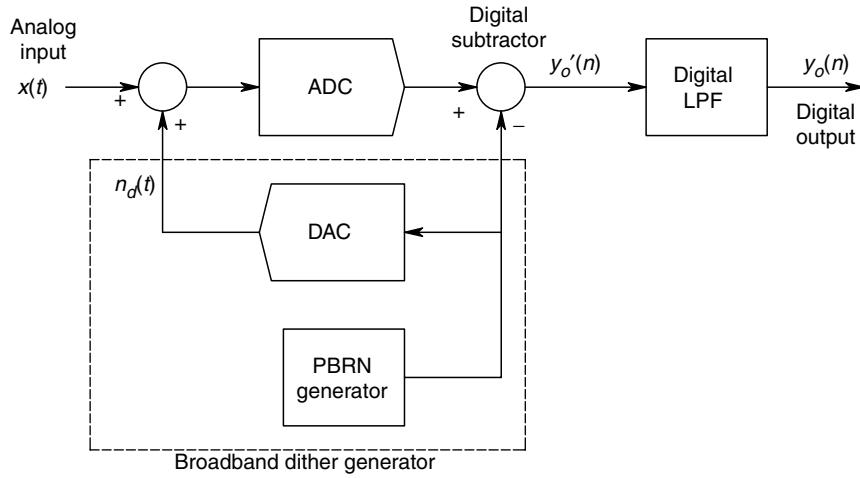
$$\text{SNRq} = (2^N - 1)/3 \text{ MSV/MSV} \quad (9.13)$$

Note that the quantizer SNR is independent of σ_x as long as σ_x is held constant under the dynamic range constraint described above. Table 9.1 summarizes the SNRq of the quantizer for different bit values. ADCs of 16 to 20-bit are routinely used in modern digital audio systems because of their low quantization noise.

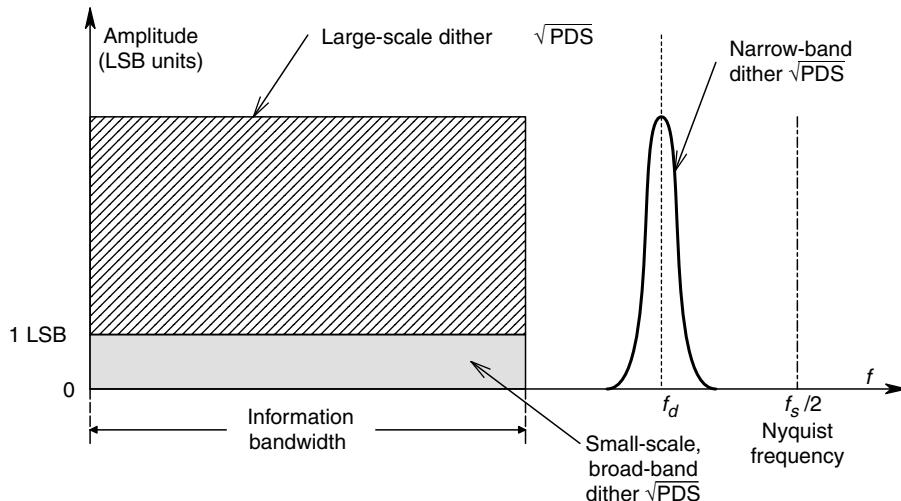
9.4 Dithering

Dithering is a method for statistically reducing the quantization errors and harmonic distortion inherent in the ADC process. In dithering, a signal uncorrelated with the input analog signal, $x(t)$, is added to it, and the sum is sampled and converted. Figure 9.6 illustrates an externally dithered ADC. The dither signal can be a narrow band (almost coherent) waveform with a power spectrum centered just below the Nyquist frequency, $f_s/2$, or a broadband noise whose spectrum covers the signal spectrum, as shown in Figure 9.7. Note that dither noise is subtracted digitally from the ADC's output after conversion.

Adding dither to a converter's analog input can provide significant improvements to the ability to extract signals below the resolution of the converter ($< \pm 1/2$ LSB) and in

**FIGURE 9.6**

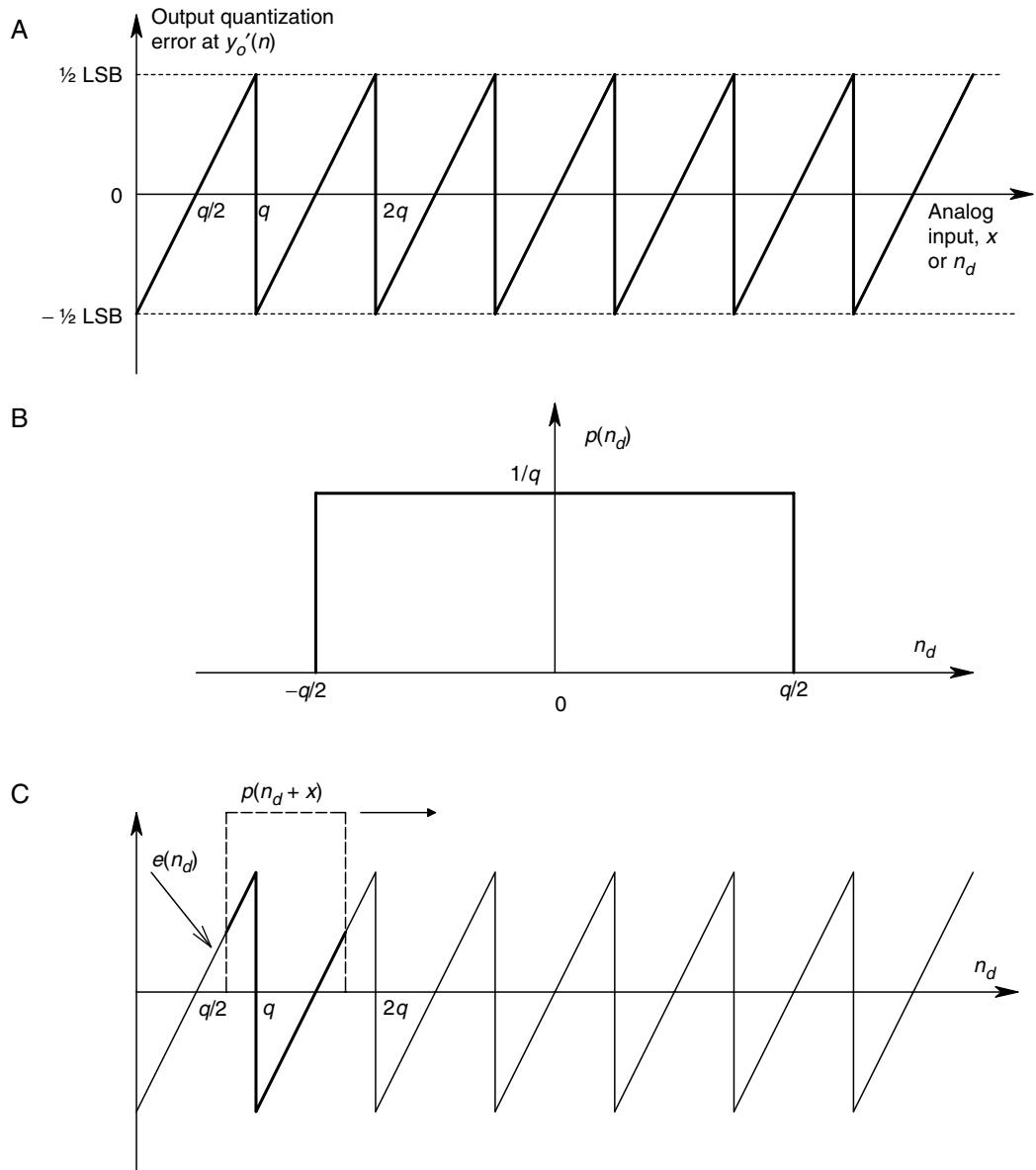
An externally dithered ADC.

**FIGURE 9.7**

RMS spectra of signal, broad and narrow band dithering signals.

linearizing an ADC's performance. The SNR improvements depend on the signal characteristics and the amount and type of dither. When combined with the time averaging property of subsequent digital low-pass filtering, dithering effectively removes, or smoothes, the quantization noise inherent in the stepwise transfer function of an ADC. To illustrate this property heuristically, let us consider the expected value of the dither noise at the output of the digital LPF (Figure 9.6). In general, the expected, or probability averaged value of the error transfer function, $e(n_d)$, is given by the well-known expression:

$$\bar{e} = \int_{-\infty}^{\infty} e(n_d) p(n_d) dn_d \quad (9.14)$$

**FIGURE 9.8**

A. Output quantization error. (Figure 9.6) B. Probability density function of the RV, n_d , the dither noise. C. Illustration of the convolution of $p(n_d)$ with the quantization error function, $e(n_d)$. The convolution acts to smooth and reduce the amplitude of the quantization error.

Here, $p(n_d)$ is the probability density function of the random variable, n_d , the dither noise. When dithering occurs, the analog signal of amplitude x is added to the RV, n_d . Now the expected transfer function value can be written as:

$$\overline{e(x)} = \int_{-\infty}^{\infty} e(n_d) p(n_d + x) dn_d \quad (9.15)$$

Equation 9.15 is mathematically a form of convolution of $p(n_d)$ with $e(n_d)$ (the quantization error as a function of noise input). The convolution process is illustrated in Figure 9.8C.

Note that the shift is x , and when the rectangular PDF of n_d is given the limits of $\pm q/2$, the integration result is zero for any $x \geq 0$.

According to Hewlett-Packard [Bartz, 1993]:

'The high order distortion terms associated with the polynomial expansion of the quantized transfer function are effectively removed by the dithering. This result is true for dither amplitudes that are integral multiples of the ADC LSB [q].' ... 'The third harmonic distortion is significantly improved by large scale dithering.'

In summary, we note that dithering provides a means to improve output SNR and reduce total harmonic distortion in a digital signal acquisition system.

9.5 Digital-to-Analog Converters

A digital-to-analog converter (DAC) is an IC device which converts an N -bit digital word to an equivalent analog voltage or current. DACs can operate at very high sampling rates, or can act as static, dc voltage sources. A DAC allows digital information which has been processed and/or stored by a digital computer to be realized in analog form. DACs are essential components in the design of CRT terminals, modern control systems digital audio systems and waveform synthesizers, to mention a few applications.

As we mentioned in the introduction to this chapter, it is necessary to understand DACs before we examine the designs of certain ADCs which use DACs as components. Most modern DACs are designed to accept straight binary inputs, although some units have been built that require BCD or gray code inputs. Most DACs use MOS or BJT switches to pass current through elected resistors in an $R-2R$ resistance ladder circuit. Either voltage or current reference sources are used to power the ladder and either internal or external op-amps are used to condition the DAC's output currents. DACs can be configured to have unipolar or bipolar outputs, depending on their application.

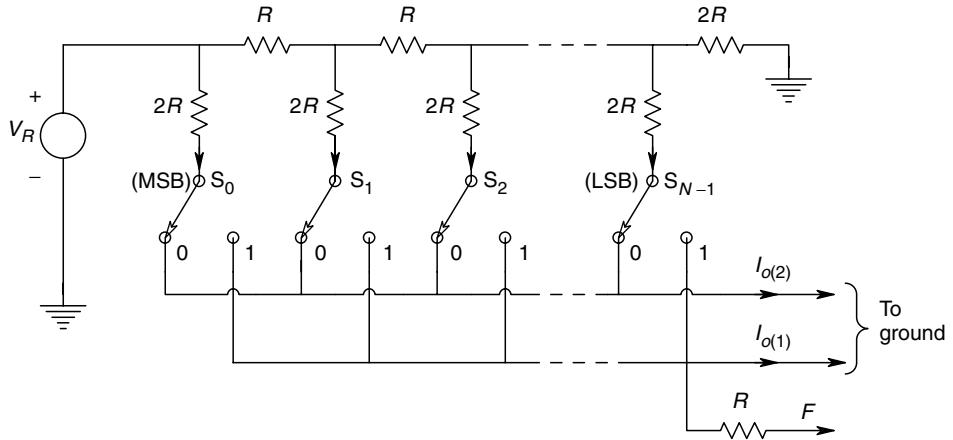
All DACs require a certain settling time for the analog output to reach steady state value following the digital input. This settling (conversion) time can be as long as 100 ms or as short as 10 ns. Immediately following a change in digital input, there is a DAC output transient, or glitch. The glitch can arise from the transient response of the op-amp(s) used to condition the output voltage, and also from internal switching transients in the DAC.

There are many designs for DACs which have evolved over the past twenty years or so. DACs can be classified by whether they use bipolar junction transistor switches, or MOS transistor technology. The following DAC architectures are found with MOS technology:

1. Binary weighted resistor DAC
2. $R-2R$ ladder
3. Inverted $R-2R$ ladder
4. Inherent monolithic ladder
5. Switched capacitor DAC

The following DAC designs generally use BJT technology:

1. Binary weighted current sources
2. $R-2R$ Ladder using current sources

**FIGURE 9.9**

An R - $2R$ current scaling DAC ladder. $S_0 \dots S_{N-1}$ are SPDT MOS switches actuated by the binary input signal. (Source: Northrop, 1990)

At present, most DACs use the current mode, R - $2R$ ladder design, as shown in Figure 9.9. This design has low switching transients, and is relatively fast. The switches are typically MOS transistors with low ON resistance. V_R is the DAC's dc reference voltage, generally +5 or +10 V, although in some designs, V_R can be a bipolar, time varying signal, allowing two-quadrant multiplication of V_R times the digital input. A DAC is called a *multiplying DAC* (MDAC) when it is operated in this mode. The DAC of Figure 9.9 has two output lines in which current must flow to ground, or an op-amp's summing junction virtual ground at all times. Due to the grounded output lines, the current in the switched $2R$ resistors remains constant, hence the node voltages, $V_1 \dots V_N$, remain constant, and there are minimum glitches caused by charging distributed capacitances from IC circuit elements to the IC substrate. Note that the V_R source sees a resistance of $R \Omega$ in the DAC and thus, the MSB current through S_1 is simply $V_R/2R$. Each successive switch current is half of that through the preceding switch. Thus, the maximum output current, $I_{o(1)MAX}$ is given by:

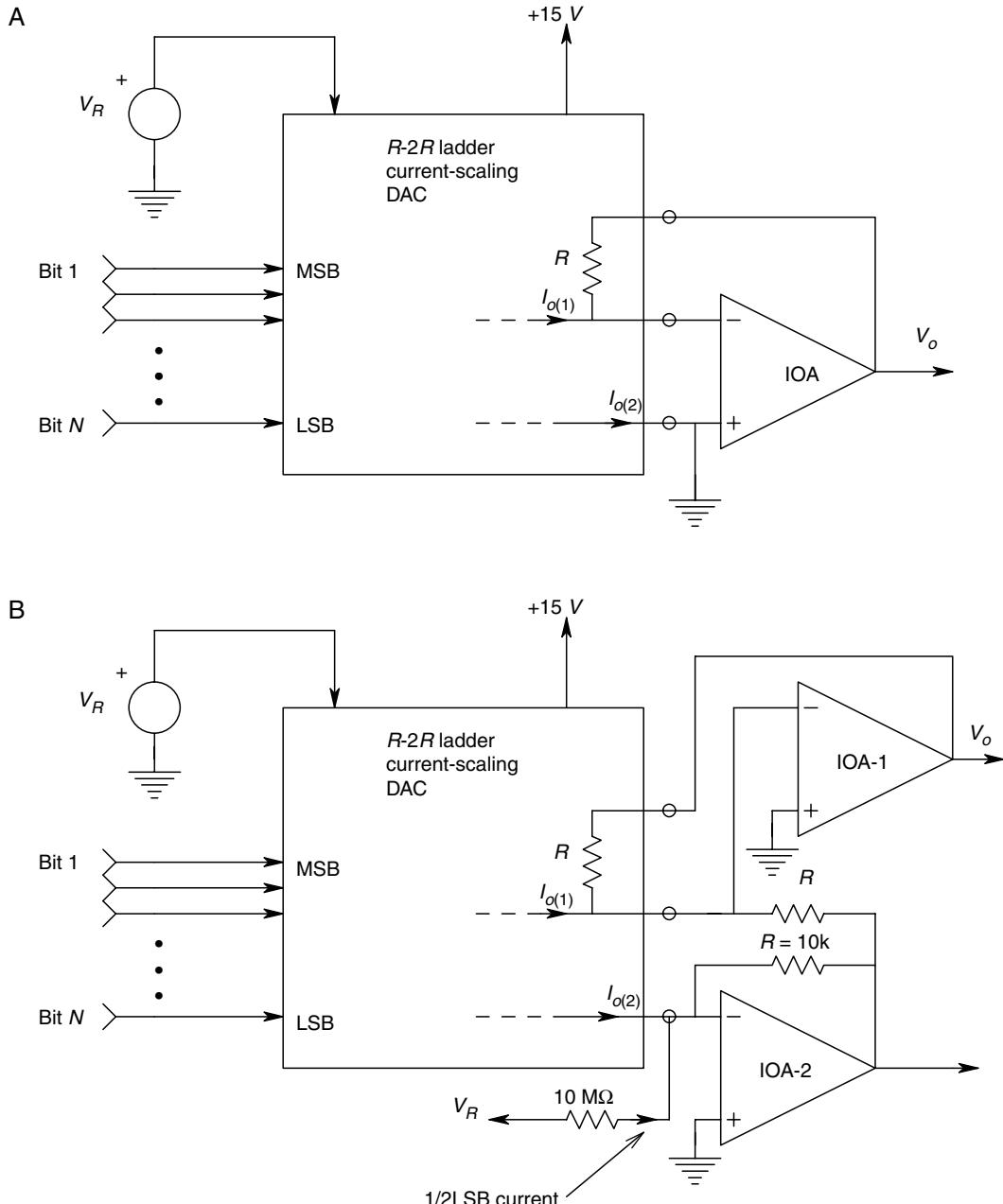
$$\begin{aligned} I_{o(1)MAX} &= (V_R/R)(1/2^1 + 1/2^2 + 1/2^3 + \dots 1/2^N) \\ &= (V_R/R) \left(\frac{2^N - 1}{2^N} \right) \end{aligned} \quad (9.16)$$

In general, the current outputs of this DAC are:

$$I_{o(1)} = (V_R/R) \sum_{k=1}^{k=N} (D_k/2^k) \quad (9.17a)$$

$$I_{o(2)} = I_{o(1)} = (V_R/R) \left[\left(\frac{2^N - 1}{2^N} \right) - \sum_{k=1}^{k=N} (D_k/2^k) \right] \quad (9.17b)$$

Here, D_k is the logic level controlling the k th switch, either 0 or 1. $k=1$ is the *most significant bit* (MSB) and $k=N$ is the least significant bit (LSB) of the input word.

**FIGURE 9.10**

A. Unipolar binary DAC circuit. The box contains the circuit of Figure 9.9. B. Circuit for offset binary (bipolar) DAC output. $R = 10\text{ k}\Omega$. (Source: Northrop, 1990)

Figure 9.10 illustrates how op-amps are used to convert the output currents given by equations 9.17a and b, to voltages. In the DAC circuit of Figure 9.9, note that there is an extra resistor of resistance R connected to the $I_{o(1)}$ node. This resistor is used as the feedback resistor for the op-amp transresistor of Figure 9.10A. Negative unipolar voltage output is obtained because $V_o = -RI_{o(1)}$, where $I_{o(1)}$ is given by equation 9.17a above. (A second inverter can be used to obtain a positive V_o .) Figure 9.10B illustrates the same DAC, given a bipolar output so that V_o ranges over approximately $\pm V_R$. In this case, the

TABLE 9.2

Coding and output for a 10-bit, offset binary DAC

<i>D</i>	<i>V_o</i>
11 1111 1111	-V _R (511/512)
10 0000 0001	-V _R (1/512)
10 0000 0000	0
01 1111 1111	+V _R (1/512)
00 0000 0000	+V _R (511/512)

Note: The 1/2 LSB offset voltage has been neglected for simplicity.

input word must be in offset binary format, as illustrated in Table 9.2. In the general case, the output of the offset binary DAC can be written:

$$\begin{aligned}
 V_o &= -RI_{o(1)} - V_2 = -RI_{o(1)} - (-RI_{o(2)}) + \frac{\Delta V_o}{2} \\
 &= -V_R \sum_{k=1}^N \frac{D_k}{2^k} + V_R \left[\left(\frac{2^N - 1}{2^N} \right) - \sum_{k=1}^N \frac{D_k}{2^k} \right] + \frac{\Delta V_o}{2} \\
 &= -2V_R \sum_{k=1}^N \frac{D_k}{2^k} + V_R \left(\frac{2^N - 1}{2^N} \right) + \frac{\Delta V_o}{2}
 \end{aligned} \quad (9.18)$$

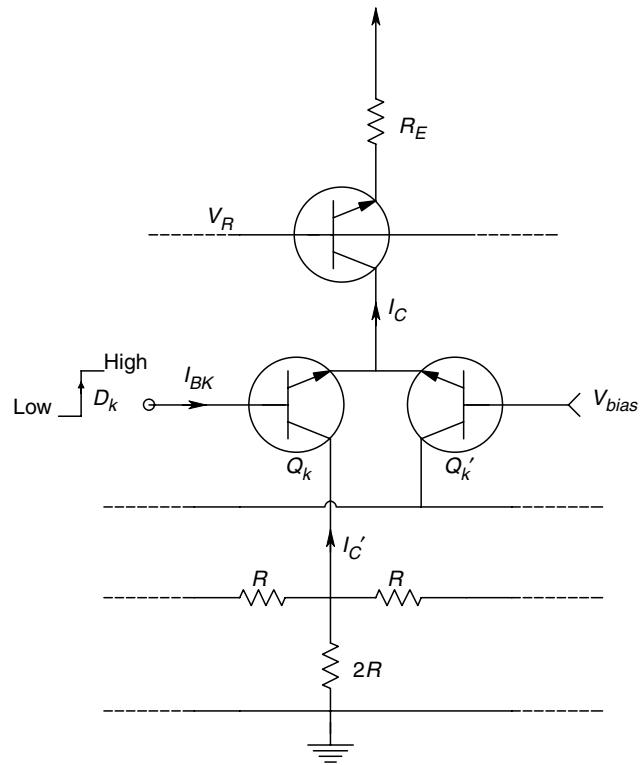
The $\Delta V_o/2$ term is a half-LSB step used to make the bipolar DAC's transfer function an odd function to minimize quantization error. Many other DAC designs exist. In many practical *R-2R* ladder circuits, especially those used for high speed operation, bipolar junction transistor (BJT) current sources are used to drive the ladder nodes. Switching of these current sources is often accomplished by differential current switches which allow the currents through the BJT current sources to remain constant whether or not they flow into the ladder nodes or ground. This design is illustrated in Figure 9.11. Switched capacitor or charge scaling DACs which use MOS IC technology are available. Instead of a ladder geometry, a capacitive voltage divider is used with the upper, switched capacitors having the values, C , $C/2^1$, ..., $C/2^{N-1}$. The lower capacitor of the divider is C pF. This circuit is shown in Figure 9.12. It can be shown [Northrop, 1990] that the voltage output of the charge scaling DAC is given by:

$$V_o = \frac{V_R}{C_T} \sum_{k=1}^N D_k (C/2^{k-1}) \quad (9.19)$$

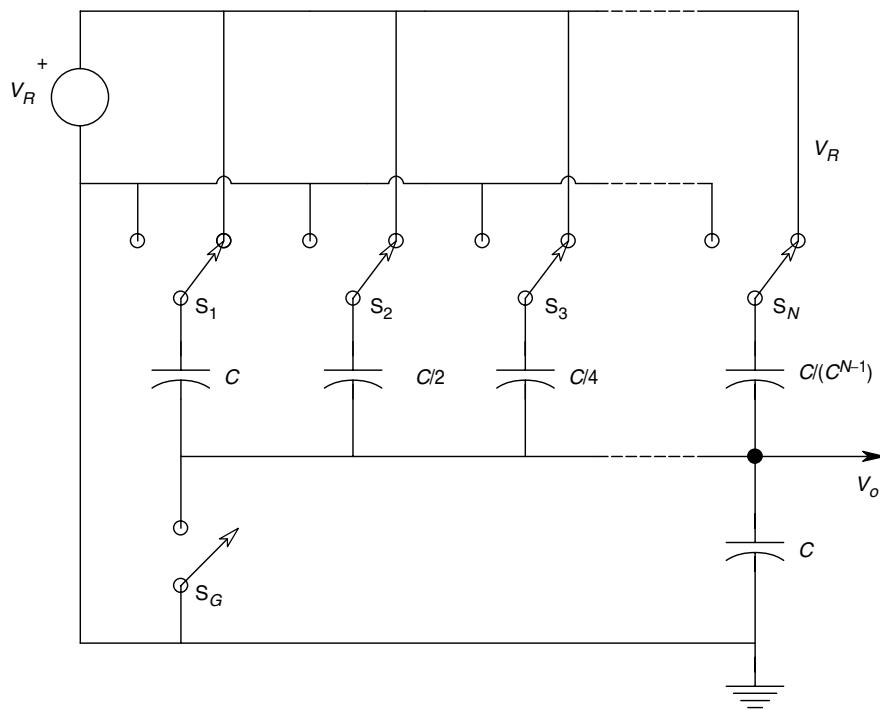
where C_T is the total capacitance of the switched array, given by:

$$C_T = \sum_{k=1}^N (C/2^{k-1}) + C/2^{N-1} = 2C \quad (9.20)$$

It is not practical to build switched capacitor DACs for N s much above 8 because of the large range in sizes required between the MSB and LSB capacitors.

**FIGURE 9.11**

A differential BJT current switch commonly used in R - $2R$, current scaling DACs. D_k = kth digital input. The actual kth ladder current is $I'_C = I_C - I_{BK}$. (Source: Northrop, 1990)

**FIGURE 9.12**

Simplified circuit of a switched capacitor or charge scaling DAC. The entire circuit is realized with MOS IC technology. (Source: Northrop, 1990)

Readers interested in the details of DAC designs and how DACs are specified, should consult Chapter 10 in Northrop (2004), Chapter 14 in Northrop (1990), Chapter 11 in Franco (1988), Chapter 4 in Tompkins and Webster (1988) and Chapter 2 in Zuch (1981).

9.6 The Hold Operation

Just as the sampling process can be interpreted in the frequency domain, so can the process of sample (or track) and hold of analog signals from a DAC. Note that the digital input to a DAC is generally periodic, with period T_s . The DAC's analog output from the n th digital input is generally held constant until the $(n+1)$ th input updates it. This process generates a stepwise output waveform if $\{b_k\}$ is changing. This process can be viewed as a linear filtering operation. The impulse response of the *zero order hold* (ZOH) filter to a unit input word at $t=0$, and zero inputs at sampling instants thereafter, is illustrated in Figure 9.13, and can be expressed mathematically as:

$$h_o(t) = U(t) - U(t - T_s) \quad (9.21)$$

where $U(t)$ is the unit step function, defined as zero for all negative argument, and 1 for all positive argument. $h_o(t)$ is thus, a pulse of unit height and duration $(0, T_s)$, else zero. The Laplace transform of $h_o(t)$ is:

$$H_o(s) = \frac{1}{s} - \frac{1}{s} e^{-sT_s} = \frac{1 - e^{-sT_s}}{s} \quad (9.22)$$

To find the frequency response of the ZOH, we let $s \rightarrow j\omega$, and use the Euler relation for $e^{j\theta}$. The ZOH's frequency response is easily shown to be:

$$\mathbf{H}_o(j\omega) = T_s \frac{\sin(\omega T_s/2)}{(\omega T_s/2)} e^{-j\omega T_s/2} \quad (9.23)$$

Note that the zeros in $\mathbf{H}_o(j\omega)$ occur at $\omega = n2\pi/T_s$ r/s, where $n = 1, 2, 3, \dots$, and $2\pi/T_s$ is the radian sampling frequency of the system. Thus, the overall process of sampling an analog signal, $x(t)$ and reconverting to (held) analog form by a DAC can be written in the frequency domain (neglecting quantization) as [Northrop, 1990]:

$$\mathbf{X}(j\omega) = \sum_{n=0}^{\infty} \frac{\sin(\omega T_s/2)}{(\omega T_s/2)} \mathbf{X}(j\omega - jn\omega_s) e^{-j\omega T_s/2} \quad (9.24)$$

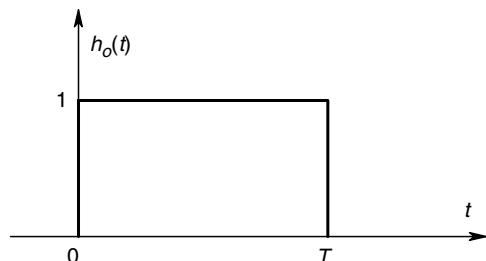


FIGURE 9.13

Impulse response of a zero-order hold filter.

Note that other more sophisticated holds exist. However, their realization is at the expense of some circuit complexity. For example, the first difference extrapolator hold which generates linear slope transitions between sampling instants (instead of steps) can be realized with three DACs, a resettable analog integrator and an analog adder [Northrop, 1990]. Its transfer function can be shown to be:

$$H_e(s) = \frac{1 - e^{-sT}}{s} + \frac{1 - 2e^{-sT} + e^{-2sT}}{s^2} - \frac{(1 - e^{-sT})e^{-sT}}{s} \quad (9.25)$$

It is left as an exercise for the reader to sketch and dimension $h_e(t)$.

9.7 Analog-to-Digital Converters

Various types of ADCs have evolved to meet specific applications in instrumentation, communications, control and audio. Obviously, the speed of the A-to-D conversion rate is not critical when digitizing a 'dc' measurand. However, the resolution may be very important and therefore, a large number of bits may be used. For example, a 16-bit ADC has resolution limited by $\pm 1/2$ LSB, or ± 7.63 ppm, a 20-bit ADC can resolve to ± 0.477 ppm. Commercial plug-in data acquisition cards for PCs of various types generally do not use ADCs of greater than 16 bits. Thus, custom ADC interface systems must be used if resolution better than ± 7.63 ppm is desired. The fastest ADCs are the *flash converters*, described below, which can convert 8 bits with a sampling period of less than 1 ns. Such speed is useful when measuring transient phenomena, such as partial discharges in power cable insulation, other insulation breakdown phenomena and transient events in particle physics and lasers. Commercial data acquisition cards for PCs generally have sampling rates on the order of 1 MHz for ADCs up to and including 12 bits. Signatec, Inc. offers a combined, 8-bit, 100 MHz digitizer with a 16-bit, digital signal processor on a PC/AT bus compatible card.

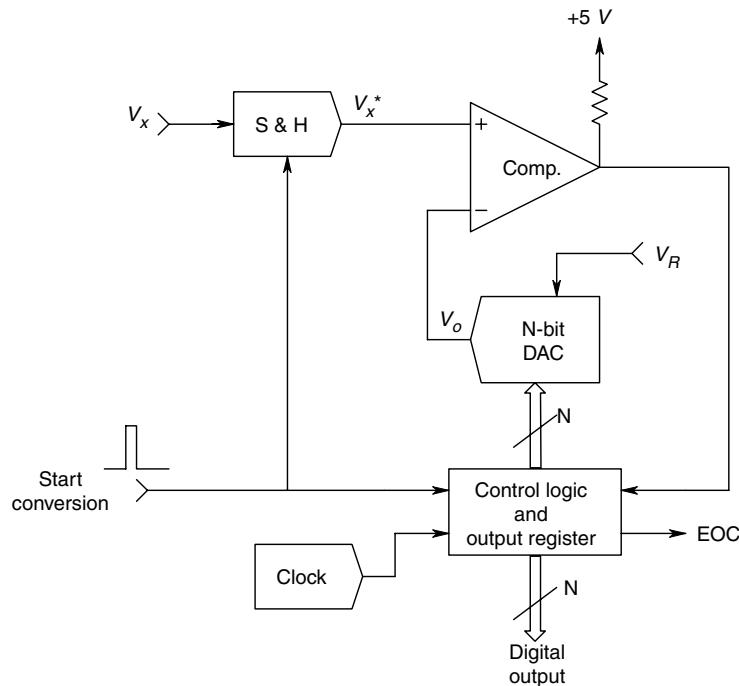
One might wonder why it is necessary to have 16-bit resolution when the display resolution of terminals and graphics plotters rarely exceeds 10 bits. The answer to this question lies in the lower quantization noise of the 16-bit systems and the fact that the statistics computed on the measurements have lower variances when calculated using 16-bit data. We will describe the organization, features and applications in measurements of five major types of ADCs:

1. Successive approximation converters
2. Tracking (servo) types
3. Dual slope, integrating converters
4. Flash (parallel) converters
5. Dynamic range, floating point converters
6. Sigma-Delta converters

DACs can be classified by whether they use bipolar junction transistor switches, or MOS transistors.

9.7.1 Successive-Approximation (SA) ADCs

SA ADCs are probably the most widely used class of ADC, with accuracies ranging 8–16 bits. They are available as single, LSI, digital/analog ICs. Their advantages include low

**FIGURE 9.14**

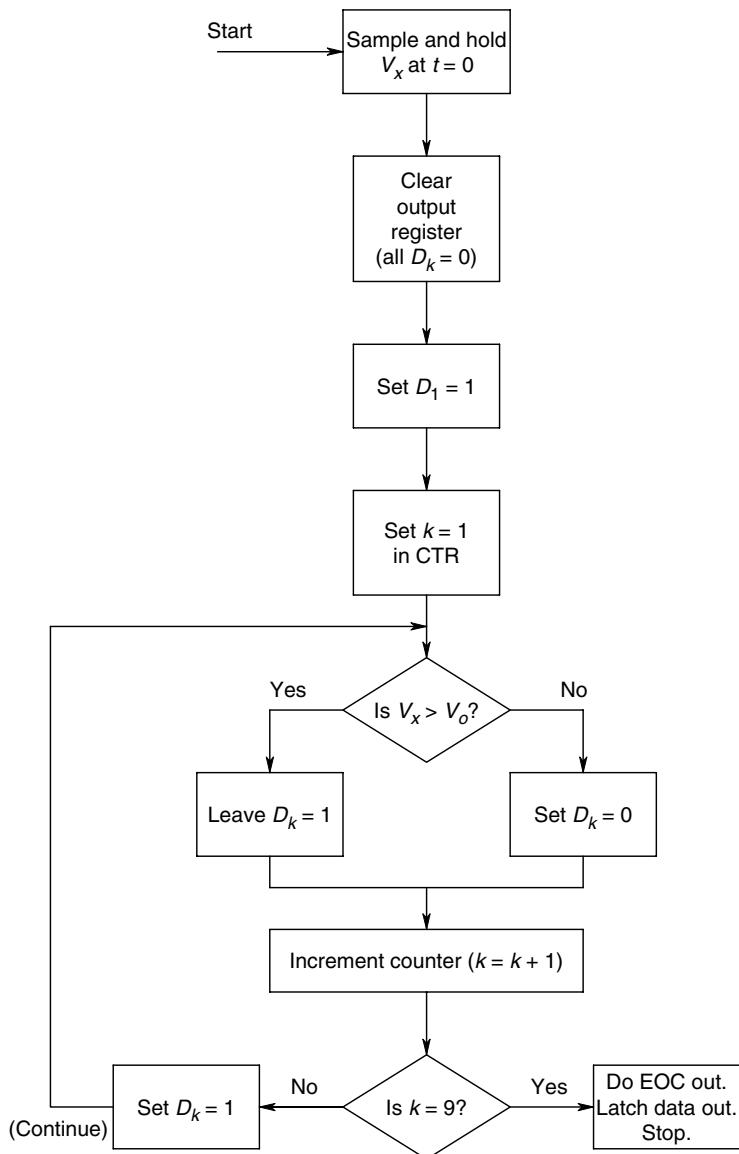
Block diagram of a successive approximation ADC. S&H is an analog sample-and-hold (or track-and-hold) circuit used to 'freeze' V_X during the A/D conversion process. (Source: Northrop, 1990)

cost and moderate conversion speed. However, they can have missing output codes, require a sample-and-hold input interface and are difficult to autozero.

A block diagram of an SA ADC is shown in Figure 9.14. Note that this type of ADC uses a DAC and an analog comparator in a logical feedback loop to control its conversion algorithm. SA ADCs are fast enough for audio frequency applications, with conversion times in the range of about 2–20 μs . As we will see, conversion time depends on the digital clock of the SA ADC and its number of bits.

The conversion cycle of an 8-bit SA ADC begins with the analog input signal being sampled and held at $t=0$. Simultaneously, the output register is cleared and all D_k are set to 0. At the next clock cycle, D_1 is set to 1, with all other $D_k=0$. This makes the DAC output $V_o=V_R(128/256)=V_R/2$. The comparator performs the operation $\text{SGN}(V_X - V_R/2)$. If $\text{SGN}(V_X - V_R/2)=1$, then D_1 is kept 1; if $\text{SGN}(V_X - V_R/2)=-1$, then D_1 is set to 0. This completes the first (MSB) cycle in the conversion process. Next, D_2 is set to 1 (D_1 remains the value found in the first cycle). The comparator output signals if $V_X > [D_1(V_R/2) + V_R/4]$. If yes, then D_2 is set to 1, if no, then $D_2=0$, completing the second bit's conversion cycle. This process continues until all N -bits are converted and then stops and signals *Data Ready*. It is easily seen that N clock cycles are required to convert V_X to an N -bit digital word. Note that the final conversion error is less than $\pm 1/2 \text{ LSB}$. The SA ADC conversion process is shown in flow chart form in Figure 9.15.

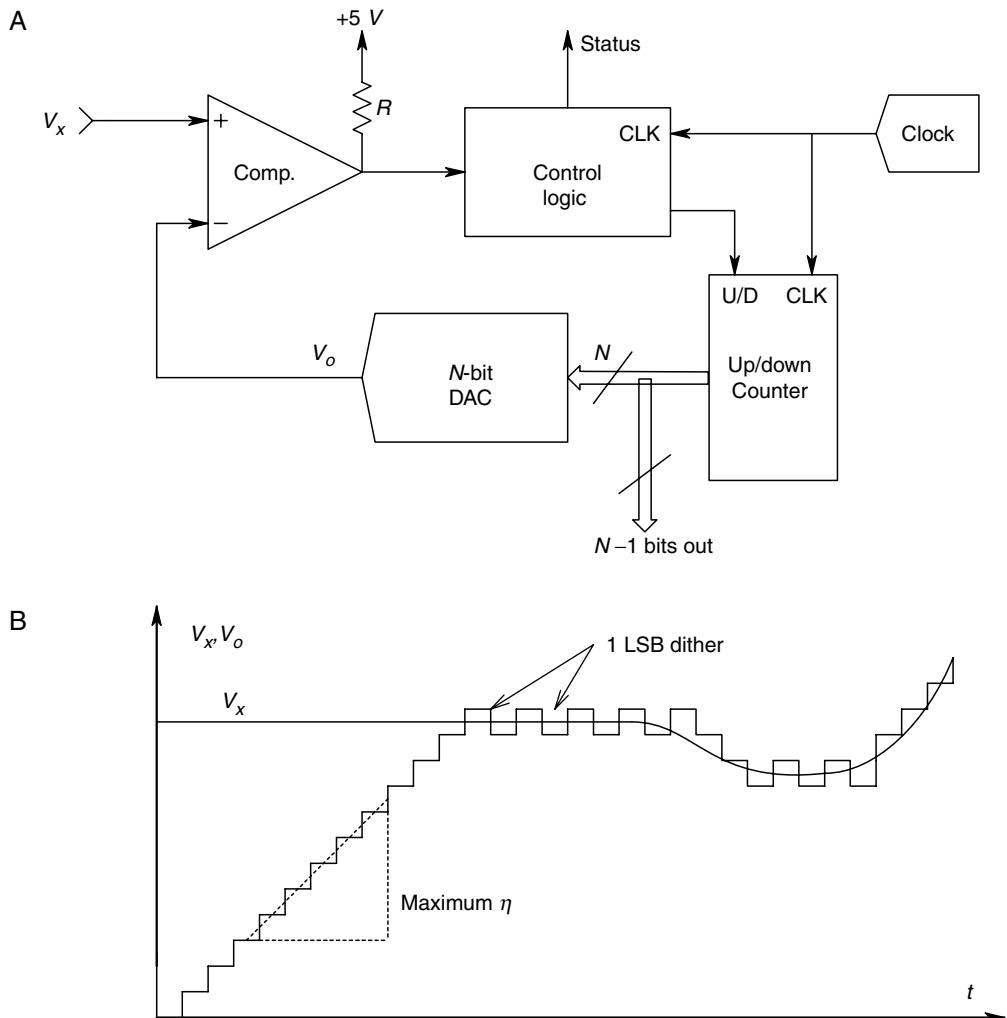
SA ADCs are designed to work at audio frequencies and higher, with their conversion cycles under computer control. Often, their output registers use tristate logic, so that an ADC can be used with other tristate SA ADCs and DACs on the same, bidirectional data bus. Note that if N , N -bit SA ADCs are on the same data bus, and their starting times are staggered one after another, an ADC output can be read into the computer at every clock cycle.

**FIGURE 9.15**

Flowchart for the operation of a typical, 8-bit, successive approximation ADC. Note: D_1 = MSB, D_8 = LSB. (Source: Northrop, 1990)

9.7.2 Tracking or Servo ADCs

The tracking ADC is a relatively slow ADC, best suited for the digitization of dc and low audio frequency measurands. Figure 9.16A illustrates the block diagram of a tracking ADC. Figure 9.16B illustrates the input voltage, $V_x(t)$, and the DAC output, $V_o(t)$ in a T-ADC, when it is first turned on. Note that the DAC output slews up to V_x at a rate set by the DAC's clock. When V_o exceeds V_x , the comparator's output goes low, which causes the counter to count down one clock cycle to bring $V_o < V_x$. Then, the comparator goes high, signaling an up count, etc. In following a very slow ac or dc V_x waveform, V_o is seen to have a ± 1 LSB limit cycle around the true value of V_x . This means that, in

**FIGURE 9.16**

A. Block diagram of a tracking (servo) ADC. B. Waveforms in a tracking ADC. (Source: Northrop, 1990)

realizing a digital output, the true LSB digit must be dropped because of the limit cycle oscillation. Thus, an 11 bit DAC is required to realize a 10-bit output T-ADC.

A T-ADC has a maximum slew rate, η , in volts per second, at which it can follow a rapidly changing $V_X(t)$ without gross error. This is:

$$\eta = \frac{V_{oMAX}}{2^N T_C} \text{ V/s} \quad (9.26)$$

As an example, let $N = 10$ bits, $V_{oMAX} = 10 \text{ V}$ and $T_C = 1 \mu\text{s}$. η is found to be $9.766 \times 10^3 \text{ V/s}$. Thus, $dV_X(t)/dt$ must be less than this η for no gross error. The T-ADC, in addition to slew rate limitations, performs poorly in the presence of high frequency noise on V_X . Thus, low-pass filtering of V_X before it is seen by the comparator improves T-ADC performance. This type of ADC is suitable for the design of low cost, digital, dc voltmeters and ammeters, etc.

9.7.3 Dual-Slope Integrating ADCs

DSI-ADCs have the advantages of high inherent accuracy (up to 22 bits output), excellent high frequency noise rejection, no requirement for a sample and hold at their inputs and no possibility of missing output codes. They are widely used in inexpensive dc digital instruments. Their conversion times are long, ranging about 10 ms – 0.333 s. They generally have an auto-zeroing mode, discussed below. Figure 9.17 illustrates a DSI-ADC which converts a unipolar, dc input. Note that a DAC is not used in this ADC design, but a high performance analog integrator is required. The control logic operates MOS analog switches, S1 and S2 (S1 at input, S2 over integrator.) The conversion cycle of a DSI-ADC begins with the control logic first clearing the counter and setting the charge on the integrator capacitor equal to zero with switch S2. Next, $V_X \geq 0$ is integrated for 2^N clock cycles. At the end of this integration time, T_1 , the integrator output, V_2 , can be expressed as:

$$V_2 = -\frac{1}{RC} \int_0^{T_1} \frac{V_X(t)}{T_1} dt \quad T_1 = -\frac{T_1}{RC} <\overline{V_X}> T_1 \quad (9.27)$$

where $T_1 = 2^N T_C$. Switch S1 now connects the integrator input to $-V_R$. V_2 now ramps linearly toward 0 V with positive slope. The counter is made to count clock cycles from T_1 to T_F , when V_2 reaches 0 V. At T_F , the counter count, M , is latched into the output register. At T_F , we can write:

$$V_2 = 0 = -(T_1/RC) <\overline{V_X}> T_i + T_2 VR/RC \quad (9.28)$$

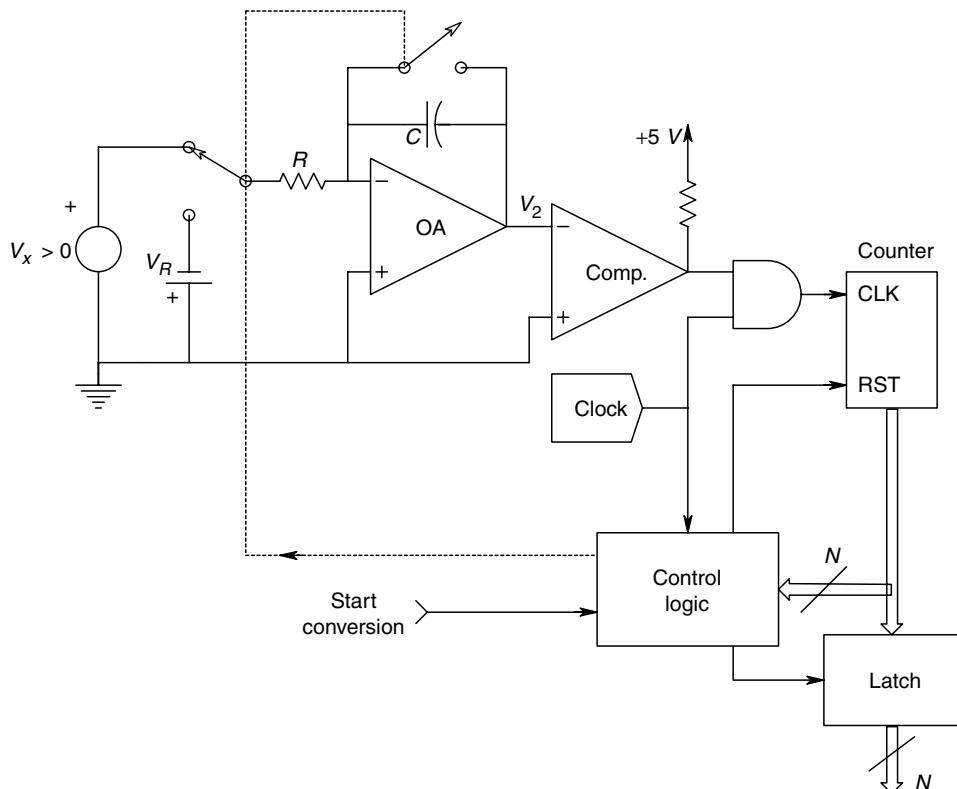


FIGURE 9.17

Diagram of a unipolar, dual-slope, integrating ADC. (Source: Northrop, 1990)

where $T_2 = T_F - T_1$. However,

$$T_2 = M T_C \quad (9.29)$$

So, we can write:

$$M T_C V_R / RC = 2^N T_C \langle \overline{V_X} \rangle_{T_1} / RC \quad (9.30)$$

which leads to:

$$M = \langle \overline{V_X} \rangle_{T_1} 2^N / V_R \quad (9.31)$$

Thus, we see that the numerical count, M , is proportional to V_X averaged over T_1 seconds.

Calibration of the DSI-ADC is seen to be independent of the values of the clock period, R , and C . It does require a temperature stabilized V_R , and an integrator op-amp and comparator with low offset voltage drift. In addition, the integrator op-amp must have negligible dc bias current. DSI-ADCs are widely used in digital multimeters. They are available as LSI ICs and generally have an autozeroing mode. In one strategy, the output count, M_o , measured with the input shorted to ground ($V_X = 0$), is read every other conversion cycle and then subtracted from M obtained with V_X connected. Another autozeroing strategy is analog based, in which a special autozero capacitor is charged to the net offset voltages of the buffer amplifier, integrator op-amp and the comparator during a special autozero cycle with inputs shorted to ground. The autozero capacitor, charged to the net offset voltage, is then switched in series with the input during the measurement cycles to effectively reduce the net ADC offset voltage to $< 10 \mu\text{V}$ (this autozero architecture is used in the Intersil ICL7106/7107 DSI-ADC).

It can be shown that if T_1 is made an integer number, K , of power line frequency periods (i.e. $T_1 = 2^N T_C = K/60$ seconds), then power line hum contaminating V_X can be rejected by as much as 70 dB. Other modifications of the DSI-ADC include a connection for bipolar input signals, which gives an offset binary output code, and charge balancing ADCs in which a reference current source switched into the integrator's summing junction replaces V_R in the DSI-ADC previously described [Northrop, 1990].

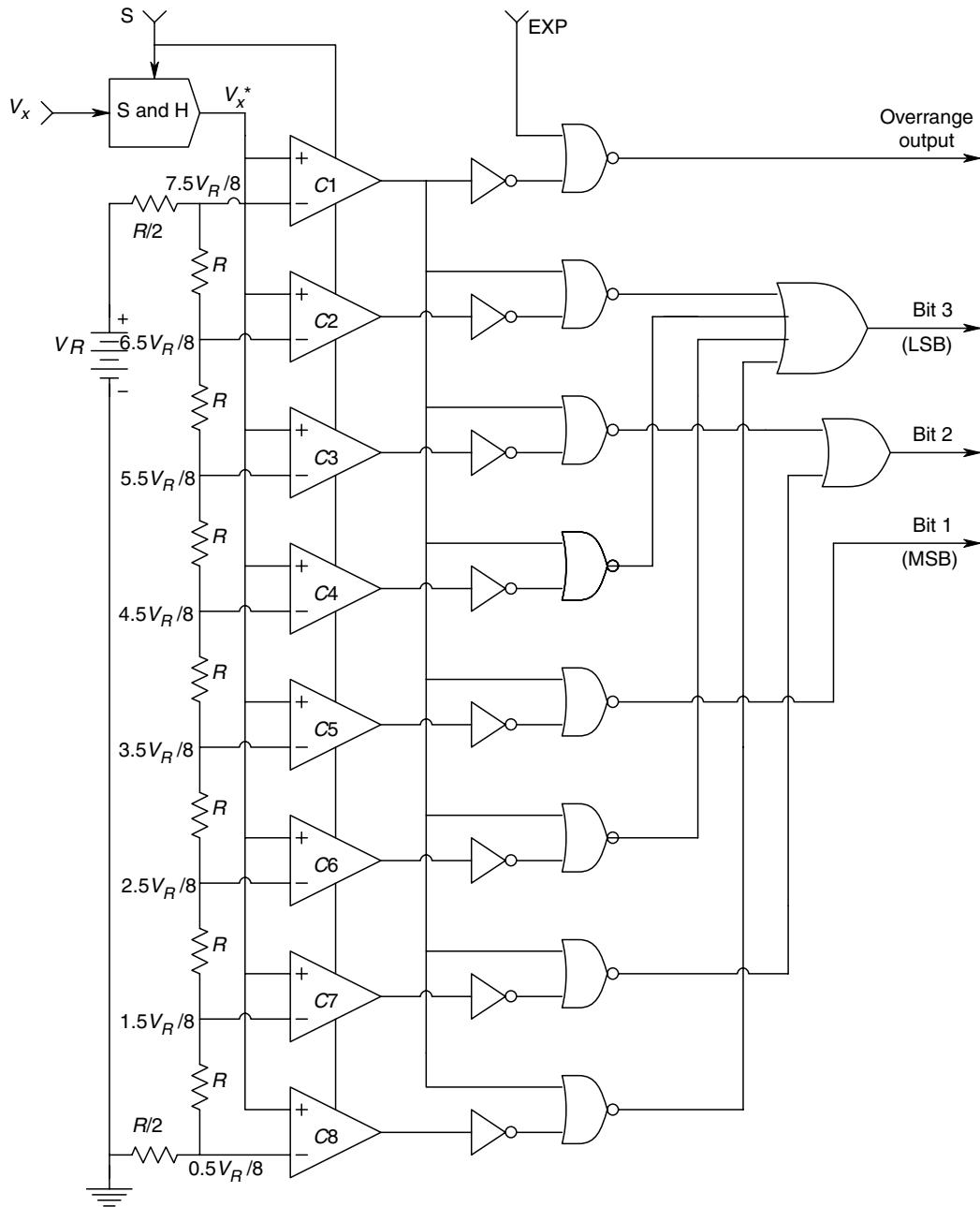
Some commercially available, dual-slope ADCs include the Analog Devices' 22-bit AD1175K; the Intersil ICL7106/7107, 3½ digit, autozeroing ADC for LCD display digital multimeters; the Intersil ICL7129, 4½ digit DMM DSI-ADC; the Teledyne TSC7135, 4½ digit DMM DSI-ADC; the Teledyne TSC808 auto-ranging, AC/DC, 3½ digit DMM IC.

9.7.4 Flash (Parallel) ADCs

Flash ADCs (FADCs) are the cornerstone of modern, ultra high speed, digitizing front ends of certain digital storage oscilloscopes (DSOs), such as those manufactured by LeCroy, Nicolet, Hewlett-Packard and Tektronix. Eight-bit digitizing rates in excess of 1 gigasamples/s (10^9Sa/s) are currently available for specialized DSOs. The digitized outputs of such flash ADCs must be buffered by ECL or proprietary logic registers, which store and download the digital data rates acceptable by computer bus structures, logic and RAM memories.

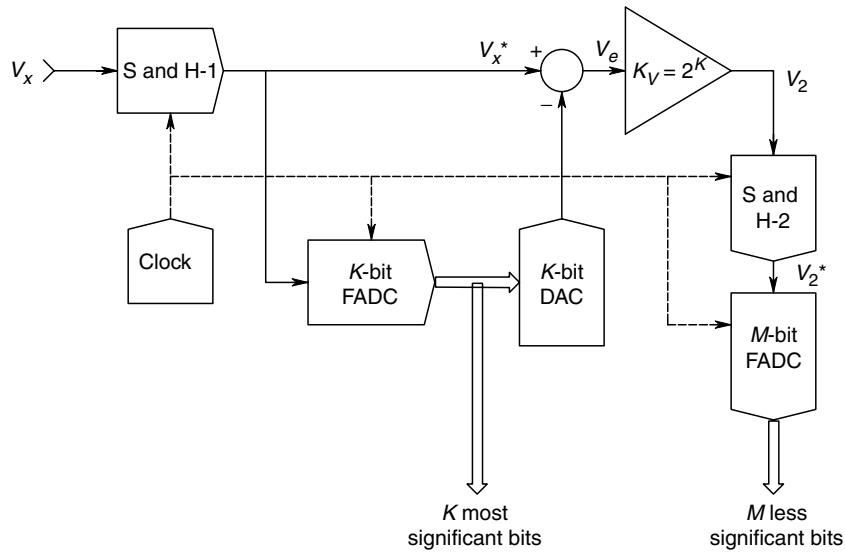
An FADC can be realized as a hybrid or LSI IC. From Figure 9.18 we see that an FADC is composed of four subsystems:

1. An analog track-and-hold or sample-and-hold circuit which is used to sample and 'freeze' the input voltage, and present that sample to the inputs of the high speed, analog, amplitude comparators

**FIGURE 9.18**

Schematic of a 3-bit, flash ADC with binary output. (Source: Northrop, 1990)

2. A stable, dc reference voltage source and a voltage divider ladder that supplies the switching reference voltages to the analog comparators
3. 2^N , high speed, analog voltage comparators having negligible dc offset voltages
4. A combinational logic circuit composed of OR, NOR and inverter gates that accepts the logic level outputs of the 2^N comparators, and that generates a parallel digital word at the FADC output.

**FIGURE 9.19**

Block diagram of a two step, flash ADC. $K + M = N$ output bits. (Source: Northrop, 1990)

Practical considerations in the design of integrated circuit FADCs having the architecture of Figure 9.18 limit the number of bits to eight. Problems exist with the input capacitances of the 28 comparators loading the track and hold circuit, signal propagation delays and maintaining low dc offset voltages on the 256 comparators.

A design for a FADC having a 12 bit output is shown in Figure 9.19. This system uses two 6-bit FADCs, a very fast 6-bit DAC, and two sample-and-hold circuits. Conversion speed of this system will be slower than either 6-bit FADCs alone because of the conversion lag and settling time of the DAC and the $K_V = 2^6$ amplifier. A conversion cycle of the 12-bit, two stage FADC operates as follows. S&H-1 samples the analog input signal, giving an analog output V_X . The V_X is flash converted by the first 6-bit FADC and simultaneously reconverted to analog form by the 6-bit DAC. The DAC output, V_1 , is subtracted from V_X , giving the analog quantization error, V_E , of the first ADC. V_E is amplified by a factor of $2^6 = 64$, again sampled and held, then converted by the second FADC to give the 6 least significant bit output. In mathematical terms:

$$V_1 = V_{X(MAX)} \sum_{n=1}^6 \frac{D_n}{2^n} \quad (9.32)$$

$$V_2 = 2^6 V_E = 2^6 (V_X - V_1) = 2^6 \left(V_X - V_{X(MAX)} \sum_{n=1}^6 \frac{D_n}{2^n} \right) \quad (9.33)$$

In general, V_E can be as large as $V_{X(MAX)}/2^6$, hence the second 6-bit converter generates a binary code on the quantized remnant, V_E . Internally, the two, 6-bit FADCs each requires $(2^6 - 1)$ comparators and thus, is far less complex than even an 8-bit FADC. It can be shown that the total coded output of this system can be written as:

$$V_X = V_{X(MAX)} \sum_{k=1}^{12} \frac{D_k}{2^k} \quad (9.34)$$

In general, FADCs are expensive, specialized ADCs whose principal application is the digitization of analog signals with frequencies in the range 10 MHz – >1 GHz. They can, of course, operate at slower sampling rates and are widely used in the front ends of digital storage oscilloscopes which operate on signals having frequencies from mHz to 500 MHz. Eight-bit resolution is typical.

Some examples of FADC ICs include the Burr-Brown ADC600, a 12-bit, 10 mega-samples/s (MSPS) circuit, which uses the two stage architecture described above, the Analog Devices AD770 8-bit, 200 MSPS IC and the Sony CXA1096P 8-bit, 20 MSPS converter. The state-of-the-art in flash converters probably lies close to the proprietary digitizing system offered by LeCroy. The LeCroy Model 6880B waveform digitizer system samples at a fixed 1.35 GSPS, giving either an 8-bit or 11-bit output, stored in a 10,016 word buffer memory holding 7.42 μ s of sampled data.

9.7.5 Dynamic Range, Floating Point ADCs

Figure 9.20 illustrates a block diagram of the Micro Networks model MN-5420, DRFP-ADC. This ADC has accuracy better than 2 ppm and can carry out 3.2×10^5 conversions/s. It has a 16-bit output consisting of a 12-bit mantissa and a 4-bit exponent. The 4-bit FADC codes the exponent and its output sets the gain of a programmable gain amplifier (PGA) having gains $K_V = 2^k$, $k = 1, 2, 4, \dots, 128, 256$. The settled amplifier output goes through a track-and-hold circuit, and is converted by a fast, 1 μ s, 12-bit, SA ADC. When the analog input signal is small, the gain of the PGA is high. For example, for $V_X < 19.5$ mV, the gain is 256. The value of a mantissa LSB at this gain is 9.5 μ V. As designed, the maximum input voltage of the MN-5420 DRFP-ADC is ± 5 V. Table 9.3 gives this ADC's gain switching points and the mantissa's LSB values. When V_X goes negative,

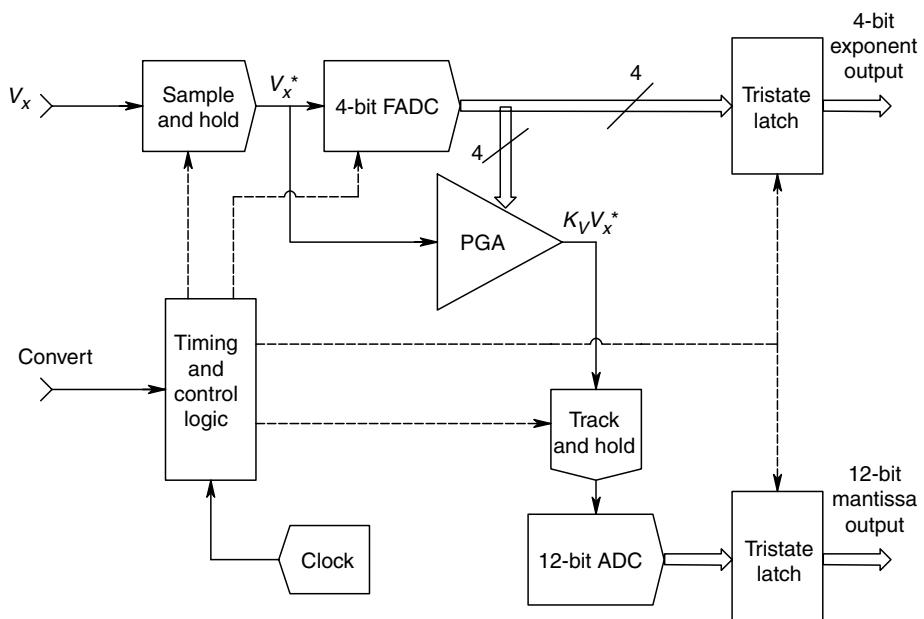


FIGURE 9.20

Block diagram of the MN-5420, dynamic range, floating point ADC. The programmed gain amplifier (PGA) has an autozeroing feature (not shown).

TABLE 9.3

MN-5420, dynamic range, floating point ADC PGA switching points and exponent coding for $V_X > 0$

PGA Switching Voltage, V_X	PGA Gain	Exponent	LSB Voltage
$+5 \leq V_X \leq 2.5$ V	1	1 0 0 0	2.44 mV
$2.5 < V_X \leq 1.25$ V	2	0 1 1 1	1.22 mV
$1.25 < V_X \leq 0.625$ V	4	0 1 1 0	610 μ V
$0.625 < V_X \leq 0.3125$ V	8	0 1 0 1	305 μ V
$0.3125 < V_X \leq 0.15625$ V	16	0 1 0 0	153 μ V
$0.15625 < V_X \leq 0.078125$ V	32	0 0 1 1	76 μ V
$0.078125 < V_X \leq 0.039063$ V	64	0 0 1 0	38 μ V
$0.039063 < V_X \leq 0.019531$ V	128	0 0 0 1	19 μ V
$0.019531 < V_X \leq 0$ V	256	0 0 0 0	9.5 μ V

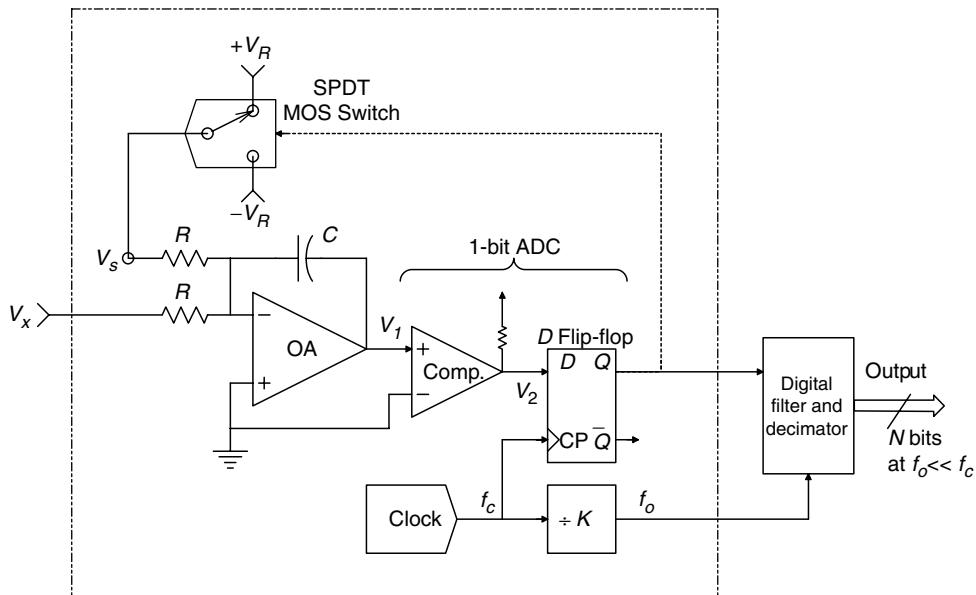
Note: V_X range is ± 5 V

the exponent coding is the same as for positive V_X . However, the MSB of the 12-bit mantissa goes from 0 to 1.

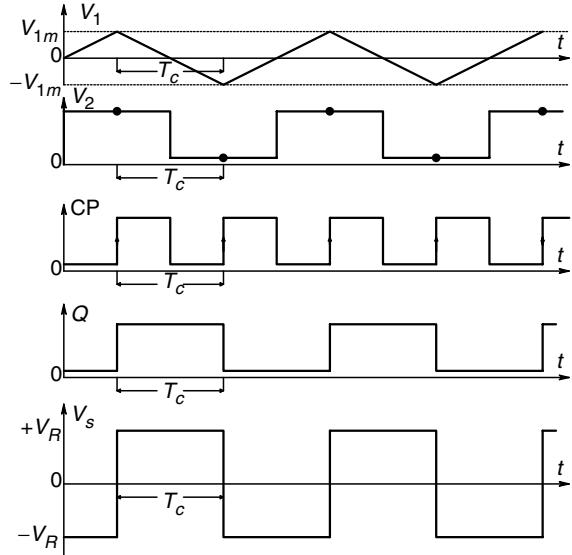
The MN-5420 DRFP-ADC design is an ingenious approach to high resolution, high speed analog-to-digital conversion. The heart of the design is the precision, autozeroing, PGA. The overall ADC speed is limited by the speed at which gains can be switched, and the speed of the 12-bit SA ADC. Note from Table 9.3, that the resolution of the DRFP ADC is 1.9 ppm on its most sensitive auto-range.

9.7.6 Delta-Sigma ADCs

Delta-Sigma ADCs (Δ - Σ ADC) are also called *oversampled* or *noise-shaping converters*. Oversampling ADCs are based on the principle of being able to trade-off accuracy in time for accuracy in amplitude. Figure 9.21 illustrates the circuit of a first order, Δ - Σ ADC.

**FIGURE 9.21**

Circuit of a first order Δ - Σ ADC.

**FIGURE 9.22**

Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = 0$.

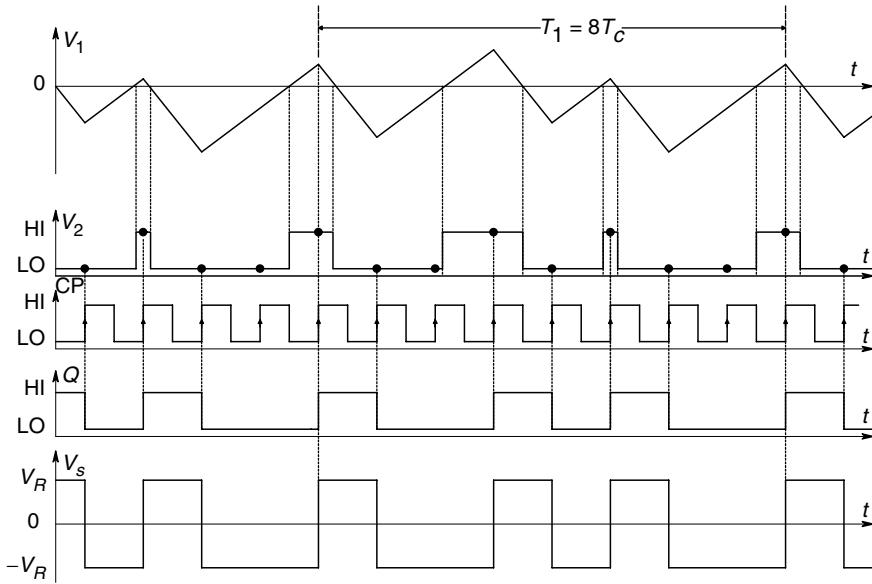
The circuit in the box is a first order Δ - Σ modulator. The digital filter and decimator produces the N -bit, digital output.

We will first analyze how the Δ - Σ modulator operates in the time domain. First, let us set $V_x \rightarrow 0$. Let the MOS switch apply $-V_R$ to the integrator so that its output, V_1 , goes positive. The comparator output, V_2 , applied to the D input of the D flip-flop (DFF) will be HI. At the rising edge of the next clock pulse, HI is seen at the Q output of the DFF. This HI causes the MOS switch to switch the integrator input to $+V_R$, so that the integrator output V_1 begins to ramp down. When V_1 goes negative, the comparator output V_2 goes LO. Again, when the clock pulse goes HI, the LO at D is sent to the DFF Q output. This LO again sets the MOS switch to $-V_R$, causing V_2 to ramp positively until it goes positive, setting V_2 to HI, etc. Figure 9.22 shows the Δ - Σ modulator's waveforms for $V_x = 0$. Q is the DFF output which is the comparator output, V_2 , latched in when the CP goes high. The peak-to-peak height of the steady state triangle wave oscillation seen at V_1 can easily be shown to be equal to:

$$2V_{1m} = \frac{V_R T_c}{RC} \quad (9.35)$$

where T_c is the Δ - Σ ADC's clock period, and the other parameters are given in the figure. The average value of $V_s = 0$.

Now let us set $V_x = +V_R/4$ and examine the Δ - Σ ADC's waveforms (Figure 9.23). Let the system start with $V_1 = 0$ at $t = 0$, and begin to integrate V_x and $V_s = +V_R$. Initially, the integrator output V_1 ramps negative as $v_1(t) = -[5V_R/(4RC)]t$. At the first positive going clock wave, the LO comparator output, V_2 , is latched to the DFF's Q output, setting it LO. When Q goes LO, the MOS switch makes $V_s = -V_R$. Now the integrator output ramps upward with slope $+3V_R/(4RC)$. When V_1 goes positive, the second positive clock transition latches the HI V_2 to Q , setting V_s to $+V_R$. Now, the integrator output, V_1 , ramps down again with slope $-5V_R/(4RC)$. When $V_1 < 0$, V_2 goes LO. The third positive clock transition latches this LO to the Q output of the DFF. As before, this sets V_s to $-V_R$ and the integrator output again ramps positive, and the cycle repeats itself. Note that there is a steady state (SS) oscillation in V_1 , V_2 , Q and V_s with a period of eight clock periods. The SS duty cycle of Q is 37.5% and $\overline{V_s} = -1/4$.

**FIGURE 9.23**

Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = V_R/4$.

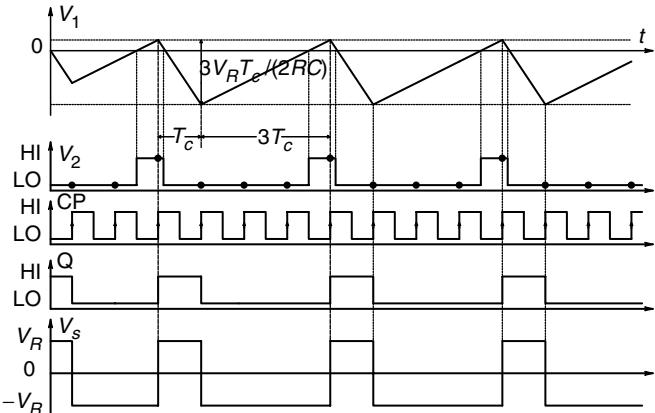
It is also of interest to examine the Δ - Σ ADC's waveforms when $V_x = +V_R/2$. Now, let the system start with $V_1 = 0$ at $t = 0$ and begin to integrate V_x with $V_s = +V_R$. Assuming V_x is constant, V_1 ramps negative as $v_1(t) = -3V_Rt/(2RC)$, causing the comparator output, V_2 , to be LO. At the first positive going clock pulse, the DFF latches the low D input to Q , causing the MOS switch to make $V_s = -V_R$. Now the net input to the integrator is $-V_R/2$ and the integrator output begins to go positive with slope $+V_R/(2RC)$, as shown in Figure 9.24. When V_1 first goes positive, V_2 goes HI. At the third clock HI transition, the DFF output goes HI, setting the MOS switch to $-V_R$. Now the net input to the integrator is $+3V_R/2$ and V_1 begins to ramp down again with slope $-3V_R/(2RC)$, and the ADC process repeats itself indefinitely until V_x changes. From the waveforms in this figure, where $V_x = V_R/2$, we see that the duty cycle of Q is 25% and $\bar{V}_s = -1/2$. From the three sets of V_s waveforms, we see that in general, $V_x = -V_R V_s$.

Now let us consider the operation of first order, Δ - Σ ADC. If the number of ones in the output data stream (i.e. the DFF's Q output) is counted over a sufficient number of samples (clock cycles), the counter's digital output will represent the digital value of the analog input, V_x . Obviously, this method of averaging will only work for dc or low frequency V_x . In addition, at least 2^N clock cycles must be counted in order to obtain N -bit effective resolution. In the Δ - Σ ADC, the modulator's DFF Q output is the input to the block labelled 'digital filter and decimator'. The digital low-pass filter precedes the decimator. The digital LPF serves two functions:

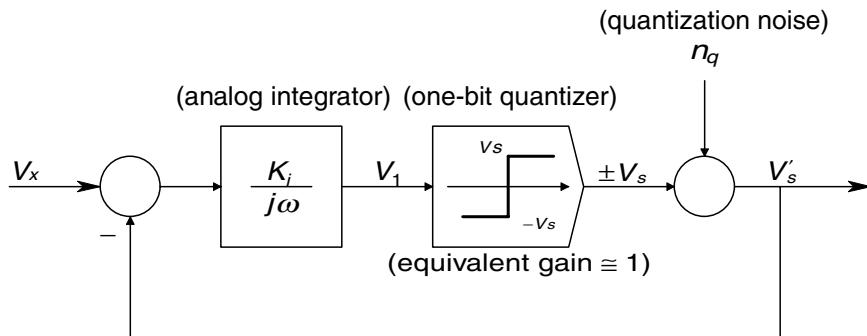
1. It acts as an anti-aliasing filter for the final sampling rate, f_o
2. It filters out the higher frequency noise produced by the Δ - Σ modulator

Quoting Analog Devices' AN-283 (*Sigma-Delta ADCs and DACs*):

'The final data rate reduction is performed by digitally resampling the filtered output using a process called decimation. The sampling rate of the input signal $x(n)$ is at a rate

**FIGURE 9.24**

Waveforms in the Δ - Σ ADC of Figure 9.21 when $V_x = +V_R/2$.

**FIGURE 9.25**

A frequency domain block diagram approximation describing a first order Δ - Σ ADC.

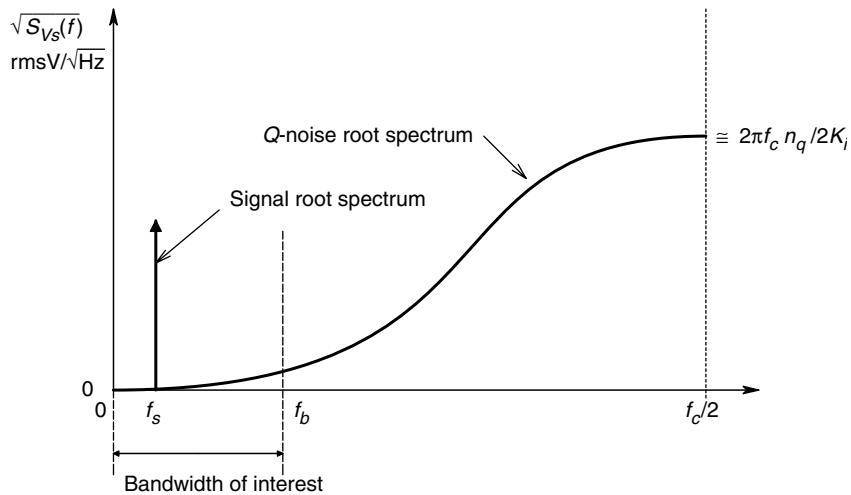
which is reduced by a factor of 4. The signal is resampled at the lower rate (the decimation rate), $s(n)$ [f_o]. Decimation can also be viewed as the method by which the redundant signal information introduced by the oversampling process is removed.

In sigma-delta ADCs it is quite common to combine the decimation function with the digital filtering function. This results in an increase in computational efficiency if done correctly'.

Figure 9.25 illustrates a heuristic, frequency domain block diagram approximation for a first order Δ - Σ ADC. To appreciate what happens in the system in the frequency domain, we can find the transfer function for the analog input, V_x , and the internal quantization noise, n_q (see the section below). Assuming a linear system, we find:

$$V'_s = \frac{V_x K_i}{j\omega + K_i} + \frac{j\omega n_q}{j\omega + K_i} \quad (9.36)$$

Thus, the 1-bit quantization noise in the output is boosted at high frequencies, so $V'_s \cong j\omega n_q / K_i$. The signal component in V'_s rolls off at -6 dB/octave above $\omega = K_i \text{ r/s}$. At low frequencies, the noise in V'_s is negligible and $V'_s \cong V_x$. Since the Δ - Σ modulator is really a sampled system at frequency f_c , the noise in V_s and Q has the root power density spectrum shown in Figure 9.26. Note that the broadband quantization noise is concentrated at the upper end of the root spectrum. The range from $0 \leq f \leq f_b$ is

**FIGURE 9.26**

Root power density spectra of signal and noise in a first order Δ - Σ ADC.

relatively noise free. Operation in this range is achieved by first having the FIR LPF operate on the data stream from Q at clock rate f_c , then decimation, so that $f_b = f_o/2$, the Nyquist frequency of the decimation frequency. Lastly, the counter counts the decimated data for $2^N f_o$ clock periods.

9.7.7 Data Acquisition Cards for PCs

It is evident that PCs and laptop computers are central to most modern instrumentation and measurement systems. Consequently, many manufacturers make plug-in interface boards that allow one or more channels of external analog data to be sampled and digitized. In fact, the ubiquity of the modern PC allows it to serve as a digital storage oscilloscope—a single specialized card is required. PC Instruments of Akron, Ohio make a series of PCI-42X and PCI-43X digital oscilloscope cards. These cards, depending on the model, have one or two 8-bit channels and bandwidths ranging 100–300 MHz. In the real time mode, maximum resolution is 80 ns/sample. In the equivalent sampling mode, maximum resolution is 5 ps/Sa, or an equivalent sampling rate of 200 gigasamples/s. Gage Applied Sciences of Montreal offer a line of super fast, PCI bus waveform digitizers and virtual oscilloscopes. Their Model 85G CompuScope is their fastest, sampling 8 bits at 5 gigasamples/s. The Gage Model 2125/ETS CompuScope does real time sampling at 250 megasamples/s in one channel and has 2 gigasamples/s equivalent time sampling on two 8-bit channels. These plug-in digital oscilloscope cards operate as the input elements for sophisticated virtual instruments. All the normal knobs and switches found on the front of a stand alone DSO are displayed on the CRT, and are mouse operated.

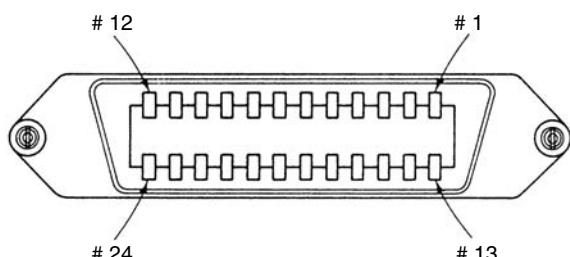
Many manufacturers make A/D boards for PCs that are not part of a virtual DSO system. For example, Delphi Engineering Group makes an 8-bit, 1 GHz PMC ADC Digitizer Board. This board has an analog –3 dB bandwidth of 1 GHz. Its sampling rate is up to 1 gigasamples/s. SuperLogics offers an extensive line of PCI data boards having 12-, 14- and 16-bit resolution on two channels, also 32 bit digital I/O. Sampling rates range from 150 kilosamples/s to 1.25 megasamples/s, depending on the board. Many other companies, including CyberResearch®, National Instruments™, Keithley Instru-

ments, etc. offer extensive lines of analog input boards. One has the option of specifying the number of input channels, single ended or differential input, the number of bits, the size of the buffer memory on board and of course, the sampling rate. The user designing an I&M system has the option of writing their own data management software, but most board vendors also sell software packages for their boards. Most notable is the LabVIEW® software sold by National Instruments.

9.8 The IEEE-488 Instrumentation Bus (GPIB)

In computer nomenclature, a 'bus' is 'a collection of unbroken signal lines that interconnect computer modules (the connections are made by taps on the lines)' [Stone, 1982]. In general, a bus contains data lines, control and signalling lines and in some cases, address lines. There are many *parallel bus* architectures and protocols which have been developed since the advent of the microcomputer. These include, but are not limited to, the Intel *Multibus* (IEEE-796 bus), the *VME bus*, the *Centronics printer bus*, the *DEC Unibus*, the *LSI-11 bus*, the *IBM PC bus* (IBM PC/XT/AT), the *Macintosh II NuBus*, the *MicroVAX Q-bus*, the *IBM PS2 Micro-Channel Architecture (MCA) bus*, the *Intel iSBX bus*, the *ISA* and *EISA PC buses*, and of course, the *IEEE-488 Instrumentation Bus (GPIB)* which we describe below.

The IEEE-488 bus was developed by Hewlett-Packard in the early 1970s as a standard, 8-bit, bidirectional, asynchronous bus (HP-IB) to enable a number of HP-IB compatible instruments to communicate with a controlling computer and with one another. Not only can measured data be sent to the host computer for storage and processing, but in certain GPIB compatible instruments, the computer can be used to set the instrument's front panel controls and to control the measurements. The original IEEE-488 standard was defined by the IEEE Standards Committee in 1975, was revised in 1978 and again in 1987 as IEEE Std. 488.2-1987. The IEEE-488.2 standard is also called the



Pin Signal	Pin Signal
1. Data 1	13. Data 5
2. Data 2	14. Data 6
3. Data 3	15. Data 7
4. Data 4	16. Data 8
5. EOI	17. REN
6. DAV	18. Gnd
7. NRFD	19. Gnd
8. NDAC	20. Gnd
9. IFC	21. Gnd
10. SRQ	22. Gnd
11. ATN	23. Gnd
	24. Logic ground

FIGURE 9.27

Standard, IEEE-488 instrumentation bus stackable connector and its pin assignments. (Source: Helfrick and Cooper, 1990)

Standard Commands for Programmable Instruments (SCPI). Its latest version is SCPI-1999. Its text is available from the IEEE as a 3.5 MB pdf file from the site (www.scpiconsortium.org/scpistandard.htm).

The GPIB is unique in that it uses unique, stackable (hermaphrodite) connectors on its cables, as shown in Figure 9.27. Many semiconductor manufacturers offer LSI IC interfaces for the GPIB. Some of these include the three Intel i829X series ICs (the i8291 talker/listener, the i8292 GPIB controller and the i2893 bus transceiver-driver), the Texas Instruments TMS9914, the Motorola 68488 and the Signetics HEF4738.

9.8.1 The GPIB Bus Structure

The GPIB consists of eight, tristate, bidirectional data lines and eight control lines that select, deselect, and otherwise coordinate the asynchronous communications between the host computer and satellite instruments. The maximum practical data transmission rate in most GPIB systems is about 250 kB/s, equivalent to about 2 Mbits/s. Practical considerations on loading a typical IEEE-488 bus limit the number of devices on the bus to about 15, all which should be located within 3 meters or so of the host computer. All lines on the GPIB use a complimentary TTL logic protocol. An active LOW is a voltage less than 0.8 V, and may be treated as a logical TRUE or '1'. An active HI is a voltage greater than 2.5 V and is a logic FALSE or '0'. In the tristate, Hi-Z condition, pull-up resistors tied to +5V make the data lines assume a high impedance, HI voltage. All commands and most data on the 8-bit data I/O lines are generally sent using the 7-bit ASCII code set, in which case, the eighth bit is used for parity, or is unused.

The eight control lines, their acronyms and functions are described below: They may be subdivided into *Handshake Lines* and *Interface Management Lines*.

- Handshake Lines
 1. **DAV** (data valid): When a selected device on the GPIB supplies an 8-bit word to the data lines (i.e. is a talker), DAV is set LOW (or TRUE) to indicate to all on the bus that the data byte on the bus is ready to be read to a listener device. DAV is one of the three 'handshaking' lines that control data transmission on the GPIB.
 2. **NRFD** (not ready for data): This second handshaking line is pulled LOW by a selected listener to indicate that it is ready to accept data. (NRFD would be better called 'RFD').
 3. **NDAC** (not data accepted): This third handshaking line is set LOW by the selected listener device when the data byte transmitted has been accepted and, that new data may now be supplied.
- Interface Management Lines
 1. **ATN** (attention): This line is pulled LOW by the bus controller (computer) to signal that it is sending a command. It drives ATN HI to signal that a talker can send it data messages.
 2. **IFC** (interface clear): The GPIB controller drives this line LOW to reset the status of all other devices on the GPIB, and to become Controller-In-Charge (CIC).
 3. **SRQ** (service request): This is the GPIB equivalent of an interrupt line. Any device on the bus (except of course, the controller) can pull the SRQ line LOW, signaling to the controller that it requires 'service'. Devices can be programmed to signal

SRQ for various reasons, such as being overloaded, out of paper, having completed an internal task such as computing an FFT spectrum, etc. The controller, upon sensing SRQ LOW, must poll the devices on the GPIB to determine which one sent the SRQ, and then take appropriate control action (service).

4. **REN** (remote enable): This controller output line is set LOW to allow the controller to take over front panel controls of a selected instrument on the GPIB (if that instrument has the capability of having its front panel settings taken over by the controller).
5. **EOI** (end or identify): This line has two functions—a talker uses EOI LOW to mark the end of a multibyte data message. EOI LOW is also used by the bus controller to initiate a parallel poll. When the controller sets both EOI and ATN LOW, a parallel poll causing devices configured for a parallel poll to present status bits on the data bus, which are read by the controller.

9.8.2 GPIB Operation

As we have already seen, the 8-bit data bus can carry data or commands. The controller pulls the ATN line LOW, to signal that it is sending commands. There are four types of command—addressed, listen, talk and universal. Commands are generally represented by 7-bit ASCII characters on the bus. The type of command being sent is signalled by bits 5, 6 and 7 on the data bus. If bits 5, 6 and 7 are HI voltages (logic 0), then the command is an addressed command. If bit 5 is LOW, and 6 and 7 are HI, then the command is universal. If bit 6 is LOW and 7 is HI, then the command is a listen command, and if bit 6 is HI and 7 is LOW, then the command is a talk command. If both bits 6 and 7 are LOW, the command byte is a *secondary command*. Secondary commands are used for sending secondary addresses or setting up devices for a parallel poll.

As suggested by their name, *universal commands* affect all devices on the GPIB. There are five universal commands:

1. **LLO** (local lockout): LLO is represented by 11 h on the bus. It is used to disable the front panel controls of all devices on the bus. Under this command, the controller will have complete control of instrument settings, etc.
2. **DCL** (device clear): DCL resets all devices on the GPIB. It is 14 h on the data bus.
3. **PPU** (parallel poll unconfigure): A PPU command (15 h on the data bus) resets the parallel poll responses of all devices on the GPIB, allowing new parallel poll responses to be specified.
4. **SPE** (serial poll enable): The Controller responds to an SRQ by sending out a SPE (18 h) on the data bus. The SPE causes each device on the GPIB to prepare a single status word to be sent back to the controller when it is sequentially interrogated by the controller.

When the controller finds the device that sent the SRQ, it sends out a SPD command. SPD (serial poll disable) is sent out by the controller (19 h) on the data bus following the SRQ identification. The SPD command resets the buses of all the devices to the normal mode following the SPE, so that the SRQ problem can be serviced.

Addressed commands are vectored to those certain devices that have been sent listen commands. There are five addressed commands in the GPIB protocol and they, too, are

represented by hex words on the data bus.

1. GTL (go to local) (01 h) cancels the Universal CommandLLO, and restores local, front panel control to addressed devices.
2. SDC (selected device clear) (04 h) clears those devices that have received listen commands.
3. PPC (parallel port configure) (05 h) command is sent to a specific device to tell it which data bit it will use to give its status. A secondary word is sent following the addressed PPC command to specify which data line will have the status signal on it, and whether a 1 or a 0 will be used to signal that the device needs service.
4. GET (group trigger) (08 h) synchronizes the operation of several instruments on the GPIB, such as starting a group of related measurements.
5. TCT (take control) (09 h) enables the main controller to pass control of the GPIB to a secondary controller. The device which is to become the new controller is first sent a listen command, followed by TCT.

Figure 9.28 summarizes the IEEE-488 bus command codes. The interested reader can explore the GPIB in greater depth in the texts by Stone (1982), Tompkins and Webster

(Sent and received with ATN = 1)

The table is organized into several groups:

- Addressed command group (ACG):** Columns 1-5.
- Universal command group (UCG):** Column 6.
- Listen address group (LAG):** Column 7.
- Talk address group (TAG):** Column 8.
- Secondary command group (SCG):** Columns 9-16.

Annotations in the table:

- ① MSG = interface message**
- ② b₁ = DI01 . . . b₇ = DI07**
- ③ Requires secondary command**
- ④ Dense subset (columns 2 through 5)**

FIGURE 9.28

Command codes for the IEEE-488 (1978) instrumentation bus. (Source: Tompkins and Webster, 1988)

(1988), Helfrick and Cooper (1990) and the National Instruments (1992) *Databook on IEEE-488 and VXIbus Control, Data Acquisition, and Analysis*.

Many manufacturers of interface cards for PCs and MACs offer IEEE-488.2 compatible cards and controlling software. Some of the newer GPIB systems have hardware and software that permit one to make a very fast and flexible measurement control system. For example, the National Instruments' NB-GPIB interface board for the Macintosh II has the latest IEEE-488.2 functions, and can read and write data at a sustained 800 kB/s. The National Instruments' LabVIEW 2 software for the Mac II is a powerful, graphical programming language that permits the creation of virtual instruments on the monitor whose settings can be altered by mouse and sent to the actual instrument by GPIB. The NB-GPIB card is compatible with NuBus DMA operations and the RTSI bus. The National Instruments' GPIB-SE card for the Mac SE is another advanced GPIB card that claims data bus transfer rates up to 1 MB/s. This card, too, is compatible with the LabVIEW software. Metrabyte offers IBM PC compatible IEEE-488 cards and handling software, the MBC-488 and the IE-488 cards. These cards support up to 15 devices on the bus and carry out DMA data transfers at 450 kB/s.

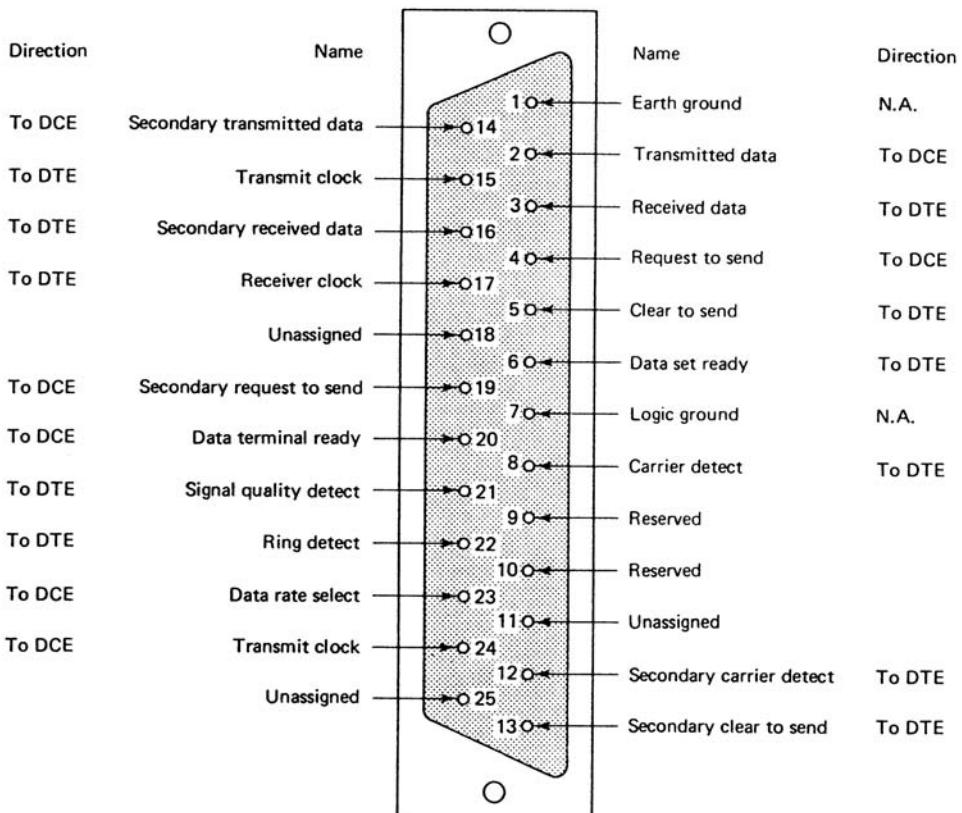
9.9 Serial Data Communications Links

While the GPIB is eminently suited for managing a large measurement system, there is often need to couple the output of a single instrument to a computer interface over long distances, using few electrical conductors. Serial data transmission has well-established protocols, and predates the GPIB by at least ten years. The RS-232 serial interface was originally developed in the early 1960s to send data to a CRT terminal or teletypewriter over telephone wires. Its low maximum bit transfer rates now makes the RS-232 serial interface generally obsolete for the control of the modern equivalents of terminals, printers and instruments. The RS-232 interface currently finds use in coupling mice, trackballs, joysticks and other slow input devices to computers, as well as computers to certain plotters and non-graphics printers. Hardly any measurement instruments are designed to use the RS-232 interface, since it lacks the speed and flexibility of the IEEE-488.2 bus. However, for specific applications, such as reading in the data from remote dc sensors and sending data to loggers, it can be useful.

Other, more modern, serial, asynchronous data transmission protocols include the RS-422, RS-423, RS-449, RS-485 and USB. We will first describe the RS-232 serial interface, as it is the oldest serial bus. It has been widely used and its protocol is utilized with little variation in the later serial bus designs. RS stands for *Recommended Standard* (of the Electronic Industries Association, Washington, DC).

9.9.1 The RS-232C and D Interfaces

In describing the RS-232C interface, we encounter some new acronyms. RS-232C serial communications are sent by *Data Terminal Equipment* (DTE) and *Data Communications Equipment* (DCE). DTE include computer terminals, teleprinters, computers and digital instruments. DCE are devices (modems) that encode the serial digital signals into low bandwidth (sinusoidal) formats, compatible with transmission on voice band telephone lines. The rate that modems transmit data is given in baud, which stands for *bit rate audio*. (Modern modems used for serial telephone line communications between PCs and instruments typically run between 2,400 baud and 56,000 baud, depending on the

**FIGURE 9.29**

Standard pin connections for the RS-232C serial data interface. Pin nomenclature based on the connections in DTE equipment. (Source: Wolf, 1983)

limiting bandwidth of the transmission system. FAX modems typically run at 24,400 baud.) Just as the GPIB has a unique, standard connector, the RS-232C interface uses a standard male or female 25-pin, DB25P (male) or DB25S (female) connector, shown in Figure 9.29.

The RS-232C interface operates either in the simplex, half-duplex, or full-duplex mode. In the simplex mode, data transmission is unidirectional (e.g. from the computer to a printer). In the half-duplex mode, serial data can be sent in both directions, but in only one direction at a time. Full-duplex operation permits simultaneous, bidirectional, serial data transmission.

Data is transmitted as 8-bit ASCII words, signalled by high or low logic voltages on the transmitted data (TD) line. An example of this process is shown in Figure 9.30. In the idle state, the TD line is held high. At the beginning of data word transmission, the TD line goes low and stays low for one clock period. This is the start bit, which is always low. The receiving equipment senses the high-low transition of the start bit, and in order to verify start, samples the received data (RD) one-half clock period later. If low, the start bit is verified. The state of the received data line is then sampled eight times at intervals of one clock period. The last (eighth) sample is the MSB and is called the parity bit. In setting up an RS-232D interface, the user can specify the use of odd, even, or no parity in the data transmission process. If even parity is used, the parity (8th) bit is set so that the total number of logical 1s in the transmitted word, including the parity bit, is even. If the parity check algorithm in the receiver counts an odd number of highs in a received word, it

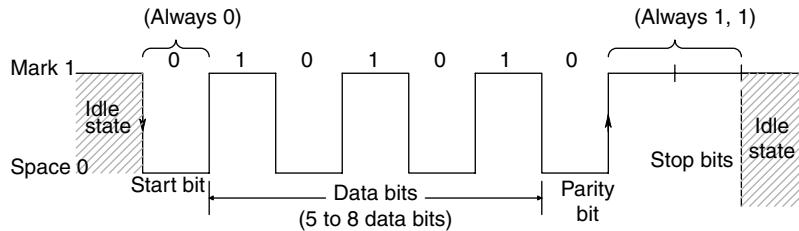


FIGURE 9.30

Example of an 8-bit, serial data signal sent on pin 2 of an RS-232C DTE.

declares a parity error. The parity error can terminate transmission, or require that the same character be sent again. Similarly, if an even number of 1s is counted in a received word when using odd parity, then an error is declared. Parity is generally used when serial data is transmitted under noisy conditions over long distances. It is normally not used in situations where the receiver is connected to the transmitter by a short cable in a low noise environment.

At the end of data transmission, after the parity bit is sent, one or two high stop bits are sent before the TD line is declared idle and is ready to transmit the next word (ASCII character). Note that there are a number of handshaking functions on the RS-232C lines which accompany the serial data transmission and receiving lines. These include—ring indicator, data terminal ready, carrier detect, signal ground, data set ready, clr to send, request to send and others. For detailed descriptions of the operation of the RS-232C interface, the reader can consult Section 5.2 in Stone (1982), Chapter 18 in Wolf (1983), Chapter 6 in Tompkins and Webster (1988), or Motorola Application Note AN-781A.

In the earliest days of RS-232 use, it was common for computers to be connected to teletypewriters such as the Western Electric ASR33. The ASR33 used no IC electronic components. All logical and switching operations were done electromechanically at very low baud rates. Modern serial data communications use UART (universal asynchronous receiver/transmitter) IC chip sets which handle the parallel to serial conversion of data for transmission and at the receiver, convert the RD to parallel form.

The latest RS-232 embodiment uses the 'D' designator. It is essentially the same as the RS-232C in its specifications and protocols. However, it uses an 8-pin, telephone type connector, the RJ-45 (similar to those used in 10-BaseT Ethernet lines). The RJ-45 pin signal descriptions and their acronyms are:

1. DCE Ready/Ring Indicator (DSR/RI)
2. Received Line Signal Detector (DCD)
3. DTE Ready (DTR)
4. Signal Ground (SG)
5. Received Data ($R \times D$)
6. Transmitted Data ($T \times D$)
7. Clear to Send (CTS)
8. Request to Send (RTS)

Like the RS-232C, The D is used for low speed, asynchronous communication between a computer and a single, slow, peripheral device.

The present state-of-the-art for the RS-232C & D protocol limits data transmission rates to less than 20 kbaud, and wire cables are seldom effective over 15 m (50 ft) in length.

Also, RS-232 protocol restricts the dV/dt of the T × D line data to a maximum of $30\text{ V}/\mu\text{s}$. To circumvent these problems, other serial data transmission standards were developed, making use of improvements in hardware to achieve higher baud rates and longer transmission lines. These interfaces are described in the next section.

9.9.2 The RS-422, RS-423 and RS-485 Interfaces

These interfaces make use of our knowledge about the transmission properties of transmission lines for transient signals and improved IC designs to realize improved data transmission rates over longer distances than possible with the 40+ year old technology of the RS-232 interface. Their properties are summarized in the following paragraphs, and are illustrated in Figure 9.31.

The *RS-422A Interface* uses a balanced, twisted pair transmission line, terminated in the characteristic impedance of the line. A balanced, or differential, amplifier line driver, as well as a differential line receiver, is used. The RS-422A interface can transmit data at up to 10 Mbaud, and can have lengths of up to 1200 m. There is a trade-off between baud rate and cable length—10 M baud is possible on a 12 m cable, 1 Mbaud on a 100 m cable, 100 kbaud on a 1 km cable and up to 80 kbaud on a 1.2 km cable. The TI 9636 and 9637 ICs, as well as the Motorola MC3487 differential receiver and the MC3486 differential driver can be used to realize an RS-422A link.

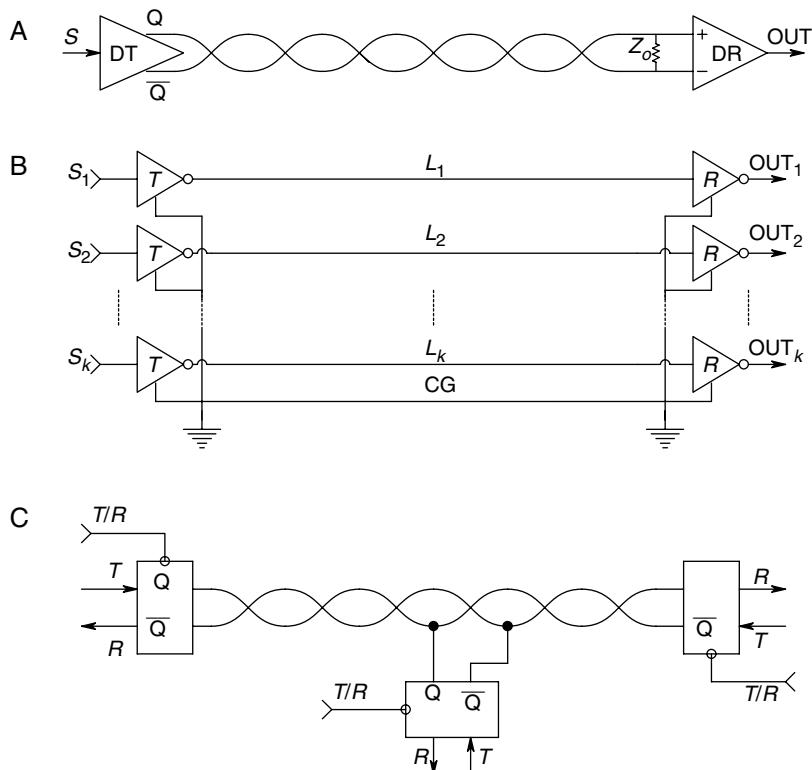


FIGURE 9.31

A. RS-422A, balanced, twisted pair transmission line. Note: DT = differential transmitter, DR = differential receiver, Z_o = characteristic impedance of the line. B. RS-423A, unbalanced serial data transmission lines. Note: CG = common ground C. RS-485, balanced, bidirectional, twisted pair serial data transmission system. Differential, tristate, transceivers are used.

TABLE 9.4

Specifications of RS serial data lines

Specifications	RS-232C&D	RS-423	RS-422	RS485
Mode of operation	Single ended	Single ended	Differential (twisted pair)	Differential (twisted pair)
Allowed Number of Tx and Rx	1 Tx, 1 Rx	1 Tx, 10 Rx	1 Tx, 10 Rx	32 Tx, 32 Rx
Maximum data rate, bps	20 k	100 k/10 M	100 k/10 M	100 k/10 M
Normal driver output voltage	± 5 to ± 15 V	± 3.6 V	± 2 V	± 1.5 V
Maximum driver output	± 25 V	± 6 V	± 6 V	± 6 V
Tx Load impedance, Ω	3–7 k	≥ 450	100	54
Rx input sensitivity	± 3 V	± 200 mV	± 200 mV	± 200 mV
Rx input voltage range	± 15 V	± 12 V	± 7 V	-7 to $+12$ V
Maximun Rx input resistance, Ω	3–7 k	< 4 k	< 4 k	≥ 12 k
Maximum cable length, feet	50	4000	4000	4000

The mechanical connections for the RS-422A interface are specified by the RS-449 standard, which specifies a 37-pin connector supporting the mandatory twisted pair lines for receive ready, test mode, data mode, request to send, clear to send, receive data and send data [Stone, 1982].

The *RS-423A Interface* uses an unbalanced single line, similar to the RS-232D link. Even so, the RS-423A link is faster than the RS-232C. According to Stone (1982), it was designed to provide a linkage between the old RS-232C and the RS-422A interfaces. Its operational protocol is very similar to RS-232, although improvements in the driver and receiver electronics allow it to operate at rates up to 100 kbaud over short cables (< 30 m) and at significantly slower rates over cables up to 1.2 km.

The *RS-485 Interface* is a balanced (twisted pair), ‘party line’, on which a number of secondary receivers and transmitters can operate. In this respect, it is effectively a data bus. Although the RS-485 interface can transmit data up to 10 Mbaud, considerations of line resistance, loading and terminations set practical limits on line lengths. The SN75172 and SN75174 differential drivers, and the SN74173 and SN75175 differential receivers are used to implement the RS-485 protocol.

Table 9.4 summarizes some important specifications for the RS-series serial data lines.

9.9.3 The Universal Serial Bus (USB)

The USB is rapidly replacing the RS-232C & D serial interfaces as a means for a computer to communicate with medium speed peripheral devices such as scanners and digital cameras. Practically all new PCs made in the last four years have two USB ports. Reasons for the popularity of USBs includes its simple connection, plug and play feature, hot insertion capability, speed and versatility.

USB v.1.0 was released in January 1996. It permitted low speed (1.5 Mb/s) and medium speed (12 Mb/s) data transfer rates. (Note that Mb/s refers to Megabits per second, not Megabytes.) The actual USB data rate is slower because of USB protocol (operating) data sent to and from a peripheral device. USB v.1.1 was released in September 1998. This edition fixed many of the problems found in release 1.0. The USB v.2.0 was released in early 2000 and has a maximum bit clock rate of 480 Mb/s. It is downward compatible with v.1.1 and v.1.0.

The USB data is clock encoded and uses NRZI with bit stuffing. A *transaction = 3* packets, with flexible packet size of up to 1023 Bytes. A *start-of-frame* marker occurs

every 1 ms and it has 125 ms microframes in SOF mode. A USB system sets up a unique *pipe* (*aka* data pipeline) between host and a device, which can be assigned bandwidths. It also has the flexibility to implement split USB transactions for 12 Mb/s operation with a 480 Mb/s (v.2.0) hub.

USB cabling is electrically simple. It uses four conductors—ground, +5 V, and a differential pair for bidirectional data transfer. There are two connectors on a USB cable, A and B. The A connector mates with the host computer's USB port. It consists of a simple, rectangular metal sheath, ca. 12×4 mm, covering a plastic insert holding four gold-plated wiper strips. The plastic insert serves to polarize the host connector so it cannot be inserted inverted. The B connector on the peripheral end of the USB cable can be smaller than the host connector and is unique to that peripheral (digital camera, scanner, external CD-ROM reader/burner, etc.). Cable length is limited to a maximum of ca. 5 m [Simpson, 2003].

Power requirements on the USB are such that a peripheral device that requires less than 100 mA can be powered from the host's hub. Devices such as external CD-RW drives that draw more than 100 mA from the host are generally self-powered (batteries, or isolated plug-in dc power supply). When a peripheral USB device is first connected, it identifies itself to the host and tells the host how much current it draws. A battery powered device, such as a digital camera, can draw power from the host if less than 100 mA is required, then switch back to its internal batteries when it is disconnected from the USB, or the host is shut down. When a device is made dormant (suspend mode), it cannot draw $> 500 \mu\text{A}$ from the bus, if it is bus powered.

By using *external expanders*, it is theoretically possible to connect a total of 127 peripheral devices to one host computer's USB port. If two host USB ports are used fully, one can have an amazing 254 peripherals attached to the computer. In practice, such a great number of USB peripherals would probably run out of USB port bandwidth and require external power. At present, two computers cannot be connected directly, USB to USB. Instead, one needs to use an external 'USB Bridge', *aka* a USB to USB adapter between them. Since most modern laptops no longer come with RS-232 ports, it is possible to transform a USB port to an RS-232 port with an external converter, such as the CyberResearch® USBS-221 device. CyberResearch also makes a series of USB adapters for RS-422/485 interfaces. Additionally, CyberResearch offers a type II, PCMCIA laptop card that has four USB 2.0/1.1 ports.

9.9.3.1 Overview of USB Functioning

The digital communication and control of USB peripheral devices is incredibly complex. What simplicity USB has electrically is more than overset by its digital protocols. All peripheral USB devices have addressable 'endpoints' which are memory buffers. An endpoint can be as simple as one register, or multiple end points which store incoming and outgoing data (endpoints 0, 1, 2, ..., N). All endpoints must be able to both send and receive data, but obviously, can only communicate in one direction at a time.

A logical connection link must be set up between the host and a peripheral device before a data transaction can take place. This connection is called a *pipe*. The pipe is set up as soon as the host recognizes the device connected through the USB. When the host responds to a connect signal from the device, the device sends its requirement for data transfer type and speed. The host can refuse to establish a pipe if it lacks the power needed by the device, or cannot meet the device's bandwidth requirement. A 'smart' device, meeting power limitations, can lower its requested data rate and again try to negotiate a pipe. Pipes can be explicitly managed—bandwidth and capacity can be assigned to each pipe by the host software. This flexibility guarantees adequate USB

bandwidth for operations such as bulk and isochronous data transfer, and provides an easy mapping from the host's application software to a specific device function interface.

Pipes connect through the second and third layers of the USB protocol. (The USB protocol layers are Function layer, USB device layer, USB interface layer.) Each device has a control endpoint (*Endpoint 0*) linked through a pipe at the device level (second layer). All other endpoint connections are at the third layer. Furthermore, there are two types of pipes—stream and message. Stream pipes are unidirectional and unstructured to pass data streams, while message pipes are two-way structured.

The USB has four types of data transfer modes:

1. Control
2. Interrupt
3. Bulk
4. Isochronous

Control mode is initiated by the host and is used to initialize the peripheral device. Bidirectional 'handshaking' traffic occurs in one direction at a time. In *Interrupt Mode*, the host queries devices to see if they need servicing. *Bulk Mode* is used when data transfer accuracy is essential, as when reading or writing to a peripheral CD-RW. *Isochronous Mode* sacrifices data accuracy for data timing. Uses include USB loudspeakers, and precision motion generation.

Quoting Simpson (2003):

"The following describes a typical data flow for a device when it is initially plugged into a host's bus when the host is active. Remember here that the host has an internal USB hub, and additional hubs may be connected downstream from the host's hub.

1. The host recognizes that a device has been attached to one of its USB hubs. It realizes this by a simple resistive divider that is connected to the differential data pair of wires in the USB bus. The resistors are inside the USB hubs and devices.
2. The host sends a *Get_Port_Status* request to the hub to find out more about what has been plugged in. It could be another hub, a device connected directly to the host's hub, or a device that has been plugged into one of the downstream hubs.
3. After receiving a response from the hub, the host issues a *Set_Port_Feature* command in which the hub issues a *Reset* over the data pair, but only to the newly connected device on the USB bus.
4. The host then checks to see if the device has come out of the reset state by issuing a *Get_Port_Status* command to the hub. After reset, the device is in the *Default State* and can only draw a maximum 100 mA. In *Default State*, the device can communicate with the host through *Endpoint 0*.
5. The hub now detects the device's speed by using the resistive dividers that are attached to the USB bus. The hub sends the speed of this device back to the host.
6. The host then sends a *Get_Descriptor* command to the hub in which the hub gets the packet size needed from this particular device and sends the result back to the host.
7. The host then sets a *Set_Address* command to the hub which sends this information to the device. The device in turn acknowledges the command back through the hub to the host and sets up this address internally.

8. To learn more about this device, the host sends a *Get_Descriptor* command to the address that the device has been given. The information that is returned to the host consists of various details of the device that the host needs to know for its operation. These queries by the host continue two more times to retrieve all the information needed.
9. Based on the information received from the device, the host determines the best (logical) device driver to use for communication with it.
10. The device driver in the host now takes over by requesting a *Set_Configuration* command. There may be several configurations for one device, and the device driver determines which to use based on the information received from the device in response to the (previous) *Get_Descriptor* command.
11. The device is now ready to use."

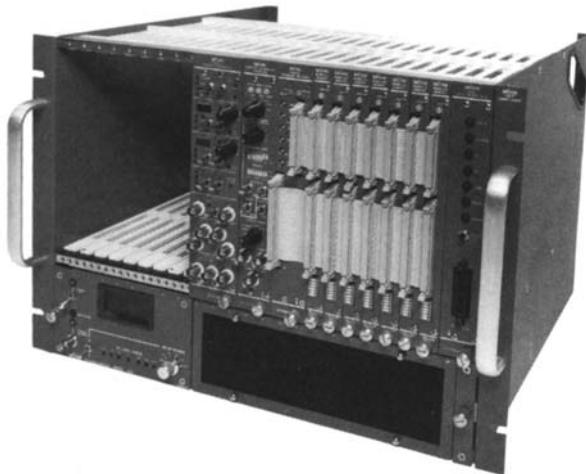
Clearly, plugging in and setting up a peripheral USB device is a user friendly process. The set-up protocol is complex, as shown above. This complexity is necessary to allow 'hot' connection and automatic device ID and initialization. It also allows the host to keep track of the different USB devices which are on the bus and operate them correctly. Included in the host's Microsoft® USB software are a spectrum of device drivers, including *Human Interface Device* (HID) drivers. HID drivers are available for devices such as keyboards, mice, game controls, remote controls and joysticks. As the use of the USB is rapidly expanding, makers of various new peripheral devices include driver software with their products.

9.9.3.2 Summary

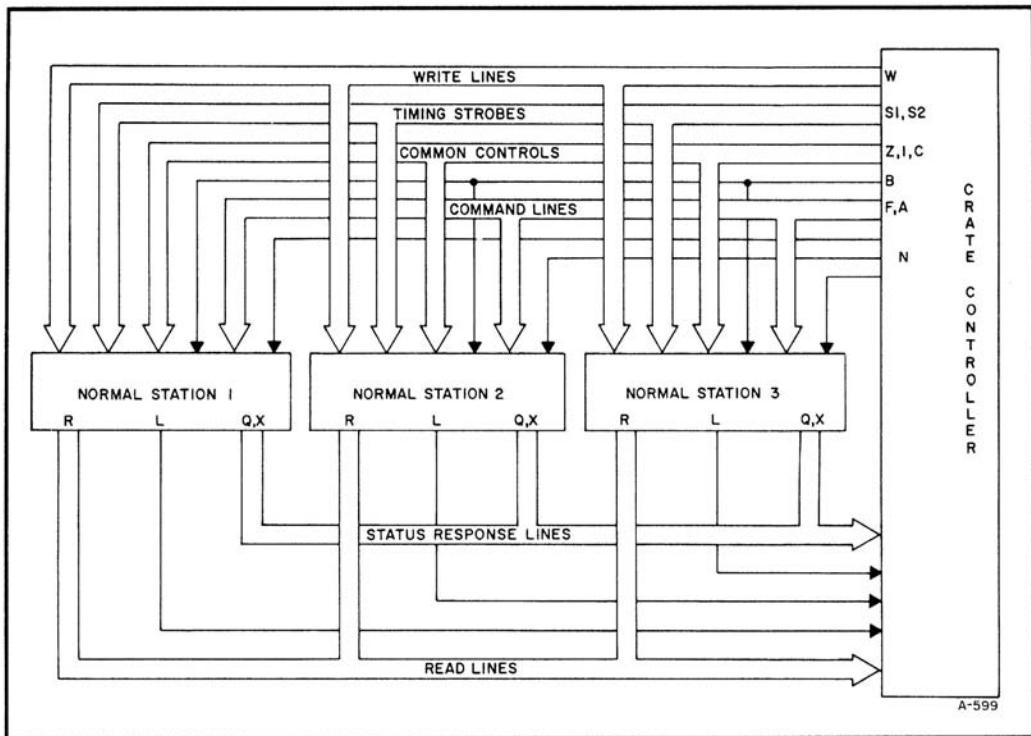
The USB architecture is here to stay. More and more computer accessory devices are designed around it—scanners, printers, digital cameras, digital telephones, wireless internet adapters, PDAs, outboard bulk storage devices (ZIP drives, TEAC® CD-RW drives, and Lexar Media™ JumpDrive™ memories), computer game devices, joysticks, flight simulator hardware, etc. A major advantage of the USB is that devices can be 'hot plugged' into the host computer, which will automatically recognize and initialize them for use.

9.10 The CAMAC (IEEE-583) Modular Instrumentation Standard, and the VXI Modular Instrumentation Architecture

The IEEE-583, -595 and -596 CAMAC standards define a hardware and data transmission system which is used to house, support and communicate with various compatible instrumentation modules. A typical CAMAC crate, illustrated in Figure 9.32, is a physical package which has 25 slots or powered stations for compatible, plug-in instruments. Generally, the two right hand slots are used to house the CAMAC controller module. The controller module allows the individual modules to be coordinated and controlled by an external computer using the IEEE-488.2 bus, or in some cases, an RS-232C interface. Commands and data transfers to and from the modules and the controller are made over the CAMAC DATAWAY, which use 24-bit data words. Figure 9.33 illustrates the (internal) DATAWAY connections to three modules in a CAMAC system. The read and write lines are each 24 bits wide. Two lines are used for the timing strobes; there are three

**FIGURE 9.32**

A partially filled CAMAC 'crate'.
(Source: Courtesy LeCroy Corp.).

**FIGURE 9.33**

DATAWAY™ signal paths to three CAMAC modules from a crate controller. (Source: Courtesy LeCroy Corp.)

common control lines, three status lines, four address bus lines, five function lines, and one station number line per module, for a total of 66 lines/module!

Lack of space prevents going into the details of the CAMAC DATAWAY control and command operation protocols. The interested reader can find their descriptions in the LeCroy instruments catalog (1990). The major advantage of CAMAC housed instrumentation is its ease of use through the GPIB and its speed in local data manipulation. Individual instrument modules are managed by the CAMAC controller over the

DATAWAY and only one GPIB bus is required. CAMAC systems represent top-of-the-line instrumentation, and as such are expensive. They are used extensively in large industrial applications and for measurements in nuclear and plasma physics.

The recently developed, VXIbus measurement systems architecture is similar in many respects to the CAMAC system, but improves on certain aspects of CAMAC system design. VXI stands for VMEbus eXtensions for Instrumentation. The VXIbus is a non-proprietary, open architecture bus for modular instrumentation systems that fully incorporates the VMEbus standard. It includes not only mechanical and electrical interfacing, but also communications protocols between compatible devices and an external controller. The VXIbus architecture was developed by a consortium of instrumentation companies, including Textronix, Hewlett-Packard and National Instruments. As in CAMAC systems, special VXI instrument cabinets are used to support a wide variety of modular instruments. All members of the HP 75000 VXIbus instrument family use the Standard Commands for Programmable Instruments (SCPI), an industry standard set by the IEEE, updated in 1999, for command, control and communication. SCPI conforms to the new IEEE-488.2 bus standard. In addition, all HP 75000 products support the HP Interactive Test Generator (HP ITG) software. There are many VXIbus instrumentation modules available from HP and other companies. All of these modules are characterized by input and output connectors, but no knobs, dials, pushbuttons or switches (scary). All settings and control operations are accomplished by software over the VXIbus. Some of the VXIbus module functions available from HP include command modules and controllers, digital multimeters, counters, function generators, amplifiers, D/A converter, digital I/O, digitizing oscilloscopes, power meter, various switches and multiplexers, a serial port adapter for RS-232C/422 protocols, and a VXIbus to MXIbus extender to tie VXI mainframes together and to an MXIbus computer interface. (MXI stands for Multisystem eXtension Interface, and was developed by, and announced as an open industry standard in April, 1989 by National Instruments Corp.)

Tektronix (1990) gave a good summary of VXI communications protocols:

Two communications protocols are defined in VXI for Message Based Devices: Word Serial Protocol (WSP) and Shared Memory Protocol (SMP). The other type of communication possible in VXI, is low level, binary data transfers which is defined for Register Based Device communication. Communication in a VXIbus system has a hierarchical commander/servant structure which is established by the system Resource Manager when the system powers up. A Commander is any device with one or more lower level devices (servants). The Commander initiates communication with its servants. A servant is a device which operates under the control of a Commander. All message based devices must respond to a WSP, but can also respond to the more advanced communication Shared Memory Protocol. With WSP, the required communication is 16-bit words transferred serially by reading or writing to one of the communication registers resident on a servant device. There is a WSP command set defined extensively in the VXIbus Specification.

SMP defines the operation of a shared memory channel established between two communicating devices with shared memory space. A channel can be asynchronous or synchronous. Asynchronous channels are used for high priority communications such as warnings, events and commands which must be immediately executed. The data transfer is accomplished over the ComputerBus in large data blocks. SMP is an efficient method of transferring large amounts of data in a VXI system whether the information is command strings or results obtained from performing a test'.

Similar to CAMAC systems, VXIbus equipment is expensive, and best suited for an industrial, ATE environment. For further details of VXIbus instrument systems and the

MXIbus, the reader should consult catalogs from National Instruments, Tektronix, and Hewlett-Packard.

9.11 Effect of Transmission Lines on the Transfer of Digital Data

Whenever we connect a computer to an instrument through a parallel or serial interface cable, we face problems associated with the transmission of digital data on a transmission line. Similar problems exist on printed circuit boards when we couple data and address lines between ICs, whenever the clock speed is greater than about 10 MHz. The main sources of difficulty come when the transmission line properties of the high speed data paths are neglected in the design of interfaces. We can observe delays, attenuation and rounding of pulses, as well as glitches caused by signal reflections at branches and terminations of the lines, which are in turn caused by improper terminating resistances and impedance mismatches at signal branch points on the lines. All these problems serve to limit the rate of data transfer on transmission lines.

In this section, we will examine some of the major properties of transmission lines used in computer interfaces. Such lines include coaxial cables, shielded twinax cables, twisted pairs of wires and ribbon cables. It will be seen that all transmission lines can be described by a common model.

9.11.1 The Transmission Line Model

In our treatment of transient signal propagation on short transmission lines, we will assume that the lines are lossless (i.e. $R\Delta x = G\Delta x = 0$ in Figure 9.34). It can be shown [Lathi, 1965] that the characteristic impedance of the transmission line is given by:

$$Z_o = \sqrt{L/C} \Omega \quad (9.37)$$

The velocity of wave propagation on the line is:

$$v = 1/\sqrt{LC} \text{ m/s} \quad (9.38)$$

where L is the inductance per unit length of the uniform line and C is the capacitance per unit length between conductors. Most coaxial cables are made to have either 50 or 75 Ω

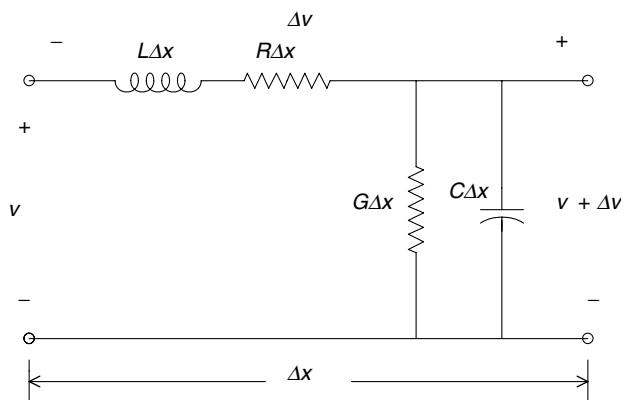
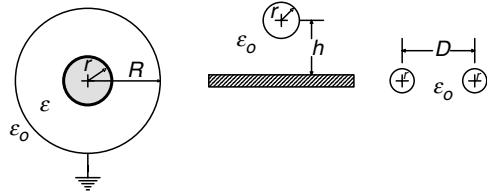


FIGURE 9.34

Lumped parameter, equivalent circuit of a length, Δx , of a transmission line.

**FIGURE 9.35**

End views of three common transmission line geometries: Left—a coaxial cable; Center—a wire over a ground plane; Right—a parallel balanced line.

characteristic impedances and they come in a variety of physical sizes and voltage ratings. For example, a coaxial cable commonly encountered in instrumentation systems is the ubiquitous RG58A/U, which has $Z_o = 50 \Omega$, $C = 29.5 \text{ pF/ft}$ and $v = 0.649c$ (c is the speed of light in *vacuo*). In the case of ribbon cables having alternate conductors grounded, wires running parallel to a ground plane and parallel wire, twin lead antenna wire, Z_o s generally ranges 100–300 Ω or more. Figure 9.35 illustrates the end views of three common transmission line geometries. From electrostatics, the capacitances per unit length of the lines can be shown to be [Taub and Schilling, 1977]:

For coaxial cables,

$$C = \frac{2\pi\varepsilon}{\ln(R/r)} \text{ F/m} \quad (9.39)$$

For round wire over a ground plane,

$$C = \frac{2\pi\varepsilon_0}{\ln(2h/r)} \text{ F/m} \quad (9.40)$$

For parallel wires in space,

$$C = \frac{\pi\varepsilon_0}{\ln(D/r)} \text{ F/m} \quad (9.41)$$

The inductance of these models can also be calculated [Krause, 1953, Motorola AN-270]. (Results are good approximations when h/r and $D/r \gg 1$.)

For coaxial cables,

$$L = \frac{\mu}{2\pi} \ln(R/r) \text{ H/m} \quad (9.42)$$

For a wire in space over a ground plane,

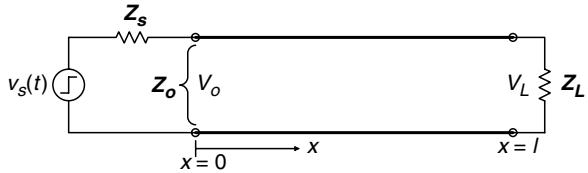
$$L = 11.7 \log 10(2h/r) \text{ nH/in} \quad (9.43)$$

For parallel wires in space,

$$L = \frac{\mu_0}{\pi} [1/4 + \ln(D/r)] \text{ H/m} \quad (9.44)$$

As an example, we use equations 9.37, 9.39 and 9.42 to find an expression for the characteristic impedance of coaxial cables:

$$Z_o = \frac{\ln(R/r)}{2\pi} \sqrt{\mu/\varepsilon} \Omega \quad (9.45)$$

**FIGURE 9.36**

Schematic of a lossless transmission line of characteristic impedance Z_o and length l .

9.11.2 Reflections on an Improperly Terminated, Lossless Transmission Line

As we will see below, the result of sending digital (pulsatile) data waveforms over improperly terminated transmission lines, is the creation of transient glitches at the receiving end, which, along with other noise on the line, can cause errors in data transfer. A transient waveform will propagate without distortion or attenuation on a lossless transmission line with a velocity $v = 1/\sqrt{LC}$ m/s. On most coaxial cables used for digital data transmission, $0.65 \leq v/c \leq 0.83$, depending on cable design. However, when the transient voltage propagating on the line reaches the terminating or receiving end, a reflection, or backwards propagating wave can arise at this end to satisfy certain boundary conditions. This first reflection propagates back to the source end of the line with velocity v , where a second reflection can arise and propagate at velocity v to the terminating end, etc. This process can be described mathematically in some detail. The interested reader is referred to the text by Lathi (1965) for an excellent derivation of the relations we use.

Figure 9.36 shows a lossless transmission line of characteristic impedance, Z_o . To understand the dynamics of reflections on the line, we consider a voltage wave (step) which has propagated to the end of the line. A_d is the instantaneous voltage of the propagating wave just before it 'sees' Z_L , and A_r is the instantaneous voltage of the back propagating, reflected wave originating at Z_L to satisfy Ohm's law boundary conditions. Thus, at the Z_L termination, at $t = \tau = L/v$, we have:

$$V_L = V_d + V_r \quad (9.46)$$

$$I_L = (A_d - A_r)/Z_o \quad (9.47)$$

The boundary condition at Z_L will be satisfied if

$$\frac{V_L}{I_L} = Z_L = Z_o \frac{A_d + A_r}{A_d - A_r} \quad (9.48)$$

Solving equation 9.39 for the ratio of A_r to A_d , we obtain the reflection coefficient, $\rho_L(s)$, defined for the *line's termination*:

$$\rho_L(s) = \frac{Z_L(s) - Z_o(s)}{Z_L(s) + Z_o(s)} = \frac{A_r}{A_d} \quad (9.49)$$

A similar reflection coefficient can be found for the *source end* of the line:

$$\rho_S(s) = \frac{Z_S(s) - Z_o(s)}{Z_S(s) + Z_o(s)} \quad (9.50)$$

where Z_S is the Thevenin source impedance of the line.

In example 1, we examine what happens on a mismatched line when a positive, 5.0 V step is generated by $v_s(t)$ at $t=0$. We assume zero initial conditions on the line and let $Z_o = 75 \Omega$, $Z_S = R_S = 75 \Omega$ and $Z_L = R_L = \infty$ (open circuit). The step voltage is acted on by a voltage divider at $x=t=0$. The source-end voltage divider attenuates by a factor of $75/(75+75)=0.5$. Hence, the initial propagating voltage, V_o , is 2.5 V. The voltage at R_L is 0 until $t=\tau=L/v$. At $t=L/v$, the wave reaches R_L and boundary conditions require that a reflected wave be generated. The termination reflection coefficient is found to be:

$$\rho_L = \frac{R_L - Z_o}{R_L + Z_o} = \frac{\infty - 75}{\infty + 75} = +1 \quad (9.51)$$

The source reflection coefficient is:

$$\rho_S = \frac{R_S - Z_o}{R_S + Z_o} = \frac{75 - 75}{75 + 75} = 0 \quad (9.52)$$

Hence the first reflected wave has a value of:

$$V_r(1) = 1.000 \times 2.5 = 2.5 \text{ V} \quad (9.53)$$

The voltage at the load is thus $V_L = 2.5 + 2.5 = 5.0 \text{ V}$, over $1/v \leq t \leq 31/v$. At $t = 2L/v = 2\tau$, the reflected wave of 2.5 V reaches the source end of the line. At this point, there will be no reflection, because $\rho_S = 0$. The steady state has been reached. In the steady state, we expect the voltage everywhere on the line to be 5.0 V, as no current flows on the line. The waveforms for V_o and V_L are shown in Figure 9.37A.

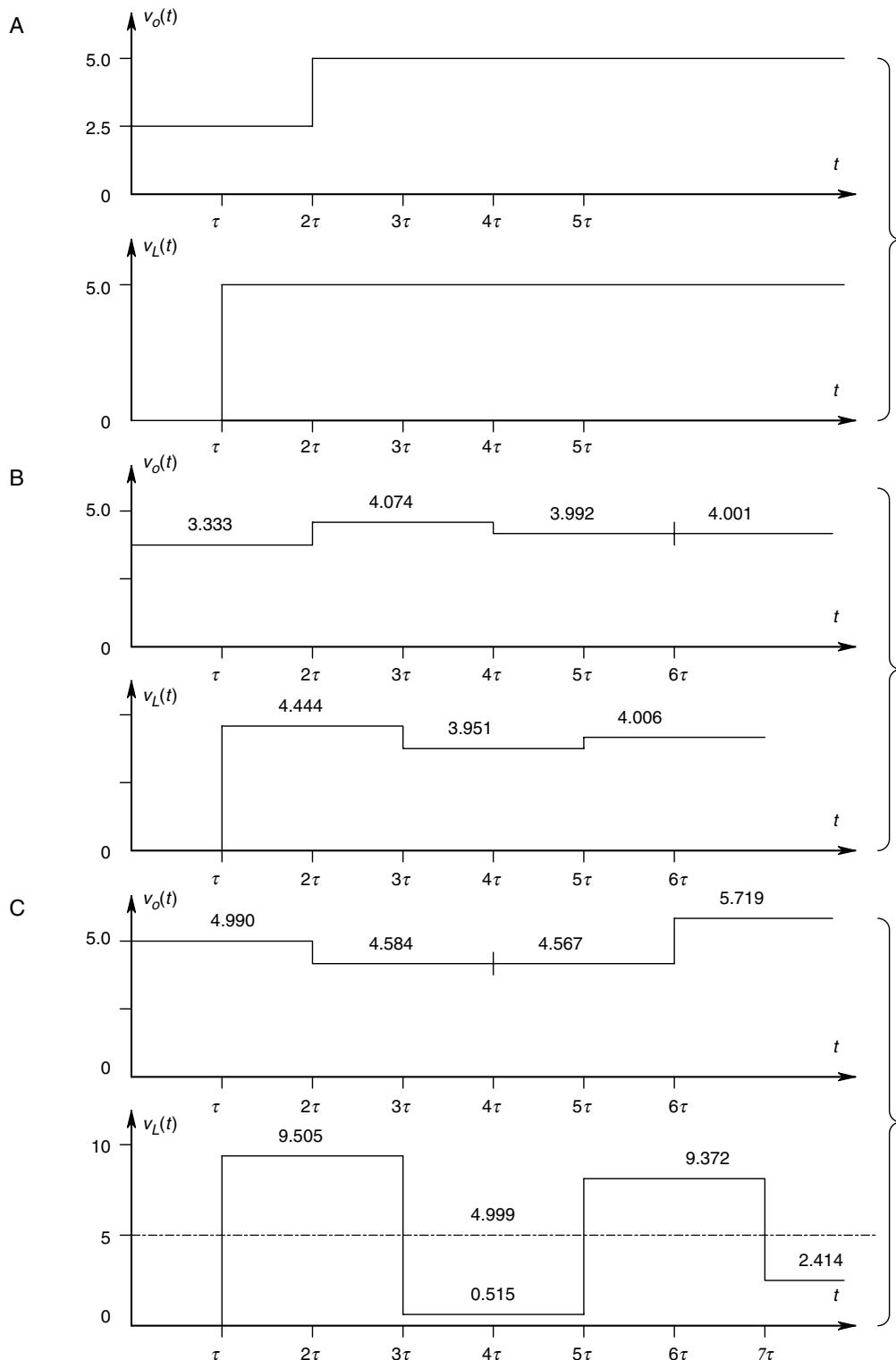
In example 2, there are reflections at both ends of the line. We let $Z_o = 50 \Omega$, $R_S = 25 \Omega$ and $R_L = 100 \Omega$, giving a moderately mismatched and underdamped line. Now, the reflection coefficients are $\rho_L = (100-50)/(100+50) = 1/3$ and $\rho_S = (25-50)/(25+50) = -1/3$. The amplitude of the initial wave at $t=0$ is $V_o(0+) = 5 \times 50/(50+25) = 3.333 \text{ V}$. The 3.33 V step travels at v m/s to the load end of the line, arriving at $t=\tau=L/v$ s. There, to satisfy boundary conditions, a reflected wave, $V_r(1)$, is generated, propagating toward the source. Its amplitude is given by:

$$V_r(1) = 3.333 \times \rho_L = 3.333/3 = 1.111 \text{ V} \quad (9.54)$$

Now, the voltage across R_L is $V_L = 3.333 + 1.111 = 4.444 \text{ V}$ for $\tau \leq t \leq 3\tau$. $V_r(1)$ arrives at the source end of the line at $t=2L/v$. A second reflection originates at $t=2L/v$ at the source end of the line, propagating toward R_L . The process of waves of diminishing amplitude being reflected from end to end of the transmission line continues, until practically, a dc steady state is reached.

Table 9.5 illustrates a systematic way to find the amplitudes of the reflections at the ends of a mismatched, lossless line. It is seen that a type of power series in the reflection coefficients describes the line's behavior. In the Table, $v_s(t)$ is the Thevenin (open circuit) source voltage, V_o is the voltage at the source end ($x=0$) end of the line and V_L is the voltage at the load end ($x=L$) of the line. $v_s(t)$ is assumed to be a 5V step for simplicity. Other numerical parameters are from example 2. Data from Table 9.5 are plotted in Figure 9.37B. Note that both $v_o(t)$ and $v_L(t)$ closely reach steady state in three reflections.

In this third example, we examine the troublesome situation where logic signals are sent to a receiver along a long, severely mismatched transmission line. We assume a

**FIGURE 9.37**

A. Voltages at the source and termination ends of a transmission line. See text Example 1 for conditions. B. Transients on the transmission line of Example 2. C. Transients on the transmission line of Example 3.

TABLE 9.5

Example of systematic means to calculate the reflections on a mismatched, lossless transmission line

Time	V_o	V_L
$\tau = L/v$		
0	$V_o(1) = V_S(0+) \frac{Z_o}{Z_o + Z_s} = 5 \frac{50}{50 + 25} V_o(1) = 3.33333$	0
τ	$3.33333 = 3.33333[1.33333]$	$V_L(1) = V_o(1)[1 + \rho_L] = 4.44444$
2τ	$V_o(2) = V_L(1) + V_o(1)[\rho_L^1 \rho_S^1] = 4.44444$ $-3.33333[0.11111] = 4.07407$	4.44444
3τ	$4.074074 = 3.9506173$	$V_L(3) = V_o(2) + V_o(1) [\rho_L^2 \rho_S^1] = 4.07407$ $-3.33333[0.037037] = 3.95062$
4τ	$V_o(4) = V_L(3) + V_o(1) [\rho_L^2 \rho_S^2] = 3.95062$ $+3.33333[0.0123457] = 3.9917696$	3.95062
5τ	3.99177	$V_L(5) = V_o(4) + V_o(1) [\rho_L^3 \rho_S^2] = 3.99177$ $+3.33333[0.0041152] = 4.0054870$
6τ	$V_o(6) = V_L(5) + V_o(1) [\rho_L^3 \rho_S^3] = 4.00549$ $-3.33333[0.0013717] = 4.00091$	4.00549
∞	$V_o(\infty) = V_L(\infty) + V_S(0+) \frac{Z_L}{Z_L + Z_s} = 5.0 \frac{100}{100 + 25} = 4.00000 \text{ V}$	

Note: Parameters are from Example 2.

TABLE 9.6

Reflections on the cable of example 3

t	V_o	V_L
0	4.99002	0
τ	4.99002	9.50480
2τ	4.58415	9.50480
3τ	4.58415	0.51537
4τ	4.56762	0.51537
5τ	4.56762	9.37157
6τ	5.71910	9.37157
7τ	5.71910	2.41447
∞	4.99950	4.99950

Note: See Figure 9.37C for plots

coaxial line is driven from an op-amp buffer having a very low output resistance (e.g. $R_S = 0.1 \Omega$). The signal is received by a linear amplifier with an input resistance of $R_L = 1000 \Omega$. The characteristic impedance of the coaxial cable is 50Ω , its length is 3 m and the propagation delay at 1 MHz is known to be 5.136 ns/m. The reflection coefficients are found to be $\rho_L = 0.90476190$, $\rho_S = -0.9960798$. The amplitude of the initial wave is:

$$V_o = 5.0 \frac{50}{50 + 0.1} = 4.99002 \text{ V} \quad (9.55)$$

Following the procedure used in Table 9.5, we can calculate and plot the voltage transient at the load end of the cable, $V_L(t)$. The time for a wave to travel down the cable is $\tau = L/v = 15.408 \text{ ns}$. For $0 \leq t \leq \tau$, $v_L(t) = 0$, as the wave has not yet reached R_L . At $t = \tau$, the V_o wave reaches R_L and the first reflection is generated, etc. Table 9.6 summarizes the V_o and V_L values for this cable.

Of particular importance in this severe case of transmission line mismatch is the fact that the received voltage goes low 46.2 ns after going high. Such behavior will render data transmission virtually impossible, or at least severely limit its rate to well below the reciprocal of the time it takes $v_L(t)$ to reach values above the logic HI level at the receiver.

By the simple expedient of making $R_S = R_L = Z_0$, all reflections on the line are eliminated and the only problem is the 15.4 ns initial delay in the signal reaching R_L . Of course, V_S should now be made 10 V because the steady state voltage at R_L will be $V_S/2$ by the R_S - R_L voltage divider.

In the examples above, we have assumed a short (< 10 m), lossless transmission line. If longer lines are used, in addition to possible reflection problems due to mismatched terminating and source impedances, there is noticeable attenuation of high frequencies due to cable losses and a resulting rounding and attenuation of pulses as they travel along the line. The longer the line, the more severe the attenuation and loss of high frequencies, and slower the pulsed data must be sent in order to resolve individual pulses. Voltage and current transfer functions for lossy transmission lines have been derived by Lathi (1965). They can be expanded into power series, and then inverse Laplace transformed to obtain the voltage and current at any distance along a finite line (i.e. $v(x, t)$ and $i(x, t)$, given an input transient at $t = 0$).

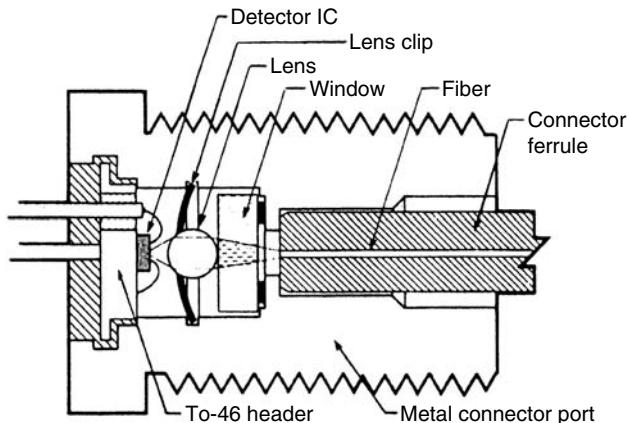
To model the behavior of a transmission line in the time domain, two approaches are commonly used. The first approach uses a series of lumped parameter, per unit length R - L / C - G circuits as shown in Figure 9.34. Their behavior is solved with an electronic circuit analysis program (ECAP) such as MicroCap™ or SPICE. The second approach is to use FFT techniques with the voltage transfer function, $Hv(x, j\omega)$ to directly find the voltage on the line, $v(x, t)$.

9.12 Data Transmission on Fiber Optic Cables

A relatively new technology, fiber optic cables (FOCs) is the emerging means for high speed, serial data transmission. It is rapidly replacing conventional wire transmission lines in point-to-point applications. FOC systems are used in increasing numbers in instrumentation systems because of their low cost, high reliability, low loss, immunity from EMI and wide bandwidth.

Fiber optic cables are dielectric waveguides in which information is generally transmitted as 0,1 amplitude modulated electromagnetic waves of light wave frequencies. A major application of FOC systems is in robust telecommunications. The channel capacity of fiber optic communications links, however calculated, far exceeds that of conventional wire communications (e.g. coaxial cables, twisted pairs) and present satellite microwave transponders. The present bandwidth of FOC data transmission systems is limited by the power, wavelength and speed of the LED or laser diode (LAD) sources, and the sensitivity and response time of the PIN or avalanche photodiode detectors coupled to the FOC. Theoretically, FOCs can support data bandwidths approaching a terahertz (THz); the present state-of-the-art, limited by terminal photonic devices, is in the low GHz range. Such high bandwidths are seldom needed in instrumentation systems, where 5 Hz–125 MHz signal bandwidths are more commonly used.

As in the case of transmission lines, the FOC largely behaves like a low-pass filter with a transport lag. The longer a given FOC, greater the transmission lag and the attenuation at the receiver, and lower the system bandwidth for data transmission. A practical criterion for FOC bit rate can be established by considering how close in time two input

**FIGURE 9.38**

Section through a Hewlett-Packard, model HFBR-2204 fiber optic receiver. A spherical lens concentrates the light on the sensor. (Source: Courtesy of HP)

current pulses to the LED or LAD source can be in order to just resolve the two pulses as a '101', instead of a 111 or 110, etc., at the photodiode output.

FOCs are made from glasses or plastic. Fiber diameters range widely with cable application. In telecommunications, FOCs are rapidly replacing copper twisted pairs for all but local subscriber lines. The reason why FOCs are far less expensive than wire cables is that they are immune to electrical and electromagnetic interference, and they have bandwidths greater than conductive transmission lines. A basic FOC, serial communications receiver is shown schematically in Figure 9.38.

9.12.1 Fiber Optic Cable Basics

Figure 9.39 illustrates a typical step index FOC. The core of a step index FOC has a uniform (in x , y and z coordinates) optical index of refraction, n_1 . The FOC core carries the light energy used to signal data. The light energy propagates in the core in a variety of modes, as we will see. In order to understand how light is entrained in the core, we need to review some elementary optical theory governing refraction and reflection at plane interfaces of two dielectric media having different indices of refraction. The index of refraction of an optical medium is the ratio of the speed of light *in vacuo* (free space) to the speed of light in the medium. The refractive index of the FOC core is:

$$n_1 = c/v_1 \quad (9.56)$$

The refractive index of the cladding material surrounding the core is:

$$n_2 = c/v_2 \quad (9.57)$$

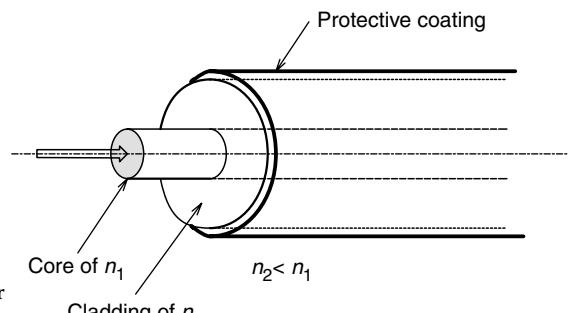
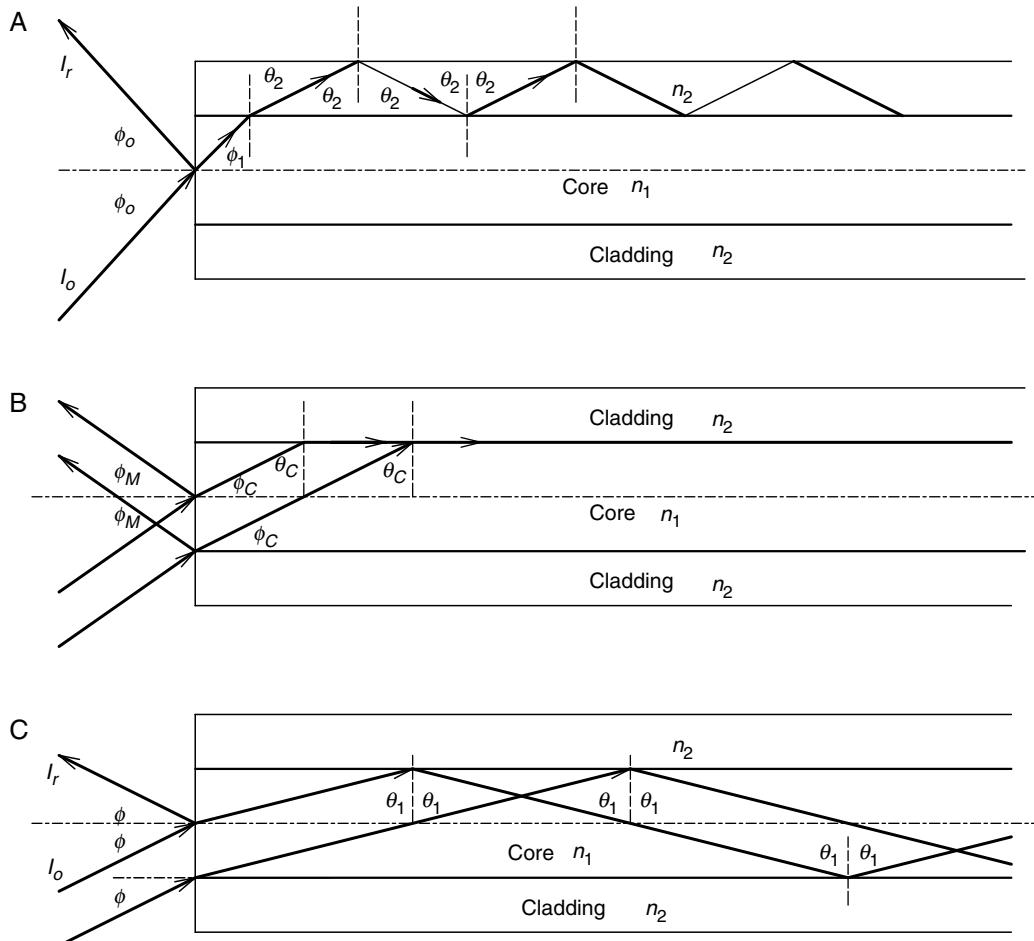
**FIGURE 9.39**

Diagram of the cut-back end of a step index fiber optic cable.

**FIGURE 9.40**

A. Lateral view of a step index, FOC. Here, an incident ray enters the cable at a very steep angle, ϕ_o . The angle of incidence of the internal ray on the core/cladding boundary is less than the critical angle, ϕ_c , and most of its energy enters the cladding and is trapped there. B. Critical angle condition for the core/cladding boundary. C. Now, the ray is sufficiently normal to the end of the core so that most of its energy propagates down the core by alternately reflecting off the core/cladding boundary. (The situation is made more complex by the cylindrical geometry.)

Obviously, $n_0 \approx 1$ for air. In a FOC, $n_1 > n_2 > 1$ by design. Figure 9.40 shows what happens to light rays in the core incident on the core/cladding interface, modelled here as a plane surface for simplicity. A critical angle, θ_C , exists such that if the angle of incidence, $\theta_1 < \theta_C$, then some fraction of light energy enters the cladding and is trapped and dissipated in it. The angle relations for this case are given by the well-known Snell's law:

$$\frac{n_1}{n_2} = \frac{\sin(\theta_2)}{\sin(\theta_1)} \quad (9.58)$$

A critical angle, $\theta_1 = \theta_C$, exists such that $\theta_2 = 90^\circ$. In this case, equation 9.58 gives us:

$$\theta_C = \sin^{-1}(n_2/n_1) \quad (9.59)$$

Rays in the core with $\theta_1 > \theta_C$ experience ‘total reflection’ at the core/cladding boundary and thus, most of their energy remains within the core, bouncing back and forth off the boundary as they propagate to the receiver end of the FOC. Such entrained rays are called *meridional rays* because they traverse the axis of the fiber.

It can be shown from detailed consideration of the Maxwell equations for EM waves incident on plane dielectric boundaries, that some portion of an incident ray’s energy will be reflected, even if $0 \leq \theta_1 \leq 90^\circ$, while some portion will be transmitted. The proportions reflected and refracted can be shown to depend on the polarization state of the incident light, on the indices of refraction of the two media, and on the angle of incidence. For example, the reflected fraction of unpolarized light normally incident on the end of a step index FOC is given by [Sears, 1949]:

$$I_r = I_i \frac{(n_1 - 1)^2}{(n_1 + 1)^2} \quad (9.60)$$

For $n_1 = 1.5$, $I_r / I_i = 0.040$, or 4% of the normally incident light energy does not enter the cable core. In passing from a higher to lower refractive index medium, there is a very steep increase in I_r / I_i as the angle of incidence approaches θ_C . This behavior is the reason why light traveling in the FOC core tends to remain in it and does not enter the cladding. On the other hand, there is a slow increase in I_r / I_i as the angle of incidence approaches and passes through θ_C as unpolarized light traveling in a low refractive index medium is incident on a plane boundary with a high refractive index medium. Thus, some fraction of the light trapped in the cladding can re-enter the core. The reflection *vs* refraction relations (Fresnel’s formulae) are given in detail in Sears (1949).

For the special case of light entering on the z-axis of the FOC ($\phi = 0^\circ$, $\theta_1 = 90^\circ$), it takes the axial ray $\tau_a = Ln_1/c$ seconds to reach the end of the FOC at $z = L$ meters. From simple geometry, photons following a bouncing path can be seen to take

$$\tau_{\theta_1} = Ln_1/[c \sin(\theta_1)] \text{ s} \quad (9.61)$$

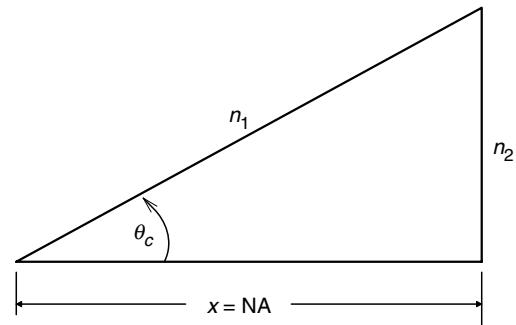
to reach the end of the cable. The difference between τ_{θ_1} and τ_a is called the *modal dispersion*. Modal dispersion causes broadening of transmitted pulses and thus, is a factor reducing the transmission bandwidth of multimode, step index FOCs. The reader who is interested in a rigorous treatment of modal propagation in plane (rectangular) waveguides using Maxwell’s equations should consult texts on FOCs by Cheo (1990), Gowar (1984), or Keiser (1983). The mathematical treatment of modal propagation in cylindrical waveguides (FOCs) is complicated by the fact that all field components are coupled. The text by Cheo (1990) gives such analyses.

Another geometrical optical consideration of FOCs is the numerical aperture (NA) of a cable. The NA of a FOC is related to the largest value of the ray entrance angle to the core from space, ϕ_M , that will permit a ray to propagate down the core. In other words, a ray entrance angle greater than ϕ_M results in $\theta_1 < \theta_C$, and nearly all its energy is coupled into the cladding and lost. Total internal reflections occur for $\phi < \phi_M$ and modal propagation occurs. At the critical entrance angle, we have, by Snell’s law:

$$n_0 \sin(\phi_M) = n_1 \sin(\phi_C) \quad (9.62)$$

Also, at the critical angle in the core, from equation 9.59, we can write:

$$\sin^2(\theta_C) = (n_1/n_2)^2 \quad (9.63)$$

**FIGURE 9.41**

A geometrical interpretation of a fiber's numerical aperture (NA).

By trigonometric identity,

$$\cos(\theta_C) = [1 - \sin^2(\theta_C)]^{1/2} \quad (9.64)$$

and from the right triangle,

$$\sin(\phi_C) = \cos(\theta_C) \quad (9.65)$$

Substituting equations 9.62, 9.63 and 9.65 into 9.64, we obtain an expression for the step index FOC's NA (note $n_o = 1$):

$$NA = \sin(\phi_M) = (n_1^2 - n_2^2)^{1/2} \quad (9.66)$$

Figure 9.41 shows a geometrical interpretation of NA in terms of θ_C , n_1 and n_2 . Note that the FOC NA is not a function of the core diameter.

FOCs can be subdivided into multimode and single mode fibers. Multimode fibers generally have core diameters ranging from 50 μm to as great as 1000 μm . It is physically easier to couple light into and out of multimode fibers because of their greater core diameters. Single mode FOCs have core diameters of 3–12 μm and consequently, I/O coupling requires great precision. The maximum number of modes in a planar, dielectric waveguide (as a first approximation to a cylindrical FOC) can be shown to be approximated by [Stark, Tuteur and Anderson, 1988]:

$$M_{MAX} = 2d(\text{NA})/(\lambda n_1) - 1 \quad (9.67)$$

where d is the core diameter, λ is the wavelength of light on the core, and NA is the fiber's numerical aperture, given by equation 9.66. From equation 9.61, the maximum modal dispersion can be shown to be:

$$\Delta\tau = (\tau_c - \tau_a) = \frac{Ln_1}{cn_2} (n_1 - n_2) s \quad (9.68)$$

While single mode fibers do not suffer from modal dispersion, they, as well as multimode fibers, exhibit material dispersion, which is caused by the fact that n_1 is not a constant, but is, in fact, a function of the wavelength of the light in the core. Cheo (1990) shows that the per unit length difference in group delay, $\Delta\tau_G$, for transmitted components separated

by $\Delta\lambda$ in wavelength is given by:

$$\Delta\tau_G = \frac{\lambda\sigma_S}{c} \frac{d^2n}{d\lambda^2} \text{ s/m} \quad (9.69)$$

where σ_S is the RMS spectral width of the source, given by:

$$\sigma_S = \sqrt{\int_0^\infty (\lambda - \lambda_o)^2 S(\lambda) d\lambda} \quad (9.70)$$

$$\lambda_o = \int_0^\infty \lambda S(\lambda) d\lambda \quad (9.71)$$

$S(\lambda)$ is the spectral distribution of the source. Thus, the material dispersion is proportional to the source's spectral width and the second derivative of $n_1(\lambda)$ around the mean wavelength, λ_o . Interestingly, $\Delta\tau_G \rightarrow 0$ at λ of around 1300 nm for silica fiber cores. Hence, sources operating around this wavelength offer improved bandwidth for single mode FOCs.

A detailed, theoretical derivation of FOC bandwidth, based on a two pulse (101), intersymbol interference under noise free conditions is given by Stark, Tuteur and Anderson (1988). These authors assumed a source λ_o of 1000 nm, the second derivative of the phase constant, $\beta''(\lambda_o) = 3 \times 10^{-26} \text{ s}^2/\text{m}$ and an interpulse interference ratio, $\delta = 0.2$, at a point midway between the two received, unit pulses. (The tails of the unit height pulses sum to 0.2 halfway between them.) Their cable bandwidth, B , defined as the reciprocal of the interpulse interval required to make $\delta = 0.2$, increases with the spectral purity of the source and decreases with cable length. Figure 9.42, adapted from Stark *et al*, illustrates FOC bandwidth B , as a function of source bandwidth, $\Delta\lambda$ around 1000 nm, and cable length.

The *graded index fiber* (GRIN) represents a design which mitigates the effect of modal dispersion found in step index fibers and yet, keeps the large core diameter permitting simpler I/O connections. Figure 9.43 illustrates a median section of a graded index FOC

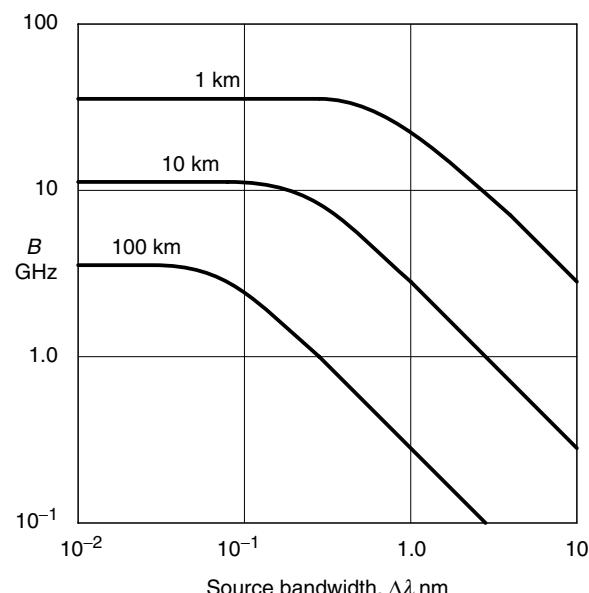
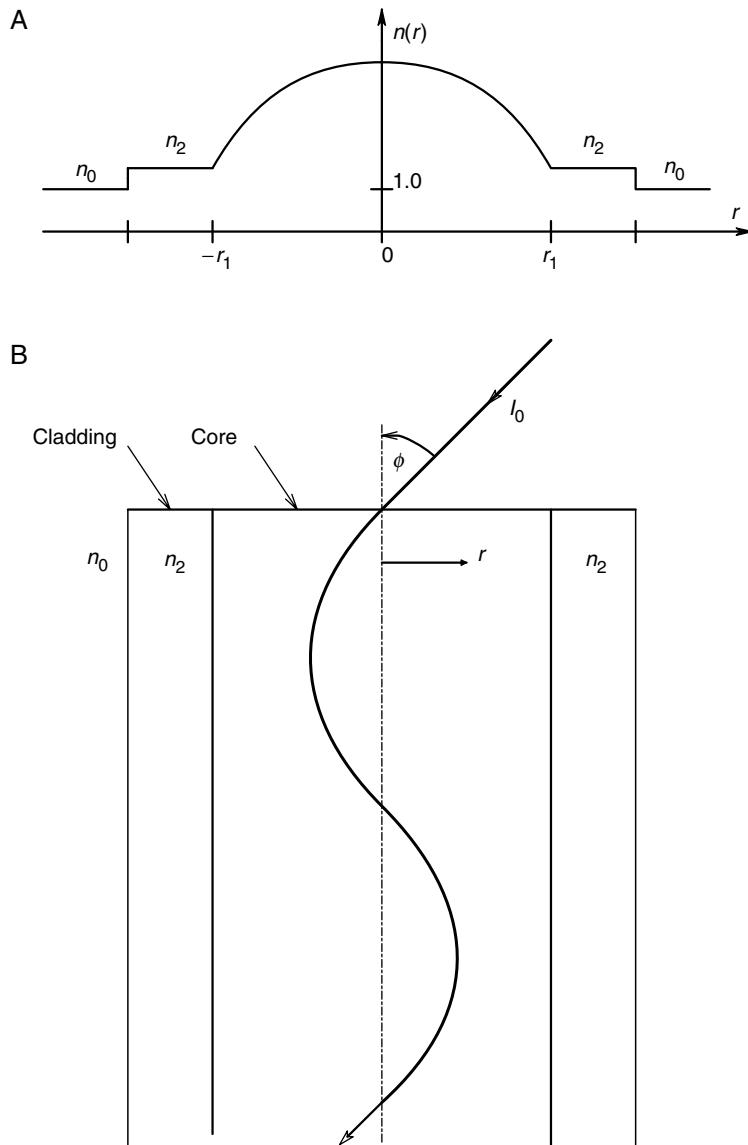


FIGURE 9.42

Plots of FOC signal bandwidth, B , in GHz as a function of FOC length and the spectral width of the 1000 nm source, $\Delta\lambda$. Note that as the optical source's bandwidth increases, the permissible transmission rate decreases, underscoring the value of IR laser diodes as FOC system sources. (Source: Adapted from Stark *et al*, 1988)

**FIGURE 9.43**

A. Plot of refractive index of a graded index optical fiber (GRIN) *vs* fiber radius. B. Path of an obliquely entering ray as it propagates down the GRIN FOC.

and its refractive index profile. There are three principal GRIN core index profiles—parabolic profile (equation 9.72), α profile (equation 9.73) and hyperbolic or ‘selfoc’ profile (equation 9.74).

$$n(r) = n_{1(MAX})(1 - \varepsilon r^2)^{1/2}, \quad 0 \leq r \leq r_1 \quad (9.72)$$

$$n(r) = n_{1(MAX)}[1 - 2\Delta(r/r_1)^\alpha]^{1/2} \quad (9.73)$$

$$n(r) = n_{1(MAX)} \operatorname{sech}(r\Delta/r_1) \quad (9.74)$$

$$\Delta = (n_{1(MAX)} - n_2)/n_{1(MAX)} \quad (9.75)$$

Modal dispersion is greatly reduced in GRIN FOCs because, instead of reflecting at the core/cladding boundary of a step index fiber, oblique rays are bent so that they follow a smooth, periodic pathway down the core, as shown in Figure 9.43B. An axial ray travels down the shortest path in the center of the core at the slowest velocity, $v = c/n_{1(MAX)}$. An oblique ray travels a longer path, most of which lies in the higher velocity periphery of the core. Thus, both oblique and direct rays tend to arrive at the receiving end at the same time, minimizing modal dispersion.

The differential delay between the fastest and slowest modes in an α -profile, GRIN fiber, for a monochromatic (LAD) source can be shown to be:

$$\Delta\tau_G = (n_{1(MAX})^2)/8c \text{ s/m} \quad (9.76)$$

where

$$\alpha = 2(1 - 1.2\Delta) \quad (9.77)$$

for minimum intermodal dispersion [Stark, Tuteur and Anderson, 1988].

9.12.2 Semiconductor Sources and Detectors Used in FOC Data Transmission

Two principal sources are used for fiber optic systems, LEDs and laser diodes (LADs). LEDs used in FOC systems are generally double heterojunction designs, as shown in Figure 9.44. A thin recombination layer of *p*-doped GaAs is sandwiched between thicker

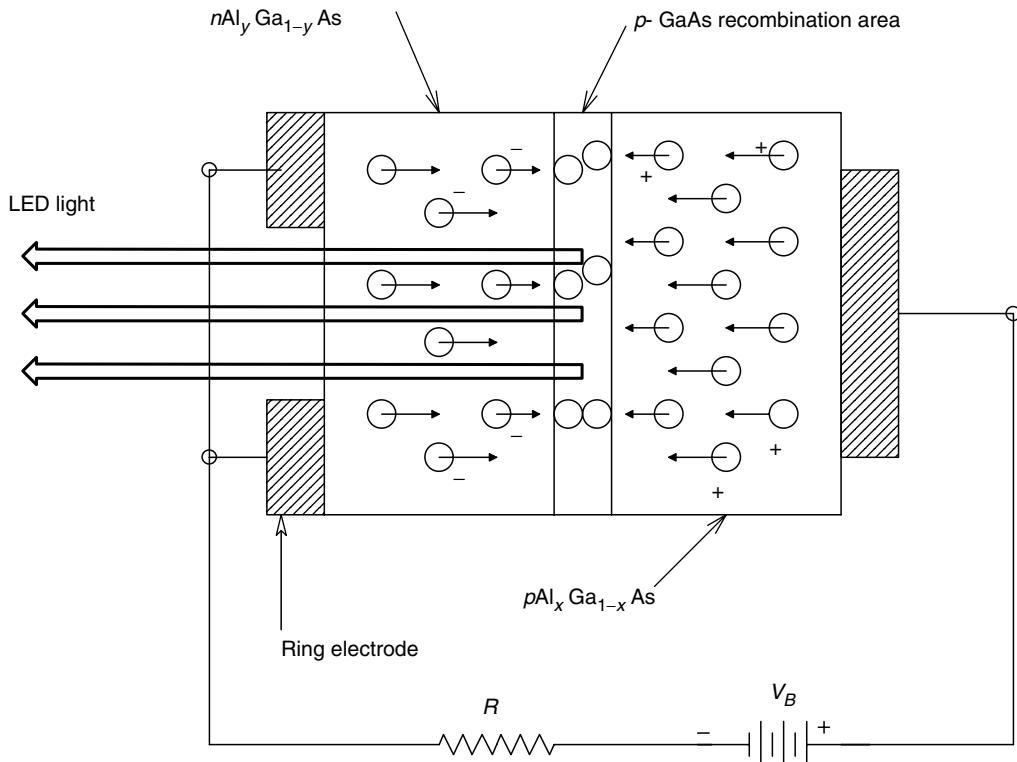
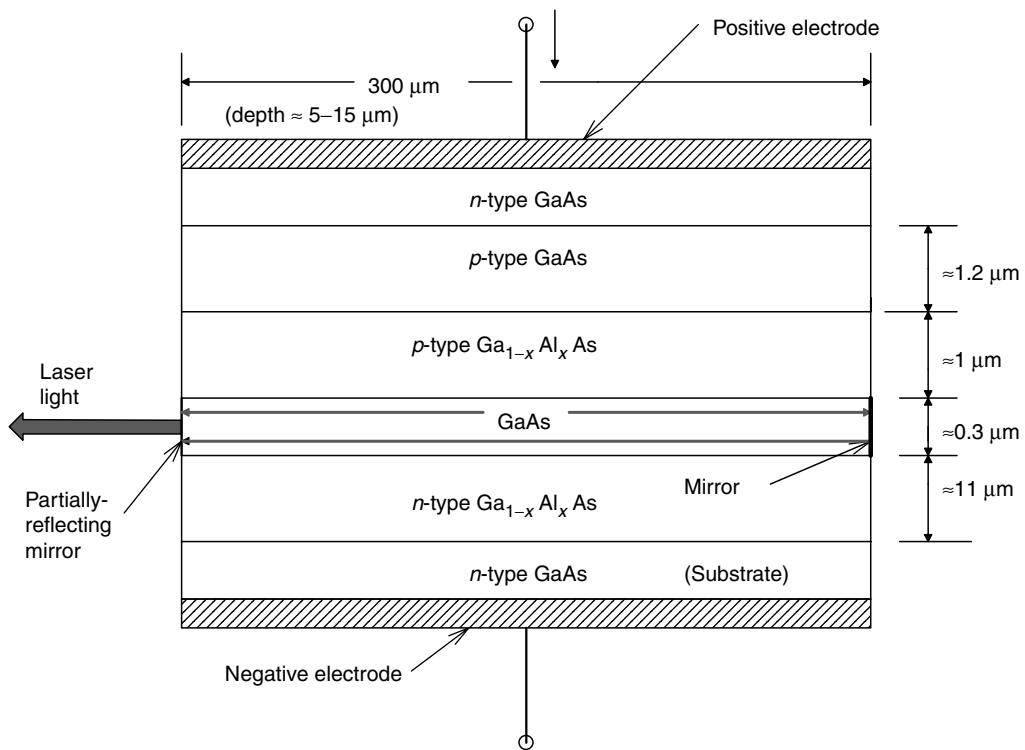


FIGURE 9.44

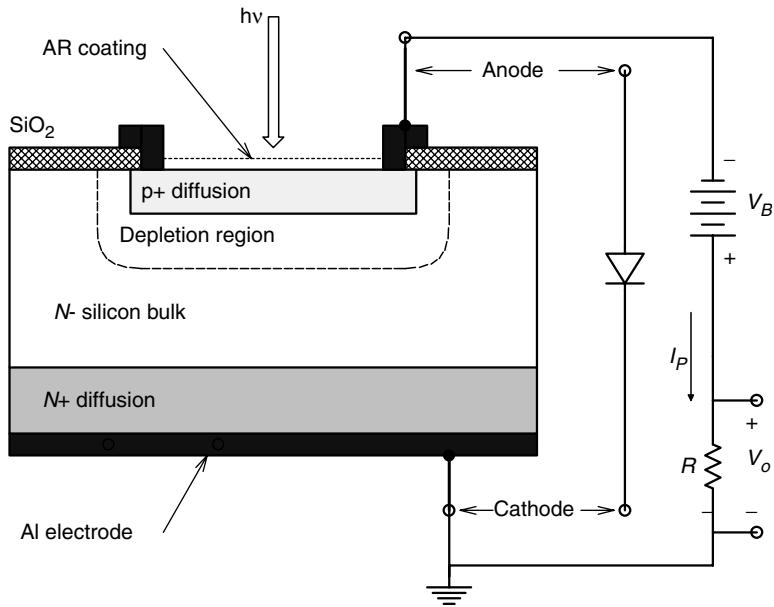
Layer cake diagram of a double heterojunction AlGaAs LED source.

**FIGURE 9.45**

Layer cake model of a stripe geometry, double heterojunction, GaAs laser diode. (Source: With permission, Stark, et al., 1988)

layers of p -type $\text{Al}_x\text{Ga}_{(1-x)}\text{As}$ and n -type $\text{Al}_x\text{Ga}_{(1-x)}\text{As}$. The Hewlett-Packard HFBR-1204 high efficiency fiber optic transmitter is a good example of a commercial, double heterojunction FOC LED. This LED emits at 820 nm and can supply 316 μW peak optical power into a 100 μm core diameter, step index, FOC. Its 10 to 90% rise and fall time is 11 ns. The HP HFBR-1100 general purpose LED transmitter uses an InGaAsP LED which emits at 1300 nm with a spectral bandwidth of 130 nm. It has an optical rise time of 1.2 ns and a fall time of 2.3 ns. It supports a maximum signaling rate of 200 MBd. Modulation of the light intensity is accomplished by switching the diode's current on and off.

Laser diodes are used in single mode FOC systems designed to operate at extremely high data rates. An example of a double heterostructure LAD is shown schematically in Figure 9.45. As long as the LAD's supply current is held constant along with its temperature, its light output remains constant and of constant spectral purity. The envelope of an LAD's output line spectrum narrows with increased output power, and its peak shifts to a slightly longer wavelength (e.g. a spectrum peak at 830 nm at 4.6 mW output shifts to a peak at 833 nm at 12 mW output). This particular LAD had about 2.5 spectral lines/nm [Cheo, 1990]. Cheo describes a constricted double heterostructure LAD which produces a single, narrow, spectral output peak at about 837.5 nm at an input current of 100 mA. The two major types of semiconductor photodetectors used in FOC systems are the PIN photodiode and the avalanche photodiode. All photodetectors must have high quantum efficiency at the operating wavelength, high responsivity, rapid response time and small size. The quantum efficiency, η , is defined as the number of carrier pairs (hole + electron) generated per incident photon. Ideally, $\eta = 1$. In practice, we

**FIGURE 9.46**

Layer cake diagram of a PIN photodiode.

see $0.3 \leq \eta \leq 0.95$. The responsivity, $R(\lambda)$, is defined as the ratio of primary photocurrent to incident optical power.

$$R(\lambda) = \eta e/hv \quad (9.78)$$

For an ideal photodiode, $\eta=1$ and $R=\lambda/1.24 \text{ A/W}$. Typical values for $R(\lambda)$ are $0.65 \mu\text{A}/\mu\text{W}$ for Si at 800 nm , and $0.45 \mu\text{A}/\mu\text{W}$ for Ge at 1300 nm [Stark, Tuteur and Anderson, 1988].

The I in PIN stands for *intrinsic* semiconductor material, sandwiched between a *p*- and *n*-layer. The PIN photodiode is generally operated in reverse bias (Figure 9.46). The photocurrent, I_P , is added to the dark (leakage) current, I_L . The objective is to resolve a small change in the total diode reverse current (due to a pulse of absorbed photons) above the shot noise and thermal noise in the diode circuit. PIN photodiode responsivity is a function of the materials from which it is made. Silicon PINs operate effectively in the range $200\text{--}1000 \text{ nm}$, with response rising with shorter wavelengths. Germanium PINs are useful in the range $1000\text{--}1800 \text{ nm}$. PINs made from $\text{Al}_x\text{Ga}_{(1-x)}\text{AsSb}$ have η s as high as 80% in the $1.8\text{--}2.5 \mu\text{m}$ range [Cheo, 1990]. The HP HFBR-2100, 1300 nm , PIN, FOC detector has 1 ns output rise and fall time, enabling it to be used in a 1300 nm , 200 MBd , FOC system.

The avalanche photodiode (APD) is a *p+ipn*, four-layered structure, shown schematically in Figure 9.47. Its increased sensitivity results from an internal, intrinsic gain mechanism that greatly increases the number of mobile carriers above the original photo induced holes and electrons. The dc electric field conditions in the reverse biased APD are such that the depletion region reaches into the *i*-region. The electric field at the *pn* junction is only 5–10% lower than that necessary for avalanche breakdown of the junction. Mobile photo-electrons and holes gain enough kinetic energy in the high *E* field region of the *p* layer to produce additional mobile electrons and holes through inelastic collisions, which impart enough energy to promote new carriers to the conduction band.

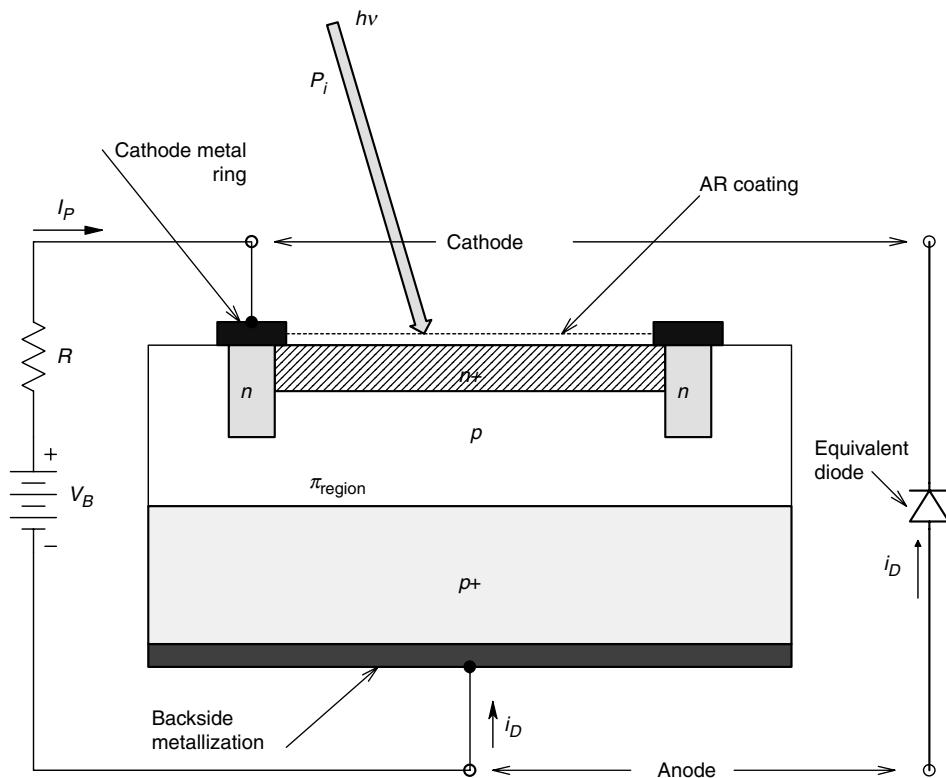


FIGURE 9.47
Layer cake diagram of an avalanche photodiode.

These secondary carriers, in turn, can gain enough kinetic energy to cause further ionization, etc. Thus n primary photo-electrons can give rise to αn electrons at the diode output. The n primary photo-holes result in αn holes, contributing to I_P . The net APD current multiplication factor, M , is a function of the internal electric field distribution, diode geometry, doping and the wavelength of the incident photons. Cheo (1990) gives a good analytical treatment of APD behavior and noise. APDs are made from Si, Ge, InGaAsP, etc., depending on the desired operating wavelength. At wavelengths below 1 μm , Si APDs have been built with net current gains of about 100, quantum efficiencies approaching 100% and response times around 1 ns. In general, APDs exhibit faster rise times than PIN diodes. However, each requires careful setting of its operating voltage and constant operating temperature.

9.12.3 FOC Systems

The instrumentation system designer who wishes to interconnect groups of instruments, such as a VXI cluster, over a long distance to a controlling computer by fiber optic link no longer has the burden of assembling custom components to build the link. Hewlett-Packard markets a wide variety of plastic and glass FOC components, including single and double (duplex) FOCs, LED transmitters and PIN receivers and signal conditioners. HP also offers FOC system evaluation kits for designers to test feasibility with 1000 μm core plastic cable and 5 MBd, TTL compatible transmitters and receivers. HP FOC link equipment ranges from 125 m, 40 kBd, TTL systems to 2 km, 200 MBd, ECL systems. This FOC equipment is described in the *HP Optoelectronics Designer's Catalog*.

9.13 Virtual Instruments

9.13.1 Introduction

With the development of high speed bus architectures, and special Analog Digital plug-in interface cards for PCs and laptop computers, there has been a trend to use the power of computer software to develop *virtual instruments* (VI) whose controls and displays are presented on a computer screen, and whose operation and settings are controlled by the computer's keyboard and mouse. Rather than purchase several, expensive, dedicated, stand alone instruments, it is more economical to purchase a computer with a multichannel, A/D interface card of appropriate speed and bit resolution, or an outboard PXI system, and the VI software to control, process and display the measured quantities.

National InstrumentsTM (NI) company has pioneered the development of computer run virtual instruments and certain digital data processing algorithms used with them. Interface A/D cards and outboard signal conditioning and A/D systems, such as PXI modules connected to the computer through RS-232D, IVI-C or GPIB buses, are offered by several vendors, including CyberResearch®, National Instruments and Keithley Instruments, to mention a few.

An example of a typical VI is a true RMS voltmeter. A time varying analog signal, such as noise, is first anti-alias filtered, then A-to-D converted at a suitable rate, then M digital samples in each of N data sample epochs are each squared, averaged and the average of the squares square rooted to generate numerical estimates of the true RMS value of the signal. For noise and noisy signals, the N averages of the squares are averaged, then square rooted to reduce the variance of the RMS estimate. The RMS voltage is presented on the computer screen as the display of a VI RMS voltmeter shown in a window. The operator can adjust the VI meter's range, the AAF's cutoff frequency, its sampling rate, the epoch length, M , and the averaging algorithm by setting the knobs and switches on the VI display with the mouse.

Another example of a user implemented VI instrument is a *pulse height window*. Such windows are used in experimental neurophysiology and nuclear physics. The VI software displays the upper and lower dc thresholds, φ_1 and φ_u , respectively, and the signal with pulses to be discriminated as a moving (strip-chart type) display. The analog signal is anti-alias low-pass filtered, sampled, $A \rightarrow D$ converted, and digitally low-pass filtered to remove some noise. The window output is a time record of when pulses on the signal cross φ_1 , do not cross φ_u and then, cross below φ_1 . The VI operator can adjust the thresholds, the AAF cut-off frequency, the sampling rate and the digital LPF parameters. The window output point process, $p_w(t) = \sum_{k=1}^{\infty} \delta(t - t_k)$, is an array of times when pulses lie within the window. $p_w(t)$ can be further processed to generate a step wise, *instantaneous frequency record* (t_k is the time of occurrence of the k th window discriminated pulse relative to a local time origin). From the point process, $p_w(t)$, the computer calculates the stepwise, continuous variable, $F_I(t)$. $F_I(t)$ is called the *demodulated instantaneous pulse frequency* of the discriminated pulses; the process defined by equation 9.79 is called instantaneous pulse frequency demodulation (IPFD).

$$F_I(t) = \sum_{k=2}^{\infty} r_k [U(t - t_k) - U(t - t_{k+1})], \quad k = 1, 2, 3, \dots, \infty \quad (t_0 \text{ occurs at } t = 0) \quad (9.79)$$

where $r_k \equiv (1/(t_k - t_{k-1}))$ is the k th element of instantaneous frequency of $p_w(t)$, $k = 2, 3, \dots, \infty$ and $U(t - t_k)$ is the *unit step*; 0 for $t < t_k$ and 1 for $t \geq t_k$. Note t_1 is the

time the first pulse occurs after $t = 0$ and $(t_2 - t_1)$ is required to define the initial element of instantaneous frequency in $F_I(t)$.

NI's LabVIEW™ offers a number of prepackaged VI modules. Signal processing VIs include FFTs, cross- and auto-power spectral analysis, Hilbert transforms and Joint Time-Frequency Analysis (JTFA) [Northrop, 2003]. Time series analysis VIs include real convolution, deconvolution, auto- and cross correlation, and derivative and integral filtering, etc. Oscilloscopes and signal generation VIs are also available. VI modules can be user linked to perform complex signal analysis and real time control tasks, including motion control. External PXI input signal conditioning modules exist for RF signals, accelerometers, velocimeters, tachometers, thermocouples, pressure sensors, force and torque sensors, phase sensitive rectifiers for carrier systems such as bridges, proximity sensors, photosensors, microphones and other sensors, permitting custom VI systems to be built that allow sound and vibration analysis of machinery and appliances, acoustics signal processing, correlation studies on noise and vibrations, etc.

As if there were not enough computer buses to consider in instrumentation, the NI-sponsored, Interchangeable Virtual Instruments (IVI) Foundation, founded in 1998, has defined two unified bus architectures for VI drivers, the IVI-C and the IVI-COM. IVI-C is based on ANSI-C and can drive over 300 compatible instruments from NI, Keithley, Pickering, etc. IVI-C source code is generally available. IVI-COM is based on Microsoft COM technology and can drive ca. 50 instruments, mainly from Agilent. IVI-COM source code is not generally available. Both IVI architectures permit remote operation of compatible instruments; their input settings are managed as VI images on the host computer and the data from the instruments is sent to the host for display, processing and logging. Thus, any analog filtering or signal processing operations, anti-alias filtering, A-to-D function and peripheral DSP is done remotely, and each individual IVI compatible instrument communicates its output to the host [National Instruments, 2003a; National Instruments, 2003b; National Instruments, 2003c].

9.13.2 PXI Systems

PXI stands for PCI eXtensions for Instrumentation. The PXI standard defines a modular instrumentation platform (hardware modules and operating software) designed specifically for robust measurement and automation applications. PXI peripheral modules are used as remote computer controlled front ends for VIs. PXI modules generally have only input signal connectors and status LEDs on their narrow front panels. There are generally no control knobs or switches.

Computer communication with the PXI box and its modules can be through the 132 MB/s standard *PCI bus*. The PXI system has over 1000 products currently available from various vendors, including NI. These products include 'crates', sensor interfaces, power supplies, analog signal conditioners, DSP, and motion control driver output modules, etc. [National Instruments *PXI Tutorial*, 2004]. The mechanical aspects of PXI and CompactPCI (connectors) are governed by Eurocard specifications, (ANSI 310-C, IEC-297 and IEEE 1101.1), which have a long history of application in industrial environments. (The reader interested in PXI hardware should consult the PXI Systems Alliance document, *PXI Hardware Specification* (v.4, February 2003) available online.)

National Instruments has also developed their own transparent, MXI-3 bus protocol to communicate with PXI peripheral modules. A PCI board is in the desktop computer, and slot 1 of the PXI crate has the peripheral MXI-3 interface card; copper or fiber optic communication is used. 132 MB/s peak and 84 MB/s data rates are possible using the PCI/MXI-3 bus. Ethernet/internet, SCSI, GPIB, VXI, VME, RS-232D and wireless

communication and control of PXI peripherals is also supported by NI (2004). Two or more PXI crates with peripheral cards can be linked by PXI-to-PXI MXI-3 connections. The PCI specification allows for up to 254 buses connected in a system via PCI-to-PCI bridges such as MXI-3.

9.13.3 Summary

Virtual instruments have found great application in science, industry, aerospace and medicine. They permit remote signal acquisition and conditioning modules to be located near the sensors, and managed remotely by a host computer. The host presents data in a VI window display that allows the operator to adjust input signal conditioning parameters and observe results on real time. In addition, the host can store incoming data on a hard disk or a CD-RW ROM disk. Links to and from the peripheral modules can be copper, fiber optic, or RF. A variety of bus protocols are available for communicating settings and data to and from modules.

VI technology has a large and growing application in engineering education. It offers an economical means to implement oscilloscopes, meters, bridges, signal generators and control signals without the expense of purchasing complete, dedicated instruments. The host computer can do spectral analysis on signals by FFT and other DSP routines, like cross correlation of two signals.

Special VI systems are used in biomedical engineering labs for the same reasons; economy and simplicity. Students can observe ECG, EEG and EMG waveforms and other physiological signals such as muscle force, blood pressure, pulse oximetry signals, etc. using medical isolation grade peripheral signal conditioners.

9.14 Chapter Summary

In this chapter, we introduced the primary limitations of converting signals from analog to digital format, namely the introduction of an equivalent broadband noise due to quantization, and the loss of high frequency information due to aliasing. Quantization was seen to be caused by the numerical round-off in representing analog data samples with digital numbers of finite bit length. Aliasing was shown to be the result of sampling a bandwidth limited signal too slowly (i.e. at a rate below twice the highest frequency in the signal's power density spectrum). Dithering was shown to be an effective way of mitigating the effect of quantization noise.

The R-2R digital-to-analog converter (DAC) architecture was introduced as representative of most modern DACs. Hold circuits were described as a means of 'freezing' a time varying input signal to an ADC while conversion takes place and also of maintaining a DAC output voltage until the output is updated. Hold dynamics were derived.

Six types of analog-to-digital converters were described and their applications were discussed relative to instrumentation systems. Mention was made of PC compatible data acquisition cards (containing ADCs, multiplexers and buffer memories).

Digital data buses were reviewed and the IEEE-488, GPIB, parallel instrumentation bus was described in some detail. Serial data communication links, including the venerable RS-232D bus and the relatively new, versatile USB were described and their limitations discussed. The CAMAC instrumentation packaging protocol and its related internal bus were covered, as well as the newer PXI system architecture.

In Sections 9.11 and 9.12, we discussed transmission lines used for digital data transfer, and showed how improper impedance termination can lead to voltage reflections on the line that cause errors and significantly slow data transmission rates. Long transmission lines pose an extra problem because line losses cease to be negligible, and there is not only delay in transmission, but transmitted pulses become rounded and attenuated as they traverse the line. The same problems of delay, attenuation and distortion are also seen on fiber optic cables, but FOCs have signal bandwidths which far exceed those of conventional, low loss, coaxial cables. Fiber optic cables too are immune to various forms of electromagnetic interference because of their dielectric waveguide properties.

Problems

9.1 An 18-bit ADC is used to convert an audio signal ranging over ± 1 V.

- (A) Find the size of the quantization step, q , required.
- (B) Find the RMS quantization noise associated with full scale operation of this ADC.
- (C) Let $v_{in}(t) = 1 \sin(\omega_0 t)$. Find the RMS SNR at the ADC output in decibels.

9.2 A 4-bit, $R-2R$, offset binary DAC is shown in Figure P9.2. Calculate the transfer characteristic between the input digital input, $\{d_3d_2d_1d_0\}$, and V_o . ($d_k = 0, 1$).

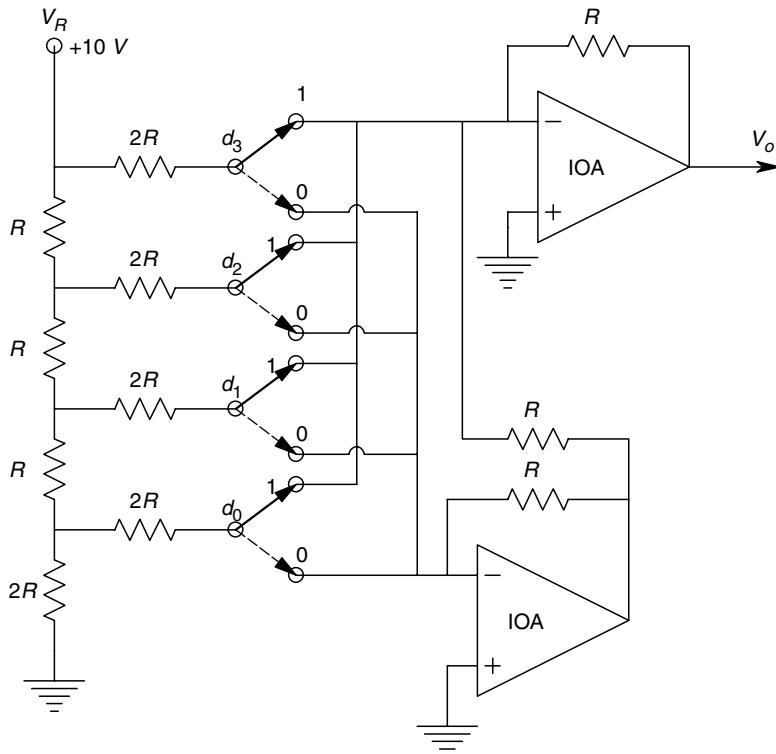


FIGURE P9.2

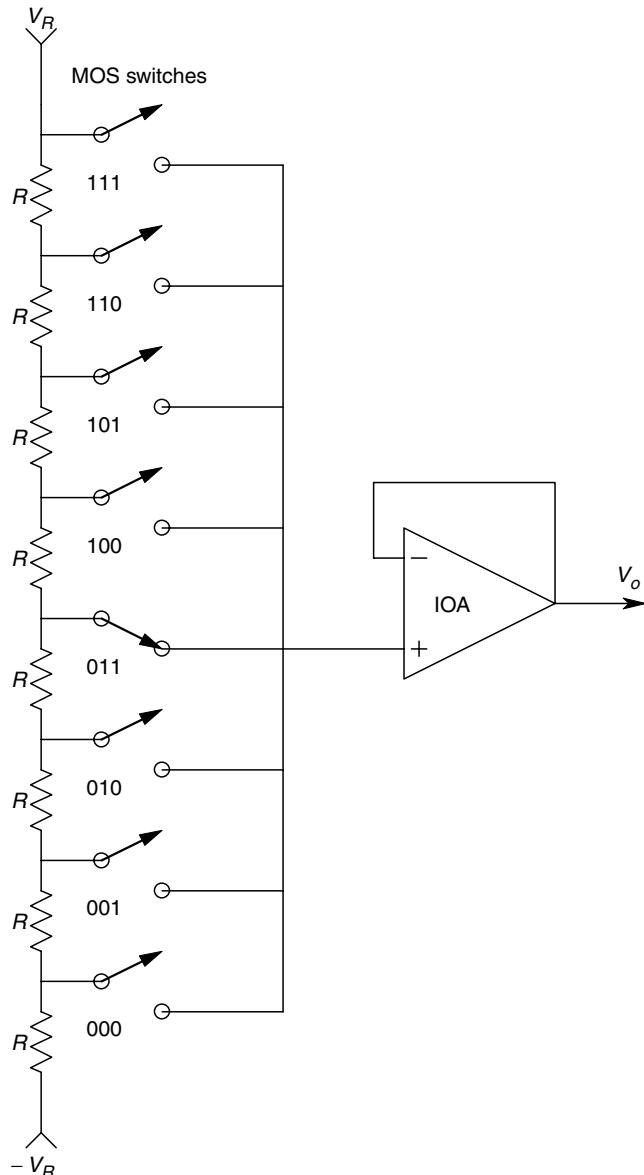


FIGURE P9.3

9.3 The simplified circuit of a 3-bit, bipolar, 'flash' DAC is shown in Figure P9.3. It is shown with the input, '011' active.

- (A) Make a table of V_o vs $\{d_2d_1d_0\}$.
- (B) Discuss the source of output 'glitches' in a practical implementation of this DAC when $\{d_2d_1d_0\}$ changes states.

9.4 Figure P9.4 shows the schematic of a design for a 3-bit, MOS, successive approximation ADC. Note that this complex design requires 2^N resistors and $[2^{(N+1)} - 2]$ MOS switches. The switches are closed for $d_k = HI$ or $\overline{d_k} = HI$. $V_R = 5V$. At the start of the SA cycle, the msb, d_2 , is HI and d_1 and d_0 are low. Show which switches are successively closed (and opened) in a conversion cycle for $V_S = -3V$ and $+4V$.

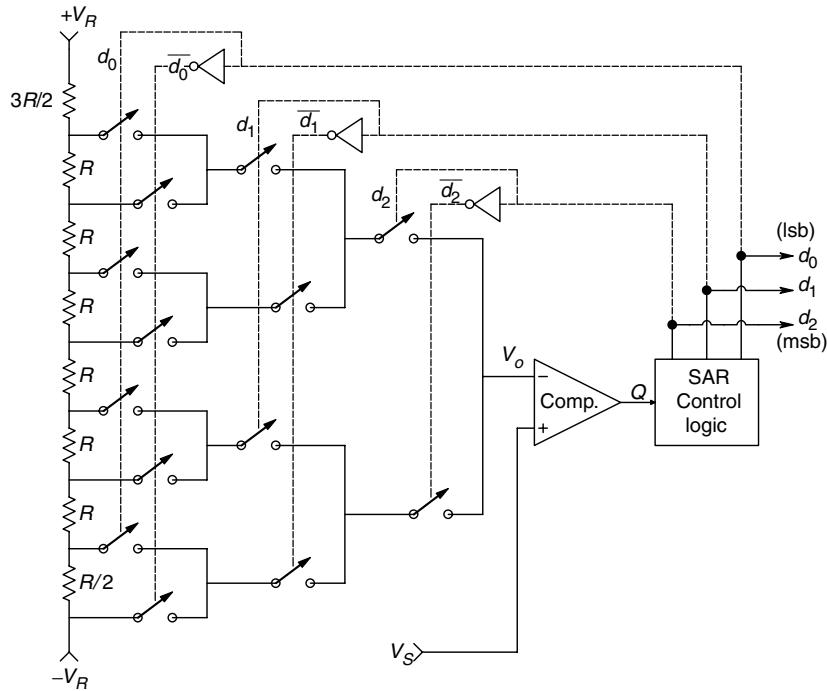


FIGURE P9.4

9.5 An FM voltage-to-frequency converter is shown in Figure P9.5. The one-shot output is periodic, narrow ($1\ \mu\text{s}$) pulses, which are counted by a digital counter (not shown) over 1 second. Each output pulse resets the integrator capacitor voltage to 0 V. Let $C = 1\ \text{nF}$, $R = 10\ \text{k}\Omega$, $V_\varphi = -5\ \text{V}$.

- Plot and dimension several cycles of $v_i(t)$ for $V_S = 1\ \text{V dc}$ and $V_B = 0$.
- Let $V_B = +5\ \text{V}$, plot and dimension the output frequency of the one-shot vs $0 \leq V_S \leq +10\ \text{V}$. Do not neglect the $1\ \mu\text{s}$ width of the output/reset pulses. Is $f = g(V_S)$ linear? What is $\Delta f / \Delta V_S$ at $V_S = +5\ \text{V}$?
- Now, with $V_B = 0$, determine the counter enable time, T_c , such that when $V_S = +10\ \text{V}$, the total count will be decimal 10^4 .

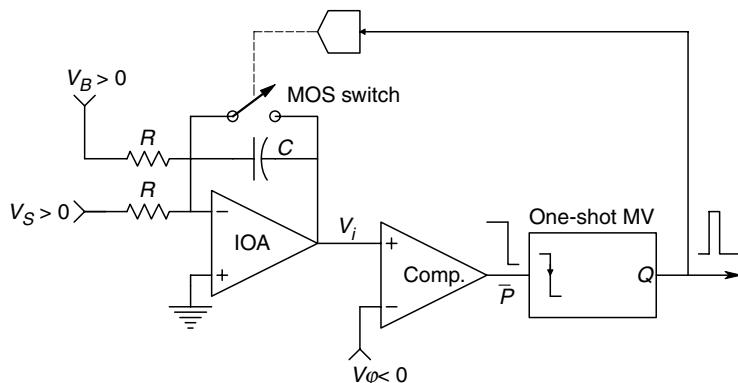


FIGURE P9.5

9.6 An 18-bit ADC with bipolar offset binary output is used to convert an audio signal ranging over ± 1 V.

- (A) Find the size of the quantization step, q , required.
- (B) Find the RMS quantization noise associated with full scale operation of this ADC.
- (C) Let $v_{in}(t) = 1 \sin(\omega_0 t)$. Find the RMS SNR at the ADC output in dB.

9.7 A 4-bit DAC using MOS switched, NPN, BJT current sources is shown in Figure P9.7. Derive an expression for $V_o = f[(d_3 d_2 d_1 d_0)]$, and other system parameters]. Let $R = 10.1 \text{ k}\Omega$, $R_E = 14.4 \text{ k}\Omega$, $BJT \beta = 100$, $V_R = +5 \text{ V}$, $-V_{EE} = -10 \text{ V}$.

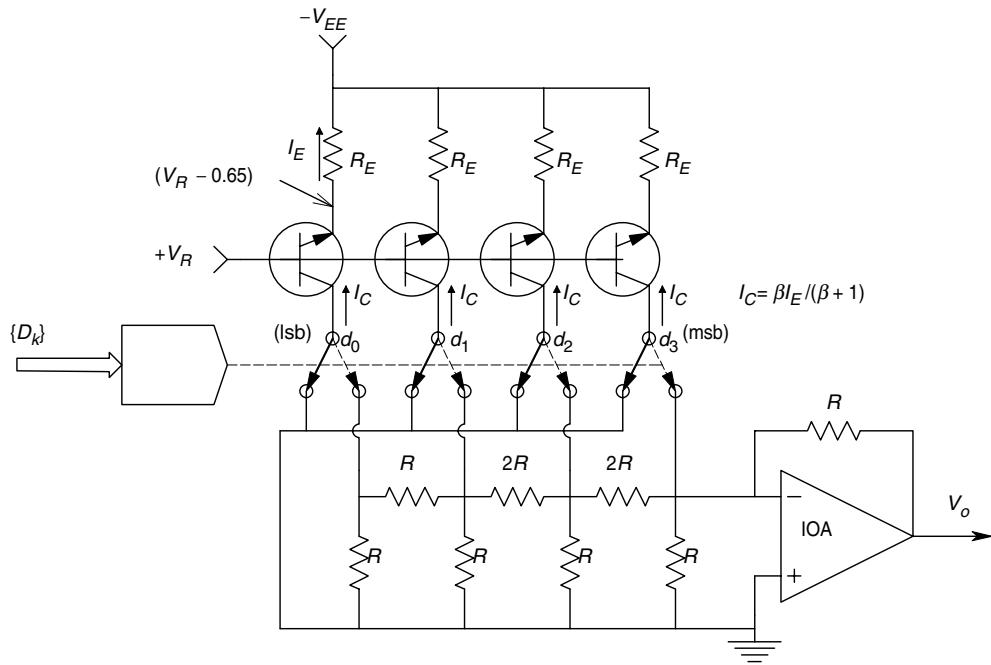


FIGURE P9.7

9.8 A unipolar, dual-slope, integrating ADC is used in a dc digital voltmeter. Figure P9.8 illustrates the system. At the convert command, the control logic resets the integrator and counter, then at local time $t=0$, allows the integrator to begin integrating the positive dc input voltage, V_x . At time $t=2^N T_c$, the counter is reset to zero. It is easy to see that at this time, $V_2 = -(V_x 2^N T_c) / (RC) \text{ V}$. The input switch is now connected to $-V_R$. The integrator now integrates $-V_R$ and ramps up to 0. When $V_2 = 0$, the comparator output Q goes from HI to LO, latching the counter count, M , to the output. Considering V_2 , it is clear that for one conversion cycle: $(V_x 2^N T_c) / (RC) = (V_R M T_c) / (RC)$ or, $M = (V_x 2^N) / (V_R)$. T_c is the clock period. Design the system (choose V_R , R , C , T_c) such that when $V_x = +10 \text{ V}$, $M = 1,000$ decimal. The total conversion time $(2^N + M) T_c$ should not exceed one second. Let $N = 10$.

9.9 Figure P9.9 illustrates a bipolar, dual-slope, integrating ADC for dc voltages. Note that $V_0 = -1/2(V_R + V_x)$. At the convert command, the capacitor voltage is set to zero by switch S_2 and the N -bit counter is set to zero. At local $t=0$, S_2 opens and S_1 connects the integrator to the negative voltage, V_o , for a fixed interval, $T_1 = 2^N T_c$ seconds. V_2 ramps positive, causing $Q = \text{LO}$. At T_1 , the integrator output equals: $V_2(T_1) = V_o(T_1) =$

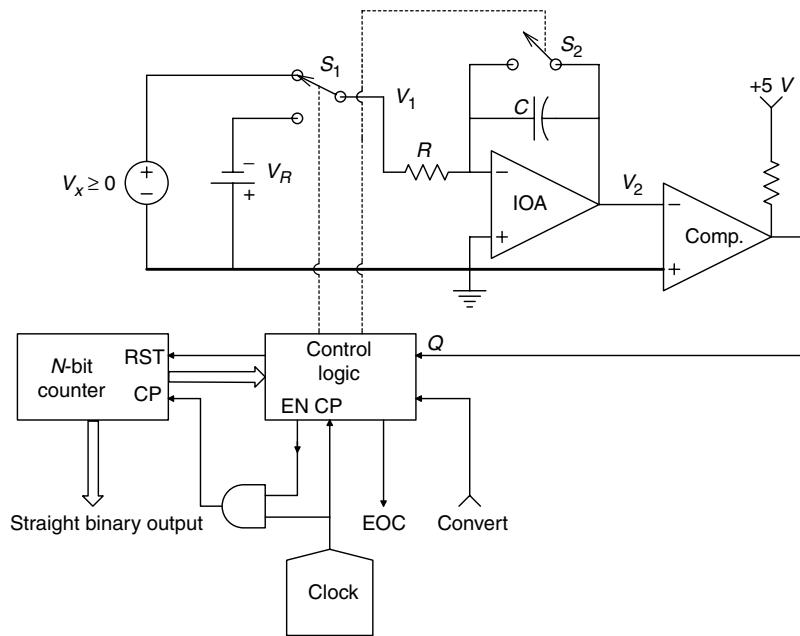


FIGURE P9.8

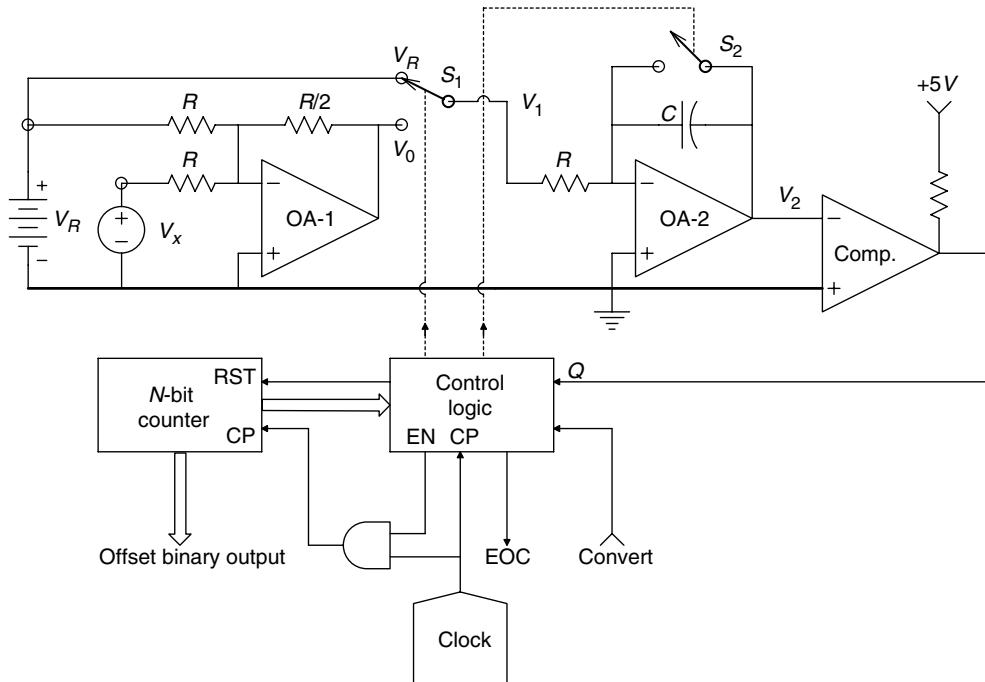


FIGURE P9.9

$+((V_R + V_x)2^N T_c)/(RC)$ V. When the counter reaches 2^N at T_1 , the counter is zeroed and S_1 connects the integrator to $+V_R$. The integrator output now ramps down until $V_2=0$. Then the comparator output Q goes HI, causing the AND gate to block the clock pulses. Also, when $Q \rightarrow \text{HI}$, the counter output, M , is latched to the output and

the counter is reset to await the next Convert command. The time it takes $V_2(T_1)$ to ramp down to zero is $T_2 = MT_c$ seconds.

- (A) Derive an expression for $M = f(N, V_X, V_R)$.
- (B) What kind of binary output code is produced by this ADC?

9.10 Consider the 4-bit resistive ladder DAC shown in Figure P9.10.

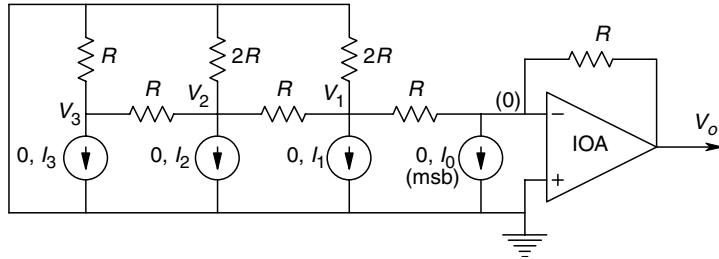


FIGURE P9.10

- (A) Use superposition to find an expression for $V_o = f(I_0, I_1, I_2, I_3)$. Let $R = 1 \text{ k}\Omega$, $I_0 = 1 \text{ mA}$ dc.
- (B) What values should I_1 , I_2 and I_3 have to make a binary DAC? Find the maximum V_o .

10

Introduction to Digital Signal Conditioning

10.1 Introduction

As we have seen in the preceding chapter, it is common practice to convey the analog conditioned output of a sensor to an analog-to-digital converter, and then pass that sampled data to a digital computer for further processing, storage and display. In this chapter, we introduce some of the basic digital signal processing (DSP) operations done on sampled data, and describe how they are accomplished. They include, but are not limited to, smoothing or low-pass filtering, integration, differentiation, computation of the RMS value of a data sequence, notch, or band reject filtering (to eliminate coherent interference in the signal bandwidth), computation of the discrete Fourier transform of a data sequence to characterize it in the (discrete) frequency domain, and interpolation and extrapolation of data sequences. Some of the DSP algorithms we have just mentioned exist as options in specialized data acquisition software packages and unless an instrumentation engineer is developing a system from scratch, there is often no need to write custom software for a specific, dedicated DSP task. One simply uses a packaged DSP routine on an input sequence to obtain the conditioned output sequence.

10.2 Digital Filters and the z -transform

Synchronous operations on sampled data can be described in the time domain as difference equations, or in the frequency domain using the complex variable, z . To illustrate how discrete DSP operations are described using the z -transform, we first consider a well-behaved, continuous analog signal, $x(t)$, defined for $0 \leq t \leq \infty$. For example, let $x(t) = A e^{-bt}$. As we have seen in chapter 9, sampling $x(t)$ is equivalent to the impulse modulation of $x(t)$. We can write the sampled $x(t)$, $x^*(t)$, in a more compact form, where T is the sample period and n is the sample number:

$$x^*(t) = \sum_{n=0}^{\infty} x(nT) \delta(t - nT) \quad (10.1)$$

if we take the Laplace transform of $x^*(t)$, we get:

$$X^*(s) = \sum_{n=0}^{\infty} x(nT) e^{-snT} \quad (10.2)$$

$x(nT)$ is the value of $x(t)$ at $t = nT$. Here, $x(nT)$ is treated as the area of the n th delta function, and the Laplace transform of the n th delta function is simply e^{-snT} . Equation 10.2 is easily changed to the open sequence form of the z -transformed $x(t)$ by letting the complex variable, z , be defined as:

$$z \equiv e^{sT} \quad (10.3)$$

Hence we can write:

$$X(z) = \sum_{n=0}^{\infty} x(nT)z^{-n} \quad (10.4)$$

For the exponential example given above, we can write:

$$X(z) = A[1 + e^{-bT}z^{-1} + e^{-2bT}z^{-2} + e^{-3bT}z^{-3} + \dots] \quad (10.5)$$

A power series of this type can be expressed in closed form as:

$$X(z) = \frac{A}{1 - e^{-bT}z^{-1}} = \frac{Az}{z - e^{-bT}} \quad (10.6)$$

A general, closed form definition of the z -transform of any Laplace transformable $f(t)$ can be derived from a consideration of the impulse modulation process, which is multiplication of $f(t)$ in the time domain with a periodic, unit delta function train, $P_T(t)$, beginning at $t = 0$. This multiplication gives $f^*(t)$. In the frequency domain, $F^*(s)$ is found by the complex convolution of $F(s)$ with the Laplace transform of the pulse train, $P_T(s)$. $P_T(s)$ can be written as:

$$P_T(s) = 1 + e^{-sT} + e^{-2sT} + e^{-3sT} + \dots = \sum_{n=0}^{\infty} e^{-nsT} \quad (10.7)$$

In closed form, $P_T(s)$ can be shown to be:

$$P_T(s) = \frac{1}{1 - e^{-sT}} \rightarrow \frac{1}{1 - z^{-1}} = \frac{z}{z - 1} \quad (10.8)$$

Now by complex convolution:

$$F^*(e^{-sT}) = \frac{1}{2\pi j} \int \frac{F(p)}{1 - e^{-sT}e^{pT}} dp \quad (10.9)$$

We use the definition of z above and treat the complex variable, z , as a constant in the convolution integral:

$$F_Z(z) = \frac{1}{2\pi j} \int \frac{zF(p)}{z - e^{pT}} dp \quad (10.10)$$

$F(p)$ is the Laplace transform of $f(t)$ with $s = p$. Thus the *contour integral*, equation 10.10, can be used to find the z -transform of any $f(t)$ having an $F(s)$. Table 10.1 shows the Laplace and z -transforms of some typical deterministic time functions defined for $t \geq 0$.

TABLE 10.1
Some common Laplace and z-transforms

Time Function ($t \geq 0$)	Laplace Transform	z-Transform
$\delta(t)$	1	1
$\delta(t - nT)$	e^{-snT}	z^{-n}
$U(t)$	$1/s$	$\frac{z}{z-1}$
$tU(t)$	$1/s^2$	$\frac{Tz}{(z-1)^2}$
$e^{-bt}, t \geq 0$	$\frac{1}{s+b}$	$\frac{z}{z-e^{-bT}}$
$te^{-bt}, t \geq 0$	$\frac{1}{(s+b)^2}$	$\frac{Tze^{-bT}}{(z-e^{-bT})^2}$
$\sin(\omega t), t \geq 0$	$\frac{\omega}{s^2 + \omega^2}$	$\frac{z \sin(\omega T)}{z^2 - 2z \cos(\omega T) + 1}$
$\cos(\omega t), t \geq 0$	$\frac{s}{s^2 + \omega^2}$	$\frac{z^2 - z \cos(\omega T)}{z^2 - 2z \cos(\omega T) + 1}$
$e^{-bt} \sin(\omega t), t \geq 0$	$\frac{\omega}{(s+b)^2 + \omega^2}$	$\frac{e^{-bT} z \sin(\omega T)}{z^2 - 2ze^{-bT} \cos(\omega T) + e^{-2bT}}$
$e^{-bt} \cos(\omega t), t \geq 0$	$\frac{s+b}{(s+b)^2 + \omega^2}$	$\frac{z^2 - ze^{-bT} \cos(\omega T)}{z^2 - 2ze^{-bT} \cos(\omega T) + e^{-2bT}}$

A digital filter is a linear, discrete operator which alters a periodic input number sequence, $\{x(k)\}$, producing a different output number sequence, $\{y(k)\}$. Digital filters are generally described as rational polynomials in the complex variable, z . They may be subdivided into two major categories:

1. *Recursive filters*, in which the present filter output, $y(t)$, depends not only on the present and past *input* samples, $[x(0), x(T), x(2T), \dots, x(kT), \dots]$, but also on filter *output* samples, $[y(0), y(T), y(2T), \dots, y(kT), \dots]$.
2. *Moving average, transversal or FIR* (finite duration impulse response) filters. FIR filters use only the present and past values of $x^*(t)$; their impulse responses are truly finite (i.e. after M sample periods following a '1' input at $t=0$ only, their output is identically zero). The impulse response of a recursive filter generally approaches zero asymptotically as k approaches infinity. In the frequency (z) domain, the z -transformed output of a linear digital filter can be written as the product of the filter's transfer function, $G_Z(z)$, and the z -transform of the filter's input, that is:

$$\mathbf{Y}(z) = \mathbf{X}(z)\mathbf{G}_Z(z) \quad (10.11)$$

In the time domain, the k th output sequence member from a digital filter can be found by the process of real, discrete convolution. This can be written as:

$$y(kT) = y_k = \sum_{i=-\infty}^{\infty} g_i x_{k-i} \quad (10.12)$$

Here, $\{g_i\}$ are the elements of the discrete filter's unit impulse response, and $\{x_{k-i}\}$ are the elements of $\{x_i\}$ reversed in time and shifted by k sample periods. The real convolution process is illustrated using a simple example in Figure 10.1.

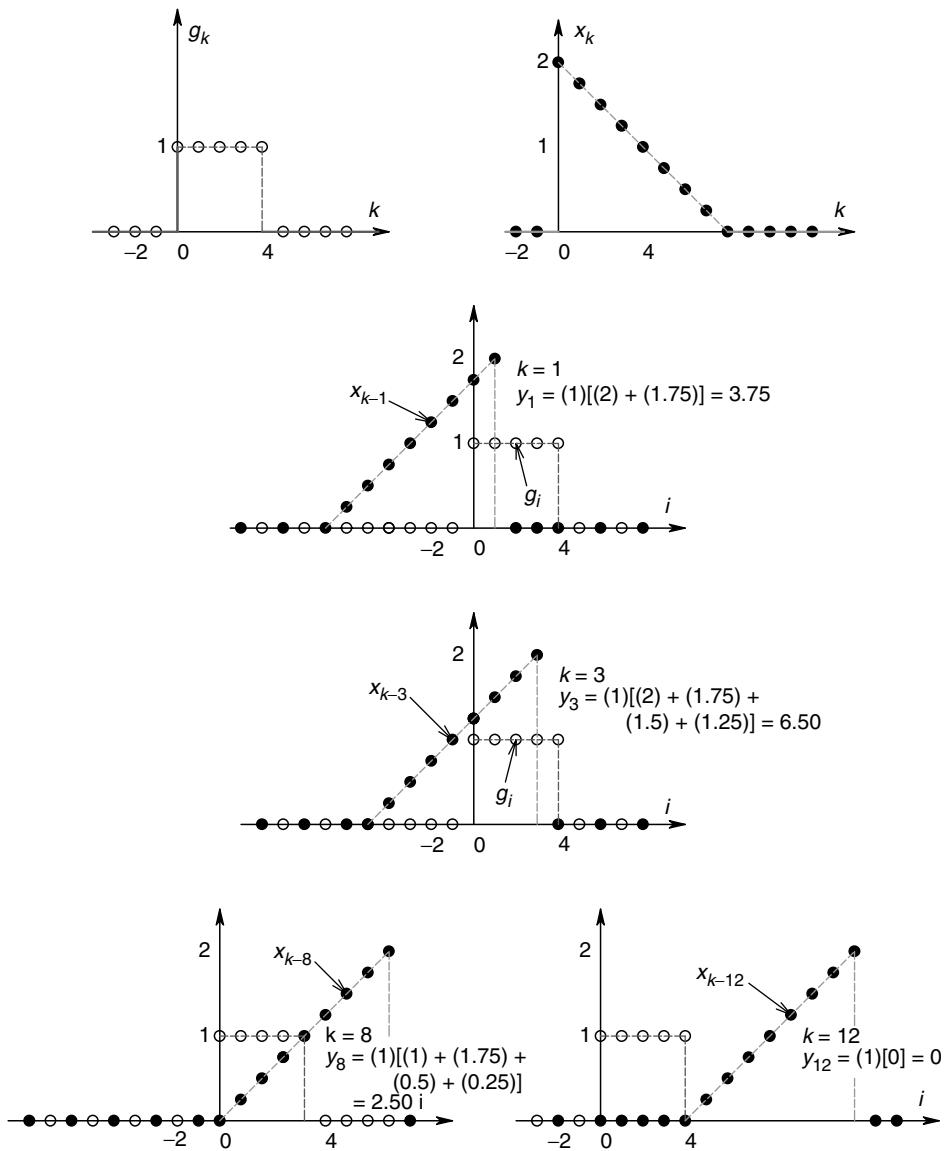
**FIGURE 10.1**

Illustration of the process of discrete, real convolution. A sampled rectangular pulse, g_k , and a sampled finite ramp waveform, x_k , are shown on the top line. In the next four lines, steps in the the real convolution operation given by equation 10.12 are shown. g_k becomes g_i , and x_k becomes x_{k-i} . That is, x_{-i} is x_k reversed in time i and shifted by constant k .

Recursive filters used in linear DSP operations can be expressed as rational polynomials in the complex variable, z . For Example, $G_Z(z)$ can be written in a general, unfactored, polynomial form as:

$$\frac{Y(z)}{X(z)} = G_Z(z) = \frac{a_m z^m + a_{m-1} z^{m-1} + \cdots + a_1 z^1 + a_0 z^0}{b_n z^n + b_{n-1} z^{n-1} + \cdots + b_1 z^1 + b_0 z^0}, \quad n \geq m \quad (10.13)$$

Both the numerator and denominator polynomial may have roots at the origin of the z plane, their roots may occur in complex conjugate pairs, or lie on the negative or positive

real axis in the z plane. The coefficients, $\{a_k\}$ and $\{b_k\}$, are real numbers. The poles (denominator roots) of a stable $G(z)$ must lie inside the unit circle in the z plane. Recursive filters are also known as infinite impulse response (IIR) filters because, although their impulse response magnitudes die out as $k \rightarrow \infty$, the terms go on forever.

To appreciate how a recursive $G_Z(z)$ can be implemented on a computer in real time, we change $G_Z(z)$ to be in powers of $1/z$ or e^{-sT_s} , the delay operator. Simple algebra gives:

$$\frac{Y(z)}{X(z)} = G_Z(z) = \frac{(a_m + a_{m-1}z^{-1} + \cdots + a_1z^{-m+1} + a_0z^{-m})z^{-(n-m)}}{(b_n + b_{n-1}z^{-1} + \cdots + b_1z^{-n+1} + b_0z^{-n})}, \quad n \geq m \quad (10.14)$$

From equation 10.14, it is now easy to write:

$$\begin{aligned} Y(z)(b_n + b_{n-1}z^{-1} + \cdots + b_1z^{-n+1} + b_0z^{-n}) \\ = X(z)(a_mz^{-(n-m)} + a_{m-1}z^{-(n-m+1)} + \cdots + a_1z^{-(n-1)} + a_0z^{-n}) \end{aligned} \quad (10.15)$$

To simplify the notation, let the k th past sample in $x^*(t)$, $x(t - kT_s)$ be written as x_{-k} , and the p th past sample in $y^*(t)$ be written as y_{-p} . Now, we can write equation 10.15 in the time domain:

$$b_n y + b_{n-1}y_{-1} + \cdots + b_1y_{-n+1} + b_0y_{-n} = a_m x_{-n+m} + a_{m-1}x_{-n+m-1} + \dots + a_1x_{-n+1} + a_0x_{-n} \quad (10.16)$$

Equation 10.16 can easily be solved for the present filter output, y , in terms of present and past weighted output and input samples. To illustrate this process, we use a simple third order example:

$$G_Z(z) = \frac{a_2 z^2 + a_1 z^1 + a_0}{b_3 z^3 + b_2 z^2 + b_1 z^1 + b_0}, \quad n - m = 1 \quad (10.17)$$

Following the procedure described above, the present output of $G_Z(z)$ can be written:

$$y = (1/b_3)(a_2 x_{-1} + a_1 x_{-2} + a_0 x_{-3} - b_2 y_{-1} - b_1 y_{-2} - b_0 y_{-3}) \quad (10.18)$$

Implementation of this recursive filter example is shown in Figure 10.2. Note that in this particular example, no *present* value of x is used in computing the present value of y .

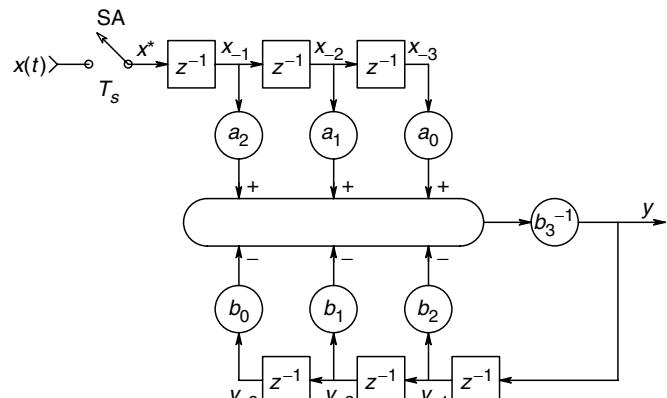


FIGURE 10.2

Block diagram showing the implementation of a third order, recursive filter. z^{-1} represents e^{-sT_s} , the unit delay operator. T_s is the system sampling period.

The impulse response of a digital filter, $\{g_k\}$, is simply its output number sequence, given an input $x=1$ at $t=0$. The $\{g_k\}$ can also be found by long division of the numerator of the filter's $G(z)$ by its denominator. This division generally gives a sequence of terms with coefficients of the form, z^{-k} . In the time domain, this is:

$$\{g_k\} = \sum_{k=0}^{\infty} g_k \delta(t - kT) \quad (10.19)$$

For stable transfer functions, $g_k \rightarrow 0$ as $k \rightarrow \infty$.

An n th order FIR digital filter's transfer function can be written in the general form:

$$\begin{aligned} \frac{Y(z)}{X(z)} &= \frac{a_n z_n + a_{n-1} z^{n-1} + \cdots + a_1 z^1 + a_0}{z^n}, \quad m = n, b_n = 1 \\ &= a_n + a_{n-1} z^{-1} + \cdots + a_1 z^{-n+1} + a_0 z^{-n} \end{aligned} \quad (10.20)$$

By taking the inverse z -transform of equation 10.20, we obtain the difference equation:

$$y = a_n x + a_{n-1} x_{-1} + \cdots + a_1 x_{-n+1} + a_0 x_{-n} \quad (10.21)$$

It is easy to see that if a '1' is input at $t=0$, the FIR filter's impulse response will have $M=(n+1)$ terms, including the present sample.

Before going on to consider some typical applications of digital filters, we remind the reader that in using DSP routines, we generally assume that the analog input signal, $x(t)$, is bandwidth limited by an anti-aliasing low-pass filter, so that it contains no significant spectral energy above the Nyquist frequency (one half the sampling frequency).

10.3 Some Simple DSP Algorithms

The first, simple DSP example we consider is the *FIR, moving average, smoothing or low-pass filter*. This type of filter is often used to 'clean up' or remove high frequency noise accompanying a slowly varying or dc signal. Moving average, FIR filters generally have an odd number of polynomial terms, and an unity dc response. For example, a three term, Hanning low-pass filter is described in the time domain by:

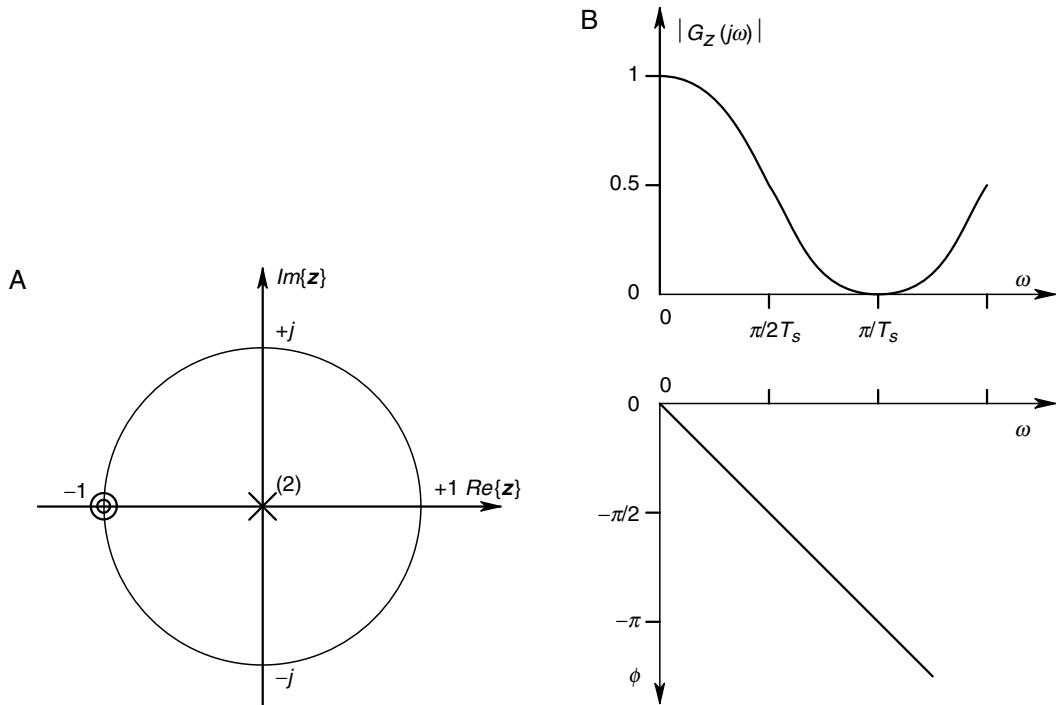
$$y(nT) = (1/4)[x(nT) + 2x(nT - T) + x(nT - 2T)]$$

or

$$y = (1/4)(x + 2x_{-1} + x_{-2}) \quad (10.22)$$

We can write the z -transform for this filter as:

$$G_Z(z) = \frac{Y(z)}{X(z)} = 0.25(1 + 2z^{-1} + z^{-2}) = \frac{0.25z^2 + 0.5z^1 + 0.25}{z^2} \quad (10.23)$$

**FIGURE 10.3**

A. A z plane plot of poles and zeros of the Hanning low-pass filter of equation 10.23. B. The magnitude and phase of $G_z(j\omega T)$ of the Hanning filter.

The numerator of $G_Z(z)$ has two real roots (zeros) at $z = -1$. A z plane plot of the roots of $G_Z(z)$ is shown in Figure 10.3A. To examine the frequency response of this $G_Z(z)$, it is convenient to substitute $z = e^{j\omega T} = e^{j\omega T}$ into equation 10.23, and vary ω

$$G_Z(j\omega T) = 0.25(1 + 2e^{-j\omega T} + e^{-j2\omega T}) = 0.5[1 + \cos(\omega T)]e^{-j\omega T} \quad (10.24)$$

The radian sampling frequency is $\omega_s = 2\pi/T$ and the Nyquist frequency is $\omega_N = \omega_s/2 = \pi/T$. Since we assume $x(t)$ is analog anti-alias filtered before sampling, we need only consider $0 \leq \omega \leq \omega_N = \pi/T$. The magnitude and phase of $G_Z(j\omega T)$ is plotted in Figure 10.3B. Note that the phase is linear in ω and the dc magnitude response of the Hanning filter is unity, decreases to zero at $\omega = \omega_N$ and then increases again.

Many other higher order, FIR low-pass or smoothing filters can be designed. In general, higher the order of the FIR filter, sharper the attenuation can be made. A number of criteria exist for the design of FIR smoothing filters. For example, Tompkins and Webster (1981) derived a five-term, FIR, low-pass filter based on a least squared error criterion to fit the five data samples used in the filter with the best parabola. This gives a time domain filter function of:

$$y = (1/35)(-3x + 12x_{-1} + 17x_{-2} + 12x_{-3} - 3x_{-4}) \quad (10.25)$$

In the z domain, this FIR algorithm becomes:

$$\frac{Y(z)}{X(z)} = G_Z(z) = (1/35)(-3 + 12z^{-1} + 17z^{-2} + 12z^{-3} - 3z^{-4}) \quad (10.26)$$

The frequency response magnitude of this $\mathbf{G}_Z(j\omega T)$ and several other FIR low-pass filters are illustrated in Tompkins and Webster (1981). These authors also derive a five point FIR lowpass filter that satisfies the criteria:

$$\mathbf{G}_Z(j0) = 1 \quad (10.27a)$$

$$\frac{d\mathbf{G}_Z(j0)}{d\omega T} = 0 \quad (10.27b)$$

$$\frac{d^2\mathbf{G}_Z(j0)}{d\omega T^2} = 0 \quad (10.27c)$$

That is, the filter's frequency response is maximally flat at dc. This filter is shown to be implemented by the difference equation:

$$y = (1/16)(-x + 4x_{-1} + 10x_{-2} + 4x_{-3} - x_{-4}) \quad (10.28)$$

Its frequency reponse magnitude is:

$$\mathbf{G}_Z(j\omega T) = 5/8 + 4/8 \cos(\omega T) - 1/8 \cos(2\omega T) \quad (10.29)$$

Clearly, $\mathbf{G}_Z(j0) = 1$ and $\mathbf{G}_Z(j\pi) = 0$ for this filter.

Often, in instrumentation applications, we wish to eliminate a specific, coherent frequency, ω_i , from a low frequency signal. We assume that $\omega_i < \omega_N$. To create a *digital notch filter*, we need to place a pair of complex conjugate zeros of $G(z)$ on the unit circle in the z plane. This can be done quite easily with a quadratic FIR filter of the form:

$$G_Z(z) = 0.5(1 + z^{-2}) \quad (10.30)$$

which is implemented as

$$y = 0.5(x + x_{-2}) \quad (10.31)$$

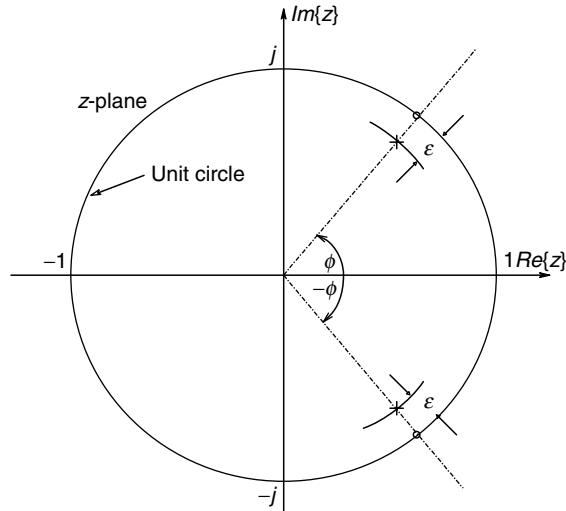
Substituting $e^{j\omega T}$ for z in equation 10.30 gives us the frequency response of this simple notch filter:

$$\mathbf{G}_Z(j\omega T) = \cos(\omega T)e^{-j\omega T} \quad (10.32)$$

It is easy to see that $\mathbf{G}_Z(j0) = \mathbf{G}_Z(j\omega_N T) = 1$ and $\mathbf{G}_Z(j(\omega_N/2)T) = 0$, that is, the notch frequency, ω_i , is half the Nyquist frequency in this example. Attenuation around the notch frequency is certainly not sharp for this filter. Tuning the notch is accomplished by adjusting the sampling period, T .

To obtain a sharp, high Q notch, we can use a recursive notch filter design, adapted from Widrow and Stearns (1985).

$$F_Z(z) = \frac{z^2 - z2 \cos(\phi) + 1}{z^2 - z2\eta \cos(\phi) + \eta^2} \quad (10.33)$$

**FIGURE 10.4**

A z -plane plot of the poles and zeros of the recursive, quadratic notch filter of equation 10.33 given by Widrow and Stearns (1985).

where $\eta = 1 - \varepsilon$, $0 < \varepsilon \ll 1$ and the angle $\phi = (\omega_i/\omega_N)\pi$ is shown on the unit circle in Figure 10.4. ω_i , as in the example above, is the frequency of the interfering sinusoid that we wish to eliminate. Consideration of the numerator of $F(z)$ shows that the filter has complex conjugate zeros on the unit circle at:

$$z_{01} = \cos(\phi) + j \sin(\phi) \quad (10.34a)$$

$$z_{02} = \cos(\phi) - j \sin(\phi) \quad (10.34b)$$

The poles of $F_Z(z)$ are easily shown to be just inside the unit circle, lying at $\pm\phi$ with respect to the real axis. Factoring the denominator yields:

$$z_{P1} = \eta(z_{01}) \quad (10.35a)$$

$$z_{P2} = \eta(z_{02}) \quad (10.35b)$$

Following Widrow and Stearns, we can show that the recursive notch filter's Q is given by:

$$Q = \frac{\text{Center frequency}}{-3 \text{ dB bandwidth}} = \frac{\omega_i T}{2(1 - \eta)} \quad (10.36)$$

Higher the Q , sharper is the notch. The Widrow and Stearns adaptive filter can be tuned to any ω_i by calculating

$$\cos(\phi) \equiv \cos[(\omega_i/\omega_N)\pi] \quad (10.37)$$

and then using this value in the recursion formula for $y(nT)$:

$$y = \eta[x + (y_{-1} - x_{-1})2 \cos(\phi) + x_{-2} - \eta y - 2] \quad (10.38)$$

Often in signal processing applications, we are interested in estimating the points in time when a bandwidth limited signal reaches its maxima or minima (i.e. the points in time

where its first derivative goes to zero). The simplest form of *digital differentiator* is the FIR two point difference equation:

$$y = (1/T)(x - x_{-1}) \quad (10.39)$$

Figure 10.5A illustrates how equation 10.39 estimates the slope of $x(t)$ by a straight line approximation using the present and first past sample of $x(t)$. The digital transfer function for this simple routine is:

$$D_Z(z) = \frac{Y(z)}{X(z)} = \frac{z - 1}{Tz} \quad (10.40)$$

$D_Z(z)$ has a zero at $z = +1$ and a pole at the origin of the z plane. By substituting $e^{j\omega T}$ for z in $D_Z(z)$, we can find the frequency response of the simple differentiator over $0 \leq \omega \leq \omega_N$:

$$D_Z(j\omega T) = (2/T) \sin(\omega T/2) \exp[j(\pi/2 - \omega T/2)] \quad (10.41)$$

For low ω such that $\sin(\omega T/2) \approx \omega T/2$, in radians:

$$D_Z(j\omega T) \cong \omega e^{j\pi/2} \quad (10.42)$$

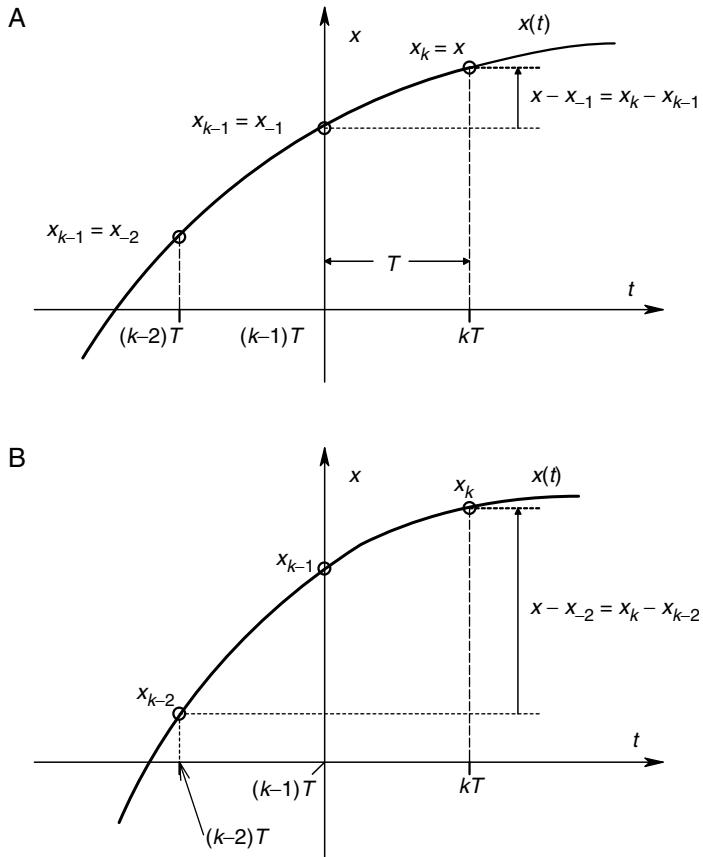


FIGURE 10.5

A. Illustration of the simple, two point, digital differentiator operating on $x(t)$. B. Illustration of the three point, central difference, digital differentiator, given by equation 10.47.

which is the frequency response function of an ideal (analog) differentiator. As ω increases, this $|D_Z(j\omega T)|$ reaches a peak at the Nyquist frequency, and then decreases.

Other digital differentiation routines are sometimes used. All digital differentiators depart from ideality as ω approaches the Nyquist frequency. In some cases, it is desireable to have the differentiator frequency response reach a peak and roll off at frequencies higher than $\omega_N/2$ to minimize the effects of noise on $x^*(t)$. These routines include [Tompkins and Webster, 1981]:

$$y = (1/2T)(x - x_{-2}) \quad (3 \text{ point central difference}) \quad (10.43)$$

$$y = 0.1(2x + x_{-1} - x_{-3} - 2x_{-4}) \quad (5 \text{ point, least squares, parabolic fit}) \quad (10.44)$$

$$y = (1/28)(3x + 2x_{-1} + x_{-2} - x_{-4} - 2x_{-5} - 3x_{-6}) \quad (7 \text{ point, LSPF}) \quad (10.45)$$

$$y = (1/60)(4x + 3x_{-1} + 2x_{-2} + x_{-3} - x_{-5} - 2x_{-6} - 3x_{-7} - 4x_{-8}) \quad (9 \text{ point}) \quad (10.46)$$

How the *three point central difference differentiator* estimates the slope of $x(t)$ is illustrated in Figure 10.5B. The transfer function of this differentiator is easily seen to be:

$$D_Z(z) = \frac{z^2 - 1}{2Tz^2} \quad (10.47)$$

This transfer function has zeros at $z = 1, z = -1$ and two poles at the origin of the z plane. The frequency response of $D_Z(z)$ is determined by substituting $e^{j\omega T}$ for z and using the Euler relations:

$$D_Z(j\omega T) = (1/T) \sin(\omega T) \exp[j(\pi/2 - \omega T)] \quad (10.48)$$

Like the two point differentiator, this algorithm has nearly ideal behaviour at low frequencies. Its magnitude response peaks at $\omega = \omega_N/2$ and goes to zero at $\omega = \omega_N$. Note that the gains of the two and three point differentiators depend on a knowledge of the sampling period, but the five and higher point, LS, parabolic fit differentiator gains do not depend on T .

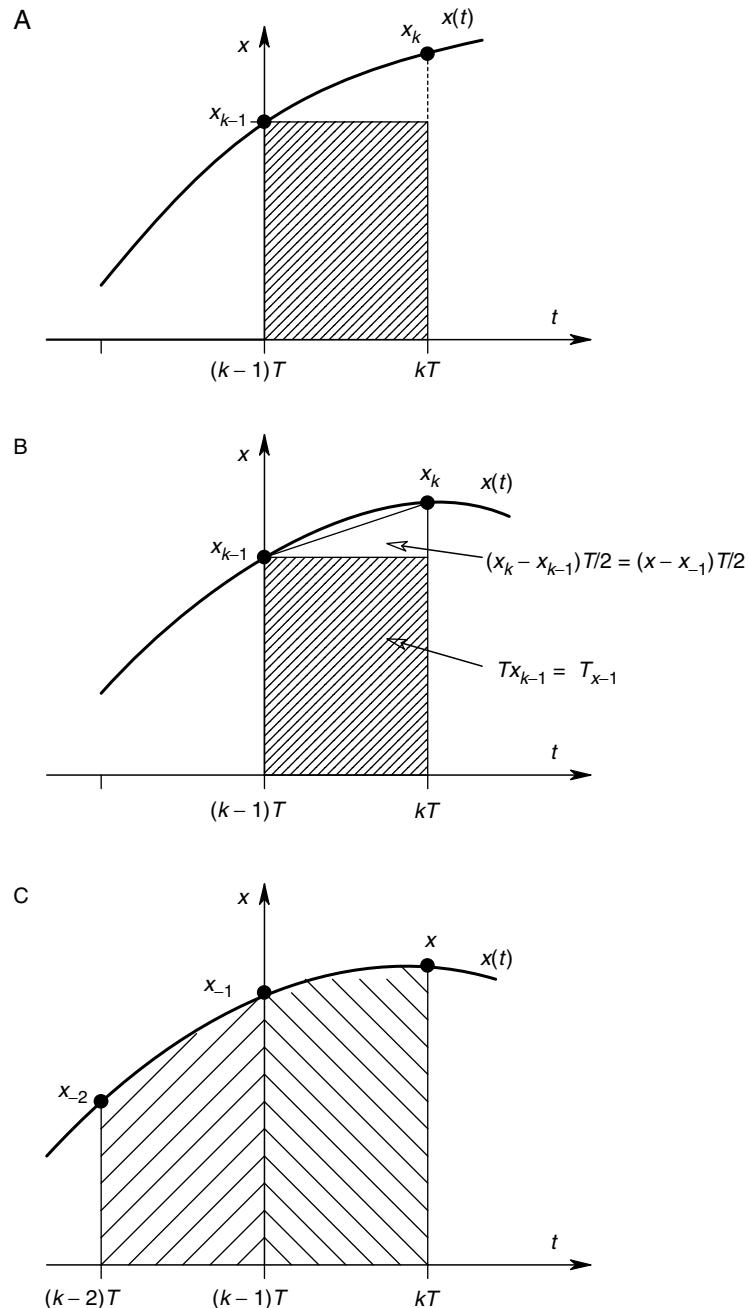
Digital integration routines allow us to estimate the area under a curve, using sampled data. In chemometric instrumentation systems, the outputs of various types of chromatographs and spectrographs are often integrated in order to estimate the quantity of a chemical species in a sample. Another obvious application of integrating a sampled variable is to compute its mean value, or RMS value.

The simplest form of digital integrator is the *rectangular integrator*, the action of which is illustrated in Figure 10.6A. All digital integration routines are recursive (i.e. they add a present estimate of area to the cumulative old estimate of area). That is, the present integrator output, $i(nT) = i$ is given by:

$$i = i_{-1} + \Delta i \quad (10.49)$$

In the case of the rectangular integrator:

$$i = i_{-1} + Tx_{-1} \quad (10.50)$$

**FIGURE 10.6**

A. Illustration of three common, digital integration routines in the time domain. Simple rectangular integration is shown here. Its transfer function is given by equation 10.51. B. Trapezoidal integration, See equation 10.57. C. Simpson's rule integration. See equation 10.58.

In terms of a z -transform transfer function, equation 10.50 can be written as:

$$\frac{I(z)}{X(z)} = H_{ZR}(z) = \frac{T}{z-1} \quad (10.51)$$

Substituting $e^{j\omega T}$ for z gives us the frequency response of the rectangular integrator:

$$H_{ZR}(j\omega T) = \frac{T/2}{\sin(\omega T/2)} \exp[-j(\pi/2 + \omega T/2)] \quad (10.52)$$

At low frequencies, $H_{ZR}(j\omega T)$ behaves like,

$$H_{ZR}(j\omega T) \cong (1/\omega) \exp[-j\pi/2] \quad (10.53)$$

which is the behavior of an (ideal) analog integrator. At the Nyquist frequency,

$$H_{ZR}(j\omega_N T) = (T/2) \exp[-j\pi] \quad (10.54)$$

Many other integration routines exist, some of them quite complex. Here, we will review two other simple integration algorithms commonly used in DSP:

$$i = i_{-1} + [Tx_{-1} + (T/2)(x - x_{-1})] \quad (\text{Trapezoidal integration, Figure 10.6B}) \quad (10.55)$$

$$i = I_{-2} + (T/3)[x + 4x_{-1} + x_{-2}] \quad (\text{Simpson's 1/3 rule, Figure 10.6C}) \quad (10.56)$$

From these difference equations, we can easily derive the *transfer function for trapezoidal integration* in z as:

$$H_{ZT}(z) = \frac{T(z+1)}{2(z-1)} \quad (10.57)$$

For Simpson's rule:

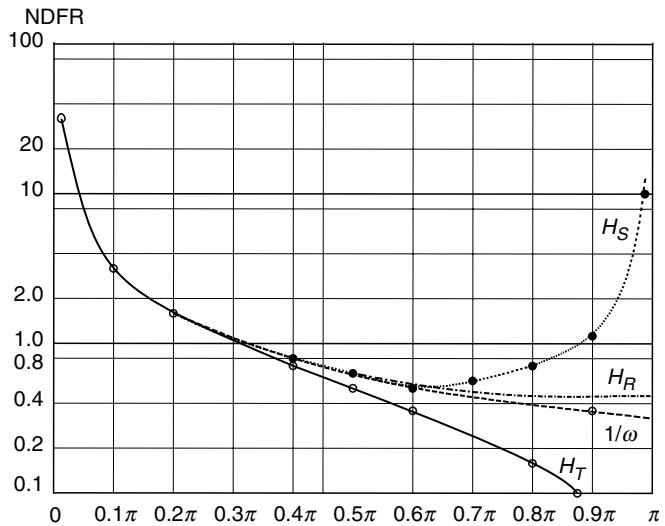
$$H_{ZS}(z) = \frac{T(z^2 + 4z + 1)}{3(z^2 + 1)} \quad (10.58)$$

The frequency responses of these integrators are easily shown to be:

$$H_{ZT}(j\omega T) = (T/2) \cot(\omega T/2) \exp[-j\pi/2] \quad (10.59)$$

$$H_{ZS}(j\omega T) = \frac{T[2 + \cos(\omega T)]}{3 \sin(\omega T)} \exp[-j\pi/2] \quad (10.60)$$

Note that the phase for both the trapezoidal and Simpson's rule integrators is that of an ideal integrator. However, their frequency response magnitudes depart from an ideal integrator at higher frequencies. Simpson's rule follows $1/\omega$ well out to $\omega = \omega_N/2$ and then reverses slope and increases sharply as ω increases above $2\omega_N/3$. Thus, Simpson's rule behaves poorly on any $x(t)$ containing noise power above two-thirds of the Nyquist frequency. The gain magnitude of the trapezoidal integrator follows the ideal frequency

**FIGURE 10.7**

Non-dimensional, frequency response magnitudes of the three integrators shown in Figure 10.6. The frequency response magnitude of an ideal (analog) integrator, $1/\omega T$, is shown for comparison. Note: H_R = rectangular integrator, H_T = trapezoidal integrator, H_S = Simpson's rule integrator.

response curve closely, up to about $\omega_N/3$, and then drops off to zero at ω_N . Thus, the trapezoidal integrator gives better accuracy when integrating noisy signals. If the power density spectrum of $x(t)$ has little power above half the Nyquist frequency, the Simpson's integration rule may perform more accurately. The frequency response magnitude curves for the three integrators we have discussed is shown in Figure 10.7.

10.4 Discrete and Fast Fourier Transforms and their Applications

Just as we can represent continuous time domain signals in the frequency domain using the continuous Fourier transform (CFT) pair, shown below,

$$\mathbf{F}(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \quad (10.61)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{F}(j\omega)e^{j\omega t} d\omega \quad (10.62)$$

we can also characterize discrete time signals of finite length in the frequency domain using the Discrete Fourier Transform (DFT) and the Inverse Discrete Fourier Transform (IDFT). In the CFT, $f(t)$ can exist over all time. In the DFT, we assume that a finite number of samples, N of $x(t)$ are taken, each at an interval of T_s seconds. The total length of time window or epoch over which $x(t)$ is sampled is $T_E = (N - 1)T_s$. In computing the DFT, we assume $x(t) = x(t + NT_s)$ (i.e. $x(t)$ is periodic with a period of NT_s seconds).

The N elements of $\{x_n\}_N$ give rise to N elements of the DFT according to the relations:

$$X_k = \sum_{m=0}^{N-1} x_m \exp\left(-j\frac{2\pi}{N}km\right), \quad k = 0, 1, \dots, N - 1 \quad (10.63a)$$

or

$$X_k = \sum_{m=0}^{N-1} x_m \cos(2\pi km/N) - j \sum_{m=0}^{N-1} x_m \sin(2\pi km/N) \quad (10.63b)$$

The N terms of the DFT are complex numbers; they have real and imaginary parts or equivalently, magnitudes and angles. The frequency spacing between adjacent terms of the spectrum, $\{X_k\}_N$, is given by equation 10.64:

$$\Delta\omega = \frac{2\pi}{T_E} = \frac{2\pi}{(N-1)T_s} \text{ r/s} \quad (10.64)$$

The total frequency span of $\{X_k\}_N$ is 0 to $2\pi/T_s$ r/s.

The DFT has several interesting properties. It is a linear operator (i.e. it obeys superposition), *sic*:

$$DFT_N\{Ap_k + Bq_k\} = A DFT_N\{p_k\} + B DFT_N\{q_k\} \quad (10.65)$$

The DFT is periodic with period N . That is:

$$X_k = X_{k+N} \quad (10.66)$$

A very important property of the DFT is that the DFT of real signals is *conjugate symmetric* [Williams, 1986]. This can be written as:

$$X_k = X_{N-1-k}^*, \quad 0 \leq k \leq N/2 \quad (10.67)$$

Another way of describing this property is to state that the real part of X_k is *even* about $k=N/2$, and the imaginary part of X_k is *odd* around $k=N/2$. Thus the magnitude of X_k is *even* around $k=N/2$, and is equal to the magnitude of X_{N-1-k}^* . This means that one half of a DFT spectrum of $\{x_n\}_N$ is redundant. Thus, N samples of $x(t)$ yield $N/2$ useful spectrum values ranging from 0 to $f_N=1/2T_s$ Hz.

The inverse DFT can be written:

$$x_k = \frac{1}{N} \sum_{m=0}^{N-1} X_m \exp\left(+j\frac{2\pi}{N}mk\right) \quad (10.68)$$

The IDFT, also, is linear and periodic in N .

As an example of setting up a DFT routine, let us assume we need to resolve and examine in the frequency domain, two coherent (sinusoidal) signals to which broadband noise has been added. The signals are at 400 and 420 Hz. We wish to examine the magnitudes of $\{X_k\}_N$ to determine the relative levels of the signals and the noise. First, let us set the Nyquist frequency of the system to be 1000 Hz. The anti-aliasing filters must have attenuated the signals plus noise to a negligible output power level at and above 1 kHz. The sampling rate required is thus 2000 samples/s. The signal spacing is 20 Hz. This implies that we should be able to resolve a $\Delta f = 2$ Hz. Thus, $T_E = 1/\Delta f = 0.5$ s and $N = T_E/T_s + 1 = 1001$ samples. (Typically, 1024 samples could be used in calculating $\{X_k\}_N$.)

10.4.1 Use of Data Windows to Improve Spectral Resolution

Direct computation of $\{X_k\}_N$ using equation 10.63 is always done with a *windowing function* operating on the sampled data, $\{x_k\}_N$. That is, the product,

$$k'_x = w_k x_k \quad (10.69)$$

is computed in the time domain, where w_k is the windowing function. The purpose of a windowing function is to reduce a phenomenon known as *spectral leakage*. If a pure, sinusoidal $x(t)$ is sampled and DFTed, we generally observe symmetrical, non-zero sidelobes in $\{X_k\}_N$ around the main peak at the sinusoid's frequency. This phenomenon is called spectral leakage because spectral energy at the sinusoid's frequency effectively 'leaks' into the adjacent frequency terms. What windows do in general is reduce the magnitude of the leakage sidelobes at the expense of the sharpness of the main spectral peak at the input sinusoid's frequency.

In general, in the frequency domain,

$${}^W X(j\omega) = W(j\omega) \otimes X(j\omega) \quad (10.70)$$

That is, the FT of the windowed, sampled data sequence can be expressed as the *complex convolution* of the FT of the windowing function with the FT of $x(t)$.

To illustrate how windowing works, we first consider the case of the simplest, rectangular window consisting of N unit impulses, spaced T_s seconds apart. The rectangular window function is a truncated unit impulse train which multiplies (modulates) $x(t)$ in the time domain to generate the sampler output. The rectangular window function can be written as:

$$w(t) = \{w_k\}_N = \sum_{k=0}^{N-1} \delta(t - kT_s) \quad (10.71)$$

A fundamental property of Fourier transforms is that the CFT of an even function in t is even in ω and has zero phase. It is therefore algebraically convenient to consider both the window and $x(t)$ to be even functions. This process is shown in Figure 10.8. We choose N to be an odd number, so M will be an integer. Note that the epoch length, T_E is:

$$T_E = (N - 1)T_s = 2MT_s \quad (10.72)$$

Hence,

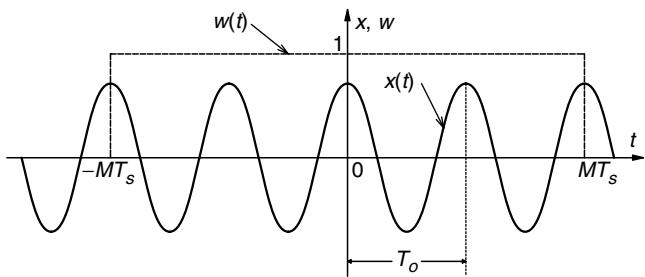
$$M = (N - 1)/2 \quad (10.73)$$

The CFT of the rectangular window can be shown to be [Papoulis, 1968]:

$$W(\omega) = \frac{\sin[\omega(2M + 1)T_s/2]}{\sin(\omega T_s/2)} = \frac{\sin(\omega NT_s/2)}{\sin(\omega T_s/2)} \quad (10.74)$$

For purposes of this example, let us assume that $x(t)$ is a cosine wave (even function): $x(t) = A \cos(\omega_0 t)$, where $\omega_0 < \omega_N$. The CFT of the cosine $x(t)$ is well-known:

$$X(\omega) = A[\pi\delta(\omega + \omega_0) + \pi\delta(\omega - \omega_0)] \quad (10.75)$$

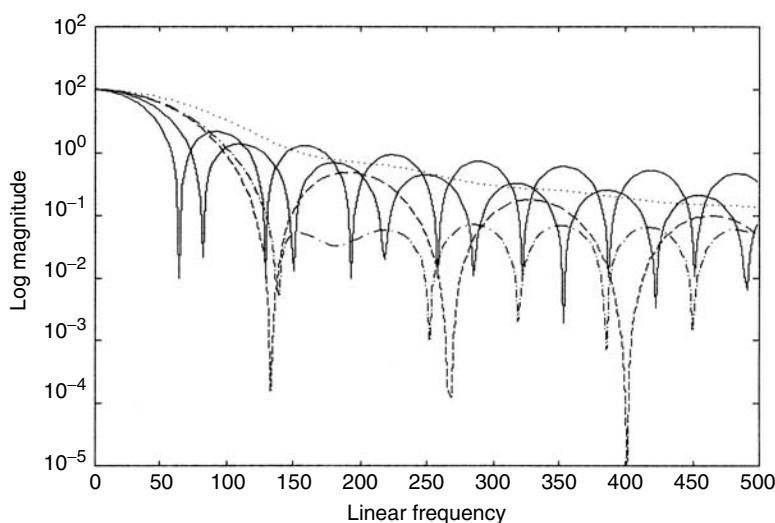
**FIGURE 10.8**

A unit rectangular (even) window function and a cosine wave. A finite number of samples of the cosine wave effectively multiplies it by the sampled, unit rectangular window function.

Complex convolution now becomes real convolution, where the kernel only exists at $\omega = \pm\omega_0$. Thus, the CFT of the finite, sampled, rectangular windowed cosine wave is:

$${}^W X(\omega) = \frac{A}{2} \left\{ \frac{\sin[(\omega + \omega_0)N T_s/2]}{\sin[(\omega + \omega_0)T_s/2]} + \frac{\sin[(\omega - \omega_0)N T_s/2]}{\sin[(\omega - \omega_0)T_s/2]} \right\} \quad (10.76)$$

The main spectral peaks in ${}^W X(\omega)$ are seen to occur at $\omega = \pm\omega_0$. The leakage is due to the side lobes of equation 10.74. Figure 10.9 illustrates normalized plots of the DFTs of several window functions (see caption for details). Clearly, to reduce leakage and to get the best resolution, we must find a window function whose CFT has a narrow main lobe and low sidelobes compared to the main lobe peak. Many such window functions have been devised and used. These include, but are not limited to [Papoulis, 1977], Bartlett, Tukey, Hamming, parabolic, maximum energy concentration, minimum energy moment, minimum amplitude moment, von Hann (Hanning), and Kaiser. (The latter two windows are described in Williams (1986).) The Hamming window is one of the most widely used windows in DFT and FFT computations. Its FT has low major sidelobes. It consists of a cosine on a pedestal. For $N \geq 64$ samples, it can be shown that the pedestal

**FIGURE 10.9**

Normalized plots of the DFT magnitudes of several commonly used windowing functions. The solid line with narrowest main lobe and highest side lobes = rectangular window. Note: Solid line with wider main lobe and lower side lobes = Kaiser window. Dashes = triangular (Bartlett) window. Dash-dot = Hamming window. Dots = triangle squared (Parzen). A 1024 point FFT was computed with Matlab™. Log magnitude vs linear frequency scale.

height approaches 0.536 and the cosine term is multiplied by 0.464. Each sample of x_k is multiplied by a corresponding w_k :

$$w_k = 0.536 + 0.464 \cos\{\pi[2k - N + 1]/N\}, \quad k = 0, 1, \dots, (N - 1) \quad (10.77)$$

In the symmetrical, even function case, we center the window function at $m=0$ and we can write:

$$w_m = 0.536 + 0.464 \cos[\pi 2m/(N - 1)], \quad -N/2 \leq m \leq (N/2 - 1) \quad (10.78)$$

Figure 10.10 illustrates the major lobes of several common window functions. The DFTs have been normalized. Note that the base width, B , of the major spectral lobe of the Bartlett (triangular) and Hamming windows is approximately given by:

$$B \approx \frac{8\pi}{(N - 1)T_S} \text{ r/s} \quad (10.79)$$

The B of the Parzen window is somewhat wider. The choice of the window algorithm to be used involves a trade-off between main lobe width and the amount of spectral leakage seen at a given distance from the frequency of the main peak. Generally, narrower the main lobe, greater is the leakage. The selection of a window algorithm can be somewhat subjective, and the window selected can influence one's ability to interpret the DFT spectrograms. Most commercial DFT spectrum analyzers offer the user a choice of plain, rectangular window, cosine window such as the Tukey, Hanning or Hamming, or the Kaiser window.

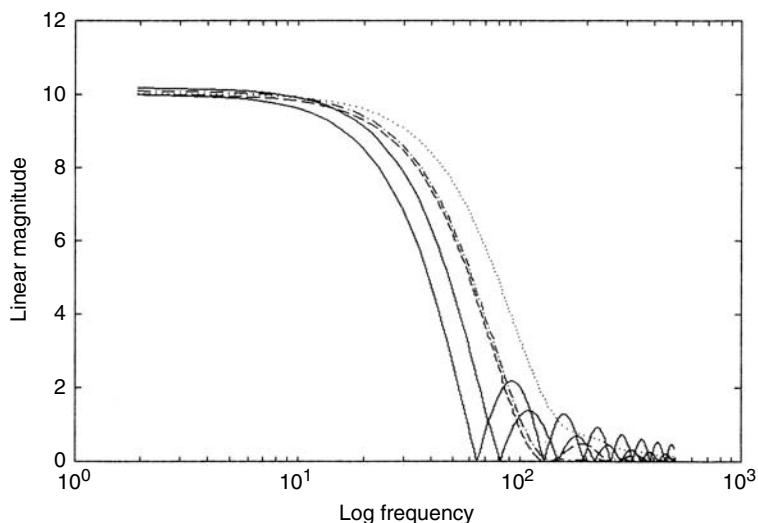


FIGURE 10.10

Normalized plots of the DFT magnitudes of the windows of Figure 10.9, except that a linear magnitude *vs* log frequency scale is used. Note the trade-off between main lobe narrowness and side lobe height. The same note used in Figure 10.9 is used.

10.4.2 Use of the DFT to Characterize Random Signals and Noise

A major use of the modern spectrum analyzer is to allow us to visualize the distribution of power with frequency in an analog waveform. This feature can be useful in the vibration analysis of mechanical structures, measurement of noise, determination of linear system transfer functions by the white noise method, etc.

In chapter 3, we saw that the power density spectrum (PDS) is used to describe the power content at a particular frequency of an infinite length, analog waveform. The units of the PDS are mean squared volts (units) per Hz. In modern engineering practice, we use the DFT with an appropriate window to estimate the square root of the PDS (RPDS) of an analog waveform. In this case, the RPDS' units are RMS volts (units)/ $\sqrt{\text{Hz}}$.

The discrete RPDS can be calculated in the following manner: First, the analog waveform, $x(t)$, is passed through an analog anti-aliasing low-pass filter to remove any significant spectral power at and above the Nyquist frequency, $\omega_N = \omega_s/2 = \pi/T_s$ r/s. Next, a finite length of the anti-aliasing filter's analog output, $x(t)$, is sampled at a rate of $1/T_s$ samples/s. N values of $x(nT_s)$ are stored in an array, $\{x_k\}_N$, in the computer. Each x_k is multiplied by the corresponding window weighting constant, w_k , giving a windowed data array, $\{\tilde{w}_k\}_N$. N elements of the DFT of $\{\tilde{w}_k\}$ are now calculated using symmetrical coefficients, giving $\{\tilde{X}_m\}_N$. Due to the conjugate symmetry property of DFT output, only $N/2$ points in $\{\tilde{X}_m\}_N$ are unique. We store a buffer array consisting of $\{\tilde{X}_m\}_{N/2}$ for $m=0$ to $m=(N/2-1)$. $m=0$ corresponds to $f=0$, and $k=(N/2-1)$ corresponds to $f=f_{MAX}=f_s/2=f_N=1/2T_s$ Hz in the spectrum.

The frequency spacing between Fourier coefficients is $\Delta f = 1/(N-1)T_s$ Hz. The coefficients of a hypothetical $N=8$ point DFT are shown in Figure 10.11. We note that the DFT coefficient magnitudes can be found from the Pythagorean theorem. Also, we observe that the important parameter, T_s , does not appear in the DFT operation, equation 10.63. T_s is used to calculate the spectrum scale factors, Δf and f_{MAX} .

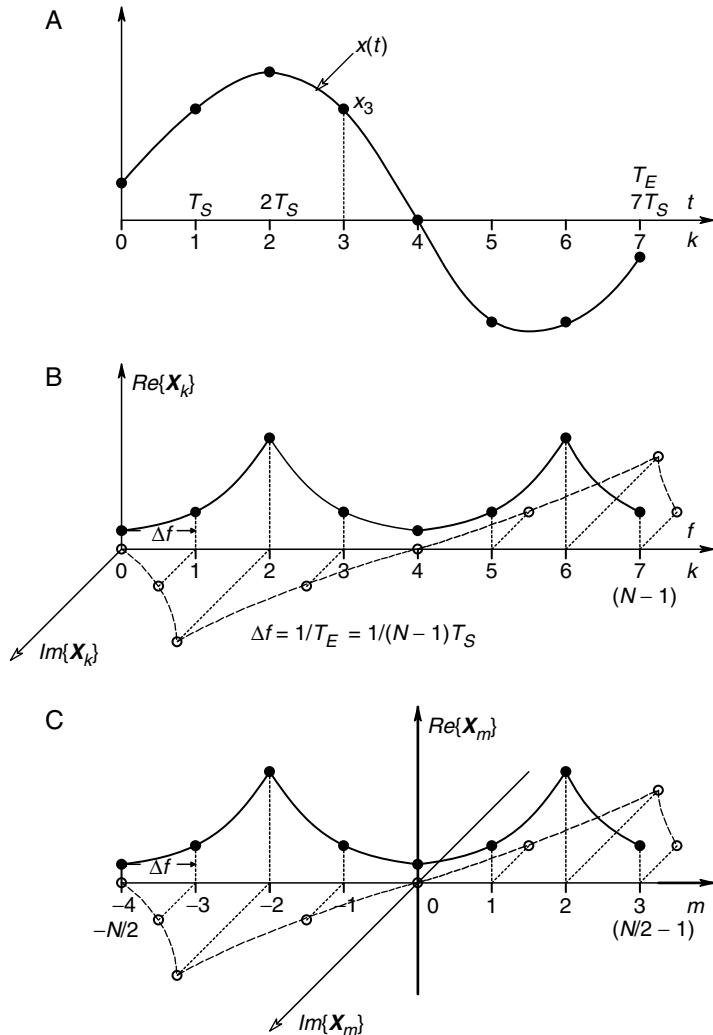
If $x(t)$ is a broadband (noisy) waveform, there will be a great deal of natural, variation between corresponding spectrum coefficient's magnitudes in the computation of M 'similar' spectra. Averaging the corresponding elements of M spectra permits a reduction of the standard deviation of a given element by a factor of about $1/\sqrt{M}$.

In the Hewlett-Packard Model 3582A DFT spectrum analyzer, noisy spectra are averaged in the following manner. M spectra are computed sequentially. Two arrays of length N are set up. One holds the average of the squares of the corresponding real parts of the DFT coefficients and the other, the average of the squares of the corresponding imaginary parts of the DFT coefficients. After M spectra are calculated and the mean squares of the corresponding coefficients are stored, the corresponding mean squares are added and the square root is taken of each corresponding sum, giving a true RMS RPDS. These steps can be shown mathematically as follows:

$$\overline{[\text{Re}\{X_k\}]^2} = \frac{1}{M} \sum_{i=1}^M [\text{Re}\{^i X_k\}]^2 \quad (10.80)$$

$$\overline{[\text{Im}\{X_k\}]^2} = \frac{1}{M} \sum_{i=1}^M [\text{Im}\{^i X_k\}]^2 \quad (10.81)$$

$$\text{RMS } X_k = \sqrt{[\text{Re}\{X_k\}]^2 + [\text{Im}\{X_k\}]^2}, \quad N/2 \leq k \leq (N-1) \quad (\text{integer } k) \quad (10.82)$$

**FIGURE 10.11**

A. Eight samples of a continuous time function, $x(t)$, are taken. A rectangular window is used. B. The hypothetical DFT of the eight point x_k is shown. Real and imaginary components of the DFT are shown for $0 \leq k \leq 7$ (eight samples). Note that the Pythagorean magnitude of the DFT is repeated for $0 \leq k \leq (N/2 - 1)$ and for $N/2 \leq k \leq (N - 1)$. This is a result of the property of conjugate symmetry. C. By translating DFT coordinates, we see that we only need half of the N -point DFT to characterize x^* in the frequency domain.

The HP 3582A spectrum analyzer also offers the option of an exponential, moving averager to reduce the uncertainty in calculating the spectra of non-stationary, noisy signals. *Non-stationarity* means the underlying physical processes producing the noise in $x(t)$ are changing with time (e.g. a resistor may be heating up and its increasing temperature produces an increasing resistance and thermal noise). The exponential averaging process weighs the most recent samples more heavily than old samples. We illustrate it below for the real parts of the k th spectral element:

$$\overline{\text{Exp}[\text{Re}\{\mathbf{X}_k\}]^2} = (1/4)\{[\text{Re}\{^1\mathbf{X}_k\}]^2 + (3/4)^1[\text{Re}\{^2\mathbf{X}_k\}]^2 + (3/4)^2[\text{Re}\{^3\mathbf{X}_k\}]^2 + \dots + (3/4)^{M-2}[\text{Re}\{^{M-1}\mathbf{X}_k\}]^2 + (3/4)^{M-1}[\text{Re}\{^M\mathbf{X}_k\}]^2\} \quad (10.83)$$

Here, the most recent DFT calculation is given the index '1' and the first, or oldest, the index M . Note that $(3/4)^8 = 0.100$ and $(3/4)^{16} = 0.0100$, so the 'tail' of the moving average has negligible weight after computing 16 consecutive DFTs from $x(t)$. As in the case of calculating the RMS X_k , the exponentially averaged RMS X_k is found by the Pythagorean theorem:

$$\text{Exp(RMS } X_k) = \sqrt{\text{Exp[Re}\{X_k\}]^2 + \text{Exp[Im}\{X_k\}]^2} \quad (10.84)$$

10.4.3 The Fast Fourier Transform

In order to obtain a good approximation to the CFT of a bandwidth limited waveform, we must calculate the DFT of the signal with a sampling rate that is greater than twice the highest frequency in the PDS of $x(t)$, and use enough samples to insure a close spacing, Δf , between adjacent values of X_k . It is common to calculate spectra using N s ranging from $256 = 2^8$ to $4096 = 2^{12}$ and higher. Use of the DFT algorithm, equation 10.63a, as written, is seen to require 2^N calculations (multiplications and additions) to calculate each X_k coefficient's real and imaginary parts. However, there are N coefficients, so a total of $2N^2$ multiplications and additions are required to obtain the DFT, $\{X_k\}_N$. If the magnitudes of $\{X_k\}_N$ are to be displayed, then we must apply the Pythagorean theorem N times as well. It is clear that a large DFT calculated by the direct method will involve many high precision computer operations, and take a long time. For example, computation of an $N = 4096$ point DFT involves over 33.5 million multiplications and additions. This is time consuming.

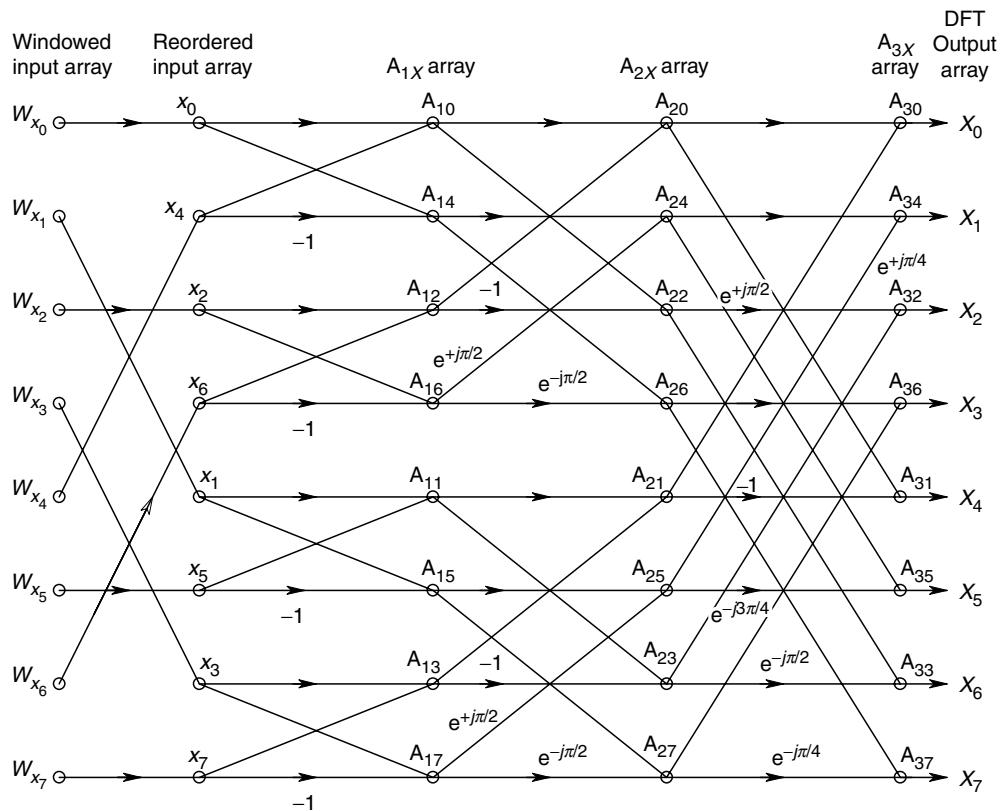
To overcome the computational burden of direct computation of large DFTs, Cooley and Tukey (1965) devised a fast Fourier transform (FFT) means of calculating the DFT. Actually, there are now many variations on the original Cooley-Tukey FFT. It is not our purpose here to describe in detail the steps involved in calculating an FFT/DFT. Rather, we will outline the strategy used and explain why it is efficient.

Using the complex form of the DFT, direct calculation of the k th complex term can be written out as:

$$X_k = x_0 e^{-j(2\pi k 0/N)} + x_1 e^{-j(2\pi k 1/N)} + \cdots + x_i e^{-j(2\pi k i/N)} + \cdots + X_{N-1} e^{-j(2\pi k [N-1]/N)} \quad (10.85)$$

Each of the N exponential terms in equation 10.85 defines an angle for the corresponding real x_k . Thus N vector summations are required in order to find X_k in the conventional DFT, and a total of N^2 vector summations must be done to find the entire DFT, $\{X_k\}_N$.

In computing an FFT, use is made of the fact that the exponential angle operator is periodic in k_i/N . This periodicity leads to redundancy in the DFT calculations which is taken advantage of in an FFT algorithm by rearranging the order of the calculations so that there are $\log_2 N$ columns of vector summers, each of which contains N summers, for a total of $N \log_2(N)$ complex summers in an FFT array. Each summer sums two vectors. In FFT operations, N is generally made a power of 2 (e.g. $N = 2^b$, where b is a positive integer, such as $b = 10$ for $N = 1024$). The angles of the vectors must still be calculated when the FFT array is set-up, and stored as constants. Their values depend only on k , i and N . It is not practical to try to illustrate the signal flow graph of an FFT operation for N s over 8. Figure 10.12 shows an implementation of an eight point FFT. Note the symmetrical rearrangement of the x_k s at the FFT input. In some versions of the FFT, the output coefficients are rearranged.

**FIGURE 10.12**

Flow graph showing an implementation of the Cooley-Tukey FFT algorithm for an eight sample time function. Note that the FFT output values in the A_{3x} array are complex (i.e. they have real and imaginary parts).

The efficiency, η , of the FFT *vs* the DFT calculated in the direct method can be calculated by assuming that computation speed is inversely proportional to the number of vector additions required in each case. So we can write:

$$\eta = \frac{S_{FFT}}{S_{DFT}} = \frac{N}{\log_2(N)} \quad (10.86)$$

For a 4096 point FFT, $\eta = 341.3$, for $N = 1024$, $\eta = 102.4$, etc.

There are many commercial FFT software routines available in various computer languages and programs, including FORTRAN, C, Matlab, etc. Also, companies offering data acquisition interface cards and DSP cards for PCs generally offer their software versions of FFT routines. Thus, there is little probability that a person setting up a DSP capability in a measurement system will have to write their own FFT routine from scratch.

10.4.4 DSP Coprocessor Boards

In the 1990s, a number of commercially available, plug-in, DSP coprocessor boards were available to compute FFTs in PCs and workstations. Designs for these boards constantly evolved, improving their speeds and capabilities. Some of the commercially available, dedicated FFT cards are quite fast. For example, the Array Microsystems

model a66540 Frequency Domain Array Processor (FDAP) card for VME bus computers can do a 1024 point real FFT in 13.3 μs . Their a66550 card for PC/AT computers will do a 1024 point FFT in 125.9 μs . The Ixthos IXD7232 VME bus DSP card can run at a sustained 50 MFLOPS and do a 1024 point, complex FFT (radix 4) in 770 μs . National Instruments' AFFA system uses two plug-in cards to make a powerful, audio frequency FFT system for the MacIntosh II PC with NuBus. The AFFA system does real time audio spectrum analysis up to 20 kHz. It can do 64, 126, 256, 512, or 1024 point FFTs at sampling rates of 22.05, 24, 32, 44.1 or 48 ksamples/s. It has a rather elegant, menu controlled operating software, which includes choice of rectangular, Hanning, Hamming, ExactBlackman, Blackman and Blackman-Harris data windows. Many other companies, too numerous to list here, make DSP coprocessor, plug-in boards for PCs, and they offer operating software that enables flexible calculation of FFTs.

The current trend is not to rely on expensive coprocessor boards for PCs; high end Pentium systems with >2GHz clocks have largely made them obsolete—they let the CPU do the FFT. For example, Datel® has phased out its line of board level data acquisition and processing products as of 28 February 2004, including its powerful PCI-431 series of analog I/O, DSP coprocessor boards for the PCI bus. However, there is still a need for DSP coprocessors and DSP accelerator boards for dedicated (non-PC), OEM DSP systems. RTD Embedded Technologies® offers a line of DSP interface, coprocessor and accelerator boards using Texas Instruments DSP chipsets.

10.5 Digital Routines for Interpolating Discrete Data

Often, in making and interpreting measurements, we need to reconstruct a smooth analog curve between discrete data points. There are a variety of means of generating a smooth, continuous curve between a set of N discrete data points, $\{x_k, y_k\}_N$. Note that for a time sampled $y(t)$, $x_k = kT_s$, $k = 0, 1, 2, \dots, N - 1$. One method of generating a smooth curve is to compute a piecewise linear approximation between the data points. In this approach, we generate a number of polynomial segments of relatively low order (0,1,2,3) between the points. The data point pairs at the ends of each polynomial segment are called *knots*. Polynomials of order 2 or 3 are generally called *splines*. Polynomial approximations of order 0 and 1 are shown in Figure 10.13. We will discuss *cubic splines* here because they are generally the most useful in terms of a trade-off between computational complexity and accuracy. The reader seeking more information on the field of splines should consult texts on numerical analysis such those by Greenspan and Casulli (1993) and Scheid (1990).

A *cubic spline* is a cubic polynomial which is made to fit three consecutive data points, (x_k, y_k) , (x_{k+1}, y_{k+1}) and (x_{k+2}, y_{k+2}) , or if y is the sampled data, then (kT_s, y_k) , $([k+1]T_s, y_{k+1})$ and $([k+2]T_s, y_{k+2})$. The polynomial may be written as:

$$y = a + bx + cx^2 + dx^3 \quad (10.87)$$

and

$$\dot{y} = b + 2cx + 3dx^2 \quad (10.88)$$

We assume that $x_0 = t = 0$ (local time origin) and that the intervals, $x_i - x_{i-1} = T_s$ over the entire dataset. Thus $x_0 = 0$, $x_1 = T_s$, $x_k = kT_s$, etc. We also assume that the first derivative, \dot{y}_0 ,

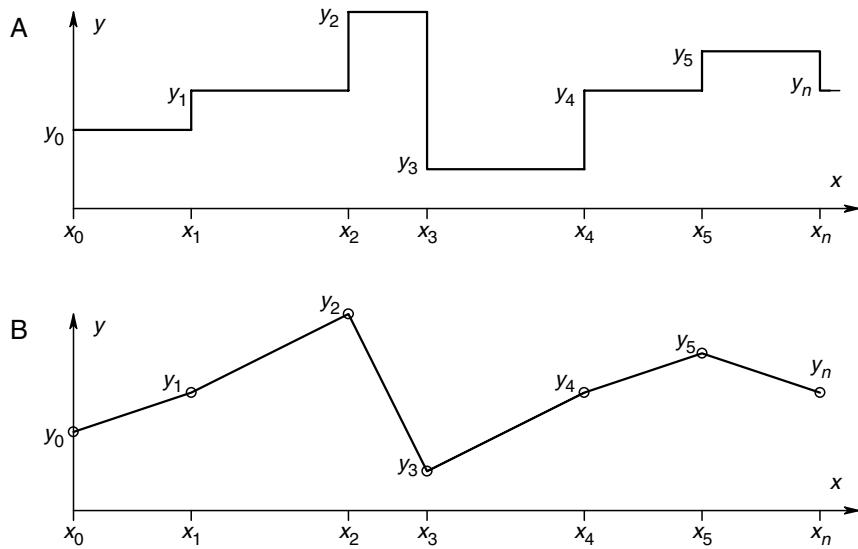
**FIGURE 10.13**

Illustration of zero order (A) and first order (B) polynomial approximations for interpolating between a set of discrete data points in order to regenerate the continuous time function from which the samples were taken.

is known or well-estimated. To calculate the cubic spline for the first three data points, $k=0, 1, 2$, we can write:

$$y_0 = a \quad (10.89a)$$

$$y_1 = a + bT_s + cT_s^2 + dT_s^3 \quad (10.89b)$$

$$y_2 = a + b2T_s + c4T_s^2 + d8T_s^3 \quad (10.89c)$$

Also, the first derivative at $x_0=t=0$ is, from equation 10.87 above:

$$\dot{y}_0 = b \quad (10.89d)$$

We wish to find the coefficients a, b, c and d for the $k=0, 1, 2$ spline. This may be done by simultaneously solving equations 10.89a–d, which gives:

$$a = y_0 \quad (10.90a)$$

$$b = \dot{y}_0 \quad (10.90b)$$

$$c = [-7y_0 + 8y_1 - \dot{y}_2 - 6T_s y_0]/(4T_s^2) \quad (10.90c)$$

$$d = [3y_0 - 4y_1 + \dot{y}_2 + 2T_s y_0]/(4T_s^3) \quad (10.90d)$$

Substituting these values into the cubic spline formula yields:

$$y(t) = y_0 + \dot{y}_0 t + \frac{-7y_0 + 8y_1 - y_2 - 6T_s \dot{y}_0}{4T_s^2} t^2 + \frac{3y_0 - 4y_1 + y_2 + 2T_s \dot{y}_0}{4T_s^3} t^3 \quad (10.91)$$

over the interval $0 \leq t \leq 2T_s$. By recursion, we can write a general cubic spline formula valid for any triplet of data points T_s seconds apart, (y_k, y_{k+1}, y_{k+2}) as:

$$\begin{aligned} S_3(t) = y_k + \dot{y}_k(t - kT_s) + & \frac{-7y_k + 8y_{k+1} - y_{k+2} - 6T_s\ddot{y}_k}{4T_s^2}(t - kT_s)^2 \\ & + \frac{3y_k - 4y_{k+1} + y_{k+2} + 2T_s\ddot{y}_k}{4T_s^3}(t - kT_s)^3 \end{aligned} \quad (10.92)$$

The first derivative estimate needed for the cubic spline of the next, adjacent triad of data points is given by:

$$\dot{y}_{k+2} = \dot{y}_k + \frac{2(y_{k+2} - 2y_{k+1} + y_k)}{T_s} \quad (10.93)$$

Note that the first 'knot' is at $k=0$, the second knot is at $k=2$, the third at $k=4$, etc. As an example of calculating the cubic splines of a small data set using equations 10.92 and 10.93, let $T_s = 1$ s, $\dot{y}_0 = 1$, $y_0 = 0$, $y_1 = 1$, $y_2 = 2$, $y_3 = 1$, $y_4 = 0$. Over the first interval, $0 \leq t \leq 2T_s = 2$ s, we have:

$$S_{0,2}(t) = 0 + (1)t + (0)t^2 + (0)t^3 = t \quad (10.94)$$

The derivative estimate at $k=2$ is:

$$\dot{y}_2 = 1 + \frac{2[2 - 2(1) + 0]}{1} = 1 \quad (10.95)$$

and over the interval, $2T_s \leq t \leq 4T_s$, the spline is:

$$S_{2,4}(t) = 2 + (1)(t - 2) + \frac{[-7(2) + 8(1) - 6(1)(1)]}{4(1)}(t - 2)^2 + \frac{[3(2) - 4(1) + (0) + 2(1)(1)]}{4(1)}(t - 2)^3 \quad (10.96)$$

Many other types of spline functions exist, as do variations on the cubic spline. For example, if both the first and second derivative of the sampled variable are known at $t=0$. then the cubic spline, $S(t)$, can be found between consecutive pairs of data points instead of between triplets of points. Using a development similar to that given above for three point cubic splines, it is possible to show that for $kT_s \leq t \leq (k+1)T_s$:

$$Sk(t) = \dot{y}_k + y_k(t - kT_s) + (\ddot{y}_k/2)(t - kT_s)^2 + \frac{y_{k+1} - y_k - T_s\ddot{y}_k - (T_s^2/2)\ddot{y}_k}{T_s^3}(t - kT_s)^3 \quad (10.97)$$

and

$$\dot{y}_k = \dot{y}_{k-1} + T_s\ddot{y}_{k-1} + 3\left(\frac{y_k - y_{k-1} - T_s\ddot{y}_{k-1} - (T_s^2/2)\ddot{y}_{k-1}}{T_s}\right) \quad (10.98)$$

$$\ddot{y}_k = \ddot{y}_{k-1} + 6\left(\frac{y_k - y_{k-1} - T_s\ddot{y}_{k-1} - (T_s^2/2)\ddot{y}_{k-1}}{T_s^2}\right) \quad (10.99)$$

Although the scope of this discussion has been an introduction to cubic splines, it should be noted that the technique of splines is easily applied to data having non-uniform sampling periods. The resulting spline approximations can then be 'sampled' at a uniform rate and these M samples can be used to calculate a conventional DFT of $\{x_k, y_k\}_N$, with $M > N$. Also, spline recursion equations can be put in tridiagonal matrix form for simultaneous solution of $S_k(t)$ from the entire data set at once [Scheid, 1990]. Some of the other interpolation methods that exist for analog data interpolation include the methods of LaGrange, Hermite, Gauss, Stirling, Newton and the least squares method of fitting a continuous linear equation to a data set, $\{x_k, y_k\}_N$ (described in chapter 1).

10.5.1 Estimating Missing Data at the Sampling Instants

We occasionally have a data sequence where one or more data points are missing, perhaps because the experimenter neglected to record them, or because a burst of noise completely masked the true data value. If data point y_k is missing in a large data set and data set does not vary abruptly in time, then the missing data point can be found by linear interpolation:

$$y_k = (y_{k+1} + y_{k-1})/2 \quad (10.100)$$

That is, the missing y_k can be approximated by the average of its nearest neighbors. In the event that the missing data point is part of a curve with non-zero second derivative, a more accurate estimate can be derived from the solution of four simultaneous, cubic algebraic equations [Williams, 1988]. Two adjacent data points on either side of the missing data point are used:

$$y_k = (-y_{k-2} + 4y_{k-1} + 4y_{k+1} - y_{k+2})/6 \quad (10.101)$$

Since cubic splines work with non-uniform data x values, they also can be used to find a missing y_k . We assume that y_{k-2} , y_{k-1} and y_{k+1} are known and can be used in the spline calculation. \dot{y}_{k-2} must also be known or estimated.

A data point missing at the beginning or end of a finite data sequence, $\{x_k, y_k\}_N$, requires numerical extrapolation, which can also be carried out by a polynomial estimation routine using existing data [Williams, 1988]. The accuracy of extrapolation decreases as the interval (x_N, x_{N-1}) increases, and as the bandwidth of $Y(j\omega)$ increases. If data is changing monotonically and is not noisy, then a reasonable value for y_N can be calculated.

We give an example of a simple extrapolator that uses the three point central difference differentiator to estimate the slope of y_k at the end of the sequence, where $k = N - 1$. This slope is found from equation 10.43 and is given by:

$$m_{k-1} = (y_{N-1} - y_{N-3})/2T_s \quad (10.102)$$

Simple algebra tells us that the extrapolated value of y at $t = NT_s$ is:

$$y_N = y_{N-1} + m_{k-1}T_s = 1.5y_{N-1} - 0.5y_{N-3} \quad (10.103)$$

This same, basic procedure can be turned around to find an estimate of y_{-1} at $t = -T_s$.

The three point, cubic spline approach can also be used to find an estimate for y_N at the end of the sequence $\{x_k, y_k\}_N$. The trick here is to redefine the local time origin so that the new $y_1 = y_N$ (value to be extrapolated), $y_0 = y_{N-1}$ (the last data sample of the sequence),

$y_{-1} = y_{N-2}$ and $y_{-2} = y_{N-3}$ in the sequence. The first derivative at y_0 is estimated by the three point central difference method:

$$\dot{y}_0 = (y_0 - y_{-2})/2T_s = (y_{N-1} - y_{N-3})/2T_s \quad (10.104)$$

Equations 10.89a, b and c can now be used with appropriate changes of indices and the derivative estimate of equation 10.104 to show that in terms of the original data set,

$$y_N = 3y_{N-2} - 2y_{N-3} \quad (10.105)$$

Proof of equation 10.105 is left as an exercise for the interested reader. Note that y_{N-1} curiously drops out of this cubic spline extrapolation formula.

10.6 Chapter Summary

Digital Signal Processing and its related mathematical field of numerical analysis is of considerable importance to all branches of engineering, particularly to electrical and systems engineers, as well as to measurement and instrumentation engineers. Many textbooks have been written in these areas and we cite a representative group in the bibliography and references for this chapter. Our purpose in this chapter was to introduce some key concepts in DSP, of particular use in instrumentation and measurement system design.

We first discussed the z-transform, and digital filters as a linear means of operating on a sequence of sampled data in order to carry out a discrete filtering operation analogous to an analog filter, such as a low-pass or bandpass filter. We showed that the relevant range of input signal frequencies in digital filtering is from 0 to the Nyquist frequency, $f_s/2$. Examples of finite impulse response (FIR) low-pass and bandpass filters were given. We also examined the time domain structure of the filtering process, showing how a computer program running in real time can implement a filtering algorithm by a series of synchronous delays, multiplications by constants and summations. Recursive, infinite (duration) impulse response (IIR) filters were introduced, and a second order, high Q , notch filter implementation was given as an example.

In measurement systems, the operations of digital differentiation and integration are also important. Several basic differentiation and integration routines were discussed and their algorithms described in the frequency domain. Integration routines were shown to be recursive, while differentiators were shown to generally use FIR algorithms.

A significant portion of this chapter was devoted to introducing the discrete Fourier transform (DFT). We examined how it can approximate the continuous Fourier transform (CFT) and how window functions which are used to prevent sidelobe 'leakage'. The basics of the fast Fourier transform (FFT) algorithm to compute the DFT were described, an eight point FFT example was illustrated and an expression for FFT/DFT efficiency was derived. The commercial availability of DSP and FFT plug-in boards for PCs was cited as the state-of-the-art way to accelerate FFT and digital filter calculations in real time.

Finally, we considered the problem of reconstructing continuous, analog data between data samples. We introduced cubic splines as an efficient way to reconstruct analog data, as well as to interpolate missing data points and to extrapolate a data point at the end of a data sequence.

Problems

10.1 A certain continuous time function, $f(t)U(t)$, has the z-transform:

$$F(z) = \frac{z}{z - 0.367879}$$

- (A) Plot its pole and zero in the complex z plane.
- (B) Use long division to find $F(z)$ as an infinite series with terms of the form z^{-n} , $n = 0, 1, 2, \dots, \infty$.
- (C) Write the sampled time function, $f(nT) = f(n)$ by inspection in closed form. What is the original continuous, $f(t)$?

10.2 Consider the discrete, linear, time invariant system (DLTIS):

$$H(z) = \frac{z^2}{z^2 - 0.6z + 0.05}$$

- (A) Factor the denominator to find the system's two real poles in the z plane.
- (B) Do a *partial fraction expansion* on $H(z)/z$ to find $H(z)$ as the sum of two realpole terms.
- (C) Write the weighting function, $h(n)u(n)$, as the sum of two inverse z-transforms.

10.3 A DLTIS, $H(z)$, has the weighting function, $h(n) = u(n)e^{-naT}$, where T is the sampling period, and $u(n) = \delta(t) + \delta(t - T) + \delta(t - 2T) + \dots + \delta(t - nT) + \dots = \sum_{n=0}^{\infty} \delta(t - nT)$. This system has the discrete input, $x(n) = u(n)e^{-nbT}$ for $n = 0, 1, 2, \dots, \infty$. Find the system's discrete, time domain output, $y(n)$. Hint: Use a partial fraction expansion on $Y(z) = X(z)H(z)$.

10.4 Use the relation $\sum_{n=0}^{N-1} \rho^n = (1 - \rho N)/(1 - \rho)$ to find the z transform, $G(z)$, of the truncated exponential decay, $g(n) = e^{-naT}$, defined for $0 \leq n \leq N - 1$, and zero for $n < 0$ and $n \geq N$. Hint: Note that there is an $(N - 1)$ th order pole at the origin of the z plane for $G(z)$.

10.5 Consider the DLTI system defined by the difference equation $y(n) = x(n) + (1/3)y(n - 1)$. $y(n)$ is the present output, $x(n)$ the present input and $x(n) = u(n)$, a unit step. There are zero initial conditions. Find an expression for, plot and dimension $y(n)$, $n = 0, 1, 2, 3, 4, \dots$

10.6 A causal FIR filter has the impulse response $y(n) = x(n) - x(n - 4)$, where $x(t) = \delta(0)$ (a '1' at $t = nT = 0$).

- (A) Write the filter's transfer function, $H(z)$, as a polynomial in z .
- (B) Plot the filter's poles and zeros in the z plane.
- (C) Find the filter's frequency response, $H(j\omega)$, by substituting $z = e^{j\omega T}$ into $H(z)$. Use the Euler relation to find, plot and dimension $|H(j\omega)|$ and $\angle H(j\omega)$ vs ω .
- (D) Find and plot the filter's output, $y(n)$, in response to a unit step input to the filter.

10.7 The DE for a causal FIR notch filter is given as $y(n) = b_0x(n) + b_1x(n - 1) + b_2x(n - 2)$.

- (A) Find the filter's transfer function, $H(z)$.
- (B) Find an expression for $H(j\omega)$, magnitude and phase. Let $b_0 = b_2$. Use the Euler relation for $\cos \theta$.
- (C) Find the b_0 and b_1 values required so that $H(0) = 1/0$ and $H(\pi/4T) = 0$.

10.8 Find the DFT, $X(k)$, of the sequence, $x(n) = \sin(n\pi/N)$, $n = 0, 1, 2, \dots, N - 1$. (Hint: Write $x(n)$ in exponential form.)

10.9 Find the DFT, $H(k)$, of the Hamming window sequence, $h(n) = 0.54 + 0.42 \cos(n\pi/N)$, $n = 0, 1, 2, \dots, N - 1$.

10.10 Find the DFT, $X(k)$, of the eight point sequence $x(n) = 0$, $0 \leq n \leq 4$; $x(n) = 1$, $5 \leq n \leq 6$; $x(n) = 0$, $n = 7$.

10.11 $x(n)$ is the sampled input, $h(n)$ is the weighting function of a DTIF and $y(n)$ is the output number sequence. Find $y(n)$ by linear, discrete convolution, following the procedure of Figure 10.1. See Figure P10.11.

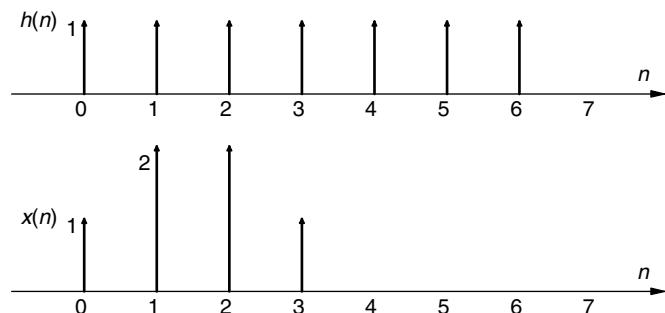


FIGURE P10.11

11

Examples of the Design of Measurement Systems

11.1 Introduction

In this final chapter, we describe the needs for, the systems design philosophies and the means of realizing systems used for measurement—a self-nulling, microdegree polarimeter to measure glucose in bioreactors; the location and intensity of partial discharges in the insulation of coaxial, high voltage, power cables; design of a closed loop, constant phase, pulsed laser velocimeter; design of capacitive sensors for the detection of hidden objects. These four systems have been chosen to illustrate certain design principles, as well as to describe diverse, measurement systems.

The author and his graduate students have designed and developed these unique measurement systems. Therefore, their designs reflect their personal bias toward certain circuit architectures and manufacturers' components. It is well appreciated that systems design in electrical engineering is an art as well as a science. The designer must balance criteria on system performance with cost and ease of manufacturing, safety, system reliability, ergonomics, aesthetics and environmental concerns. There are often several designs that can lead to the same end product, and no clear set of factors appears to favor one design over the others. It is in these fuzzy cases that a designer has greater freedom to show personal preferences (i.e. exhibit his or her 'design style').

In the first design example given below, we describe the need for, the design approach to, the specifications for, and a critique of, the performance of a partial discharge measurement system.

11.2 A Self-nulling, Microdegree Polarimeter to Measure Glucose in Bioreactors

A *polarimeter* is an instrument used to measure the angle of rotation of linearly polarized light when it is passed through an optically active material (Polarimeters were introduced in Section 6.7.1). There are many types of polarimeters, as also a number of optically active substances found in living systems. Probably the most important is D-glucose dissolved in water. Clear biological liquids such as urine, extracellular fluid and the aqueous humor of the eyes contain dissolved D-glucose, whose molar concentration is proportional to the blood glucose concentration [Northrop, 2002]. Glucose concentration is an important parameter in the fermentation industry and also in biotechnology applications where genetically engineered cells are grown in bioreactors.

Using polarimetry, D-glucose concentration can be measured by the simple relation:

$$\phi = [\alpha]_{\lambda}^T CL \text{ degrees} \quad (11.1)$$

where ϕ is the measured optical rotation of linearly polarized light (LPL) of wavelength λ passed through a sample chamber of length L , containing the optically active analyte at concentration C and temperature T . The constant, $[\alpha]_{\lambda}^T$, is called the *specific optical rotation* of the analyte. Its units are generally in degrees per (optical path length unit \times concentration unit). $[\alpha]_{\lambda}^T$ for D-glucose in 25°C water and at 512 nm, is 0.0695 millidegrees/(cm \times g/dl). Thus, the optical rotation of a 1 g/l solution of D-glucose in a 10 cm cell is 69.5 millidegrees. Note that $[\alpha]_{\lambda}^T$ can have either sign, depending on the analyte, and if N optically active substances are present in the sample chamber, the *net optical rotation* is given by superposition as:

$$\phi = L \sum_{k=1}^N [\alpha]_{\lambda}^{T_k} C_k \quad (11.2)$$

To describe how a polarimeter works, we must first describe what is meant by linearly polarized light. When light from an incoherent source, such as a tungsten lamp, is treated as an electromagnetic wave phenomenon (rather than photons), we observe that the propagating radiation is composed of a broad spectrum of wavelengths and polarization states. To facilitate the description of LPL, let us examine a monochromatic ray propagating in the z direction at the speed of light in the medium in which it is travelling. This velocity in the z direction is $v = c/n$ m/s. c is the velocity of light in *vacuo* and n is the *refractive index* of the propagation medium (generally > 1.0). A beam of monochromatic light can have a variety of polarization states, including *random* (unpolarized), *circular* (CW or CCW), *elliptical* and *linear* [Balanis, 1989]. Figure 11.1 illustrates a LPL ray in which the \mathbf{E} vector propagates entirely in the xz plane and its orthogonal \mathbf{B}_y vector lies in the yz plane. \mathbf{E} and \mathbf{B} are in phase. Both vectors are functions of time and distance. For the $\mathbf{E}_x(t, z)$ vector:

$$\mathbf{E}_x(t, z) = \mathbf{E}_{ox} \cos \left[\frac{2\pi c}{\lambda} t - \frac{2\pi}{\lambda} z \right] = \mathbf{E}_{ox} \cos[kct - kz] \quad (11.3)$$

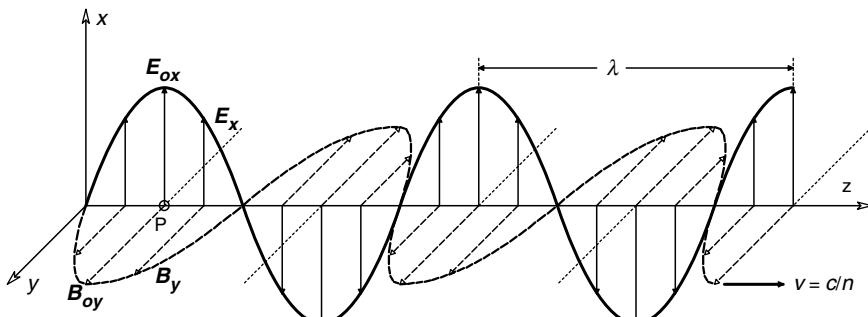
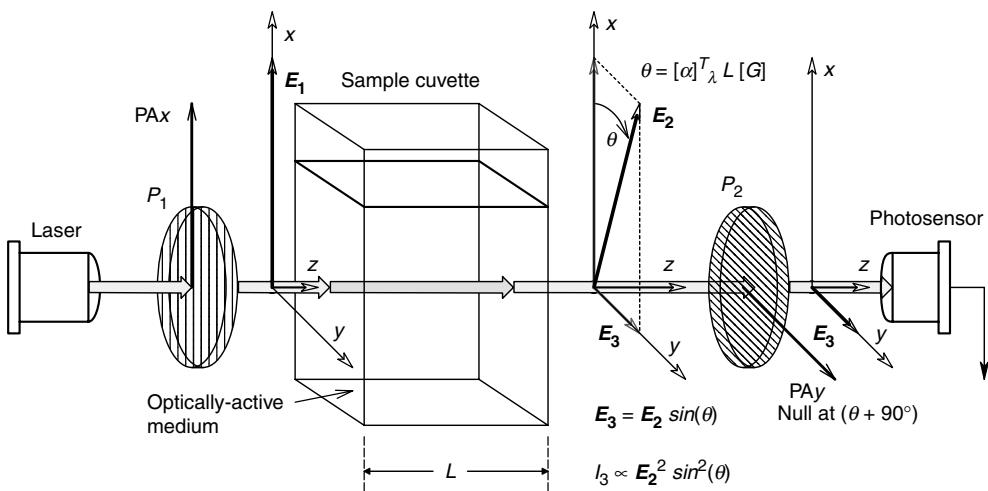


FIGURE 11.1

Propagation of a linearly polarized light wave in space. The wave is electromagnetic. It has an \mathbf{E} vector which lies in the xz plane and a \mathbf{B} vector that lies in the yz plane.

**FIGURE 11.2**

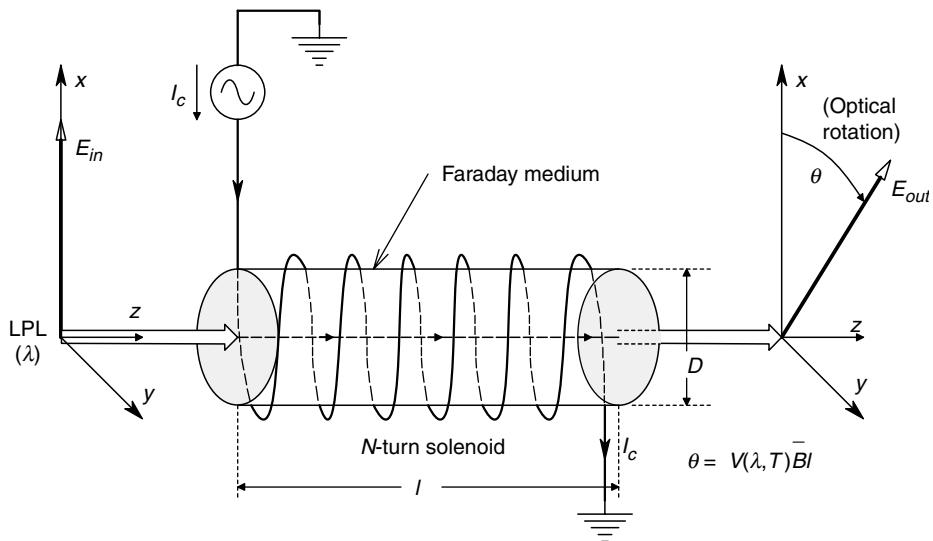
Schematic of a simple manual polarimeter. Polarizer P_1 creates linearly polarized light in which the E vector lies in the xz plane as the light propagates. As a result of propagating through a cuvette containing an optically active material (such as D-glucose), the propagation plane of the E vector is rotated by θ degrees away from the xz plane. If polarizer P_2 is rotated ($90^\circ + \theta$) from the x axis, there will be an intensity null at the photosensor. Thus, by reading θ we know the optical activity of the analyte in the cuvette.

where $k \equiv 2\pi/\lambda$. Note that the E vector need not lie in the xz plane, its plane of propagation can be tilted by some angle θ with respect to the x axis and still propagate in the z direction (the orthogonal B vector is rotated θ with respect to the y axis). When describing LPL, we examine the positive maximums of the E vector and make a vector with this maximum and at an angle θ with the x axis in the xy plane.

Figure 11.2 illustrates what happens to LPL of wavelength λ passing through a sample chamber containing an optically active analyte. Polarizer P_1 converts the input light to a linearly polarized beam which is passed through the sample chamber. As a result of the interaction of the light with the optically active solute, the emergent LPL ray is rotated by an angle, θ . In this figure, the analyte is *dextro rotary*, that is, the emergent E vector, E_2 , is rotated clockwise when viewed in the direction of propagation (along the $+z$ axis). (D-glucose is dextro rotary.) To measure the optical rotation of the sample, a second polarizer, P_2 , is rotated until a null in the light intensity striking the photosensor is noted (in its simplest form, the photosensor can be a human eye). This null is when P_2 's pass axis is orthogonal to E_2/θ . Since polarizer P_1 is set with its pass axis aligned with the x axis (i.e. at 0°), it is easy to find θ and subsequently, the concentration of the optically active analyte, $[G]$. In the simple polarimeter of Figure 11.2 and other polarimeters, the polarizers are generally made from calcite crystals (Glan calcite), and have *extinction ratios* of ca. 10^4 . (Extinction ratio is the ratio of LPL intensity passing through the polarizer when the LPL's polarization axis is aligned with the polarizer's pass axis, to the intensity of the emergent beam when the polarization axis is made 90° (orthogonal) to the polarizer's axis.)

To make an electronic polarimeter, we need an electrical means of nulling the polarimeter output, as opposed to physically rotating P_2 around the z axis. One means of generating an electrically controlled optical rotation is to use the *Faraday magneto-optical effect*. Figure 11.3 illustrates a *Faraday Rotator* (FR). The FR has two major components:

1. A core of magneto-optically active medium. Many transparent gasses, liquids and solids exhibit magneto-optical activity (e.g. a lead glass rod with optically flat

**FIGURE 11.3**

A Faraday rotator. LPL propagating in the same (z) direction as the magnetic field from the solenoidal coil in a Faraday medium will be optically rotated. Faraday media include certain gasses, liquids (e.g. water) and solids (e.g. glass).

ends, or a glass test chamber with optically flat ends containing a gas or liquid can be used).

2. A solenoidal coil wound around the rod or test chamber (core) that generates an axial magnetic field, \mathbf{B} , inside the rod or chamber, collinear with the entering LPL beam. The exiting LPL beam undergoes optical rotation according to the simple Faraday relation [Hecht, 1987] given by:

$$\theta_m = V(\lambda, T) \bar{B}_z l \quad (11.4)$$

where θ_m is the Faraday optical rotation in degrees, $V(\lambda, T)$ is the magneto-optically active material's *Verdet constant*, having units of degrees/(tesla meter). Verdet constants are given in a variety of challenging units, such as 10^{-3} minutes of arc/(gauss cm). \bar{B}_z is the average axial flux density collinear with the optical path and l is the length of the path where the LPL is exposed to the axial \mathbf{B} field. A material's Verdet constant *increases* with *decreasing wavelength* and *increasing temperature*. When a Verdet constant is specified, the wavelength and temperature at which it was measured must be given. For example, the Verdet constant of distilled water at 20°C and 578 nm is 218.3°/(tesla m) [Chapter 8, Hecht, 1987]. This value may appear large, but the length of the test chamber on which the solenoid is wound is 10 cm = 10^{-1} m and 1 T = 10^4 gauss, so actual rotations tend to be less than $\pm 5^\circ$. The Verdet constant for lead glass under the same conditions is about six times larger. Not all Verdet constants are positive. For example, an aqueous solution of ferric chloride and solid amber have negative Verdet constants.

Two well-known formulas for axial \mathbf{B} inside a solenoid are given below [Krauss, 1953]. At the center of the solenoid, on its axis:

$$\mathbf{B}_z = \frac{\mu N I}{\sqrt{4R^2 + l^2}} \text{ T} \quad (11.5)$$

At either end of the solenoid, on its axis:

$$\mathbf{B}_z = \frac{\mu NI}{2\sqrt{R^2 + l^2}} \text{ T} \quad (11.6)$$

It can be shown by averaging over l , that the \bar{B}_z over the length, l , of the solenoid is approximately:

$$\bar{B}_z \approx \frac{3\mu NI}{4\sqrt{4R^2 + l^2}} \rightarrow \frac{3\mu NI}{4l} \text{ T} \quad \text{when } l \gg 2R \quad (11.7)$$

It should be clear that a Faraday rotator can be used to electrically null a simple polarimeter. Figure 11.4 gives an illustration of a simple open loop dc, electrically nulled polarimeter. The dc FR current is varied until the output photosensor registers a null ($E_4 = 0$). P2's axis remains orthogonal to P1's axis. At null, the optical rotation from the FR is equal and opposite to the optical rotation from the optically active analyte. Thus:

$$V(\lambda, T) \frac{3\mu NI}{4l} l = [\alpha]_\lambda^T CL \quad (11.8)$$

and the concentration of OA analyte is:

$$C = \frac{3V\mu NI}{4[\alpha]L} \quad (11.9)$$

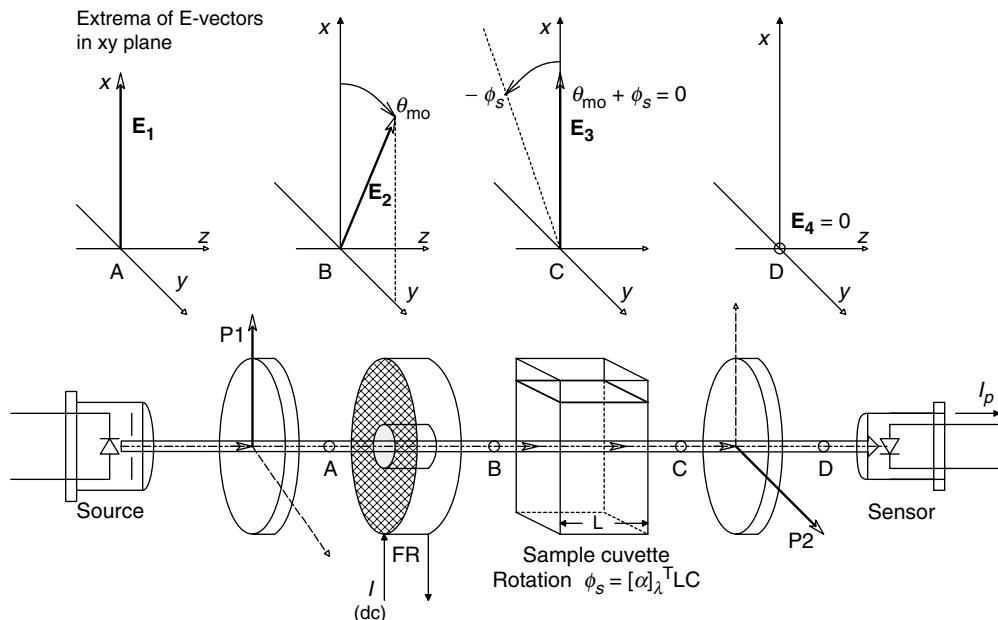
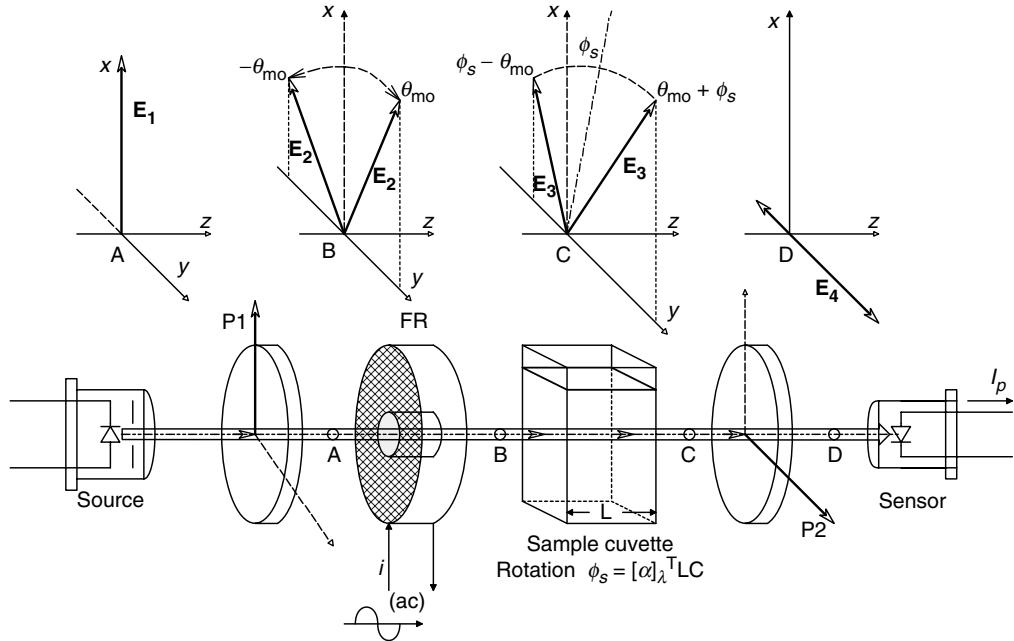


FIGURE 11.4

Diagram of a simple polarimeter in which nulling is done electromagnetically with a Faraday rotator (FR), instead of by rotating polarized P2. The optical rotation by the FR is made equal and opposite to the rotation by the sample, with P2 set at 90° to P1, to get the null.

**FIGURE 11.5**

Polarization measuring system, which uses *polarization angle modulation* of the LPL entering the sample. Sinusoidal ac current of frequency, ω_m , is passed through the coil of the FR to rock the E vector plane $\pm\theta_{mo}$ around the xz plane. The intensity of the emergent ray from P2 can be shown to have both ω_m and $2\omega_m$ components. The fundamental frequency term's intensity is proportional to ϕ_s , the sample's optical rotation of LPL.

Note that C is dependent on three well-known parameters (N , I and L), and two that are less well known and are temperature dependent (V and $[\alpha]$). This type of operation yields accuracy to the tenths of a degree.

Figure 11.5 illustrates an improvement on the accuracy of the static polarimeter shown in Figure 11.4. An audio frequency ac current is passed through the FR, causing the emergent polarization vector at point B to rock back and forth sinusoidally in the xy plane at audio frequency, $f_m = \omega_m/2\pi$ Hz. The maximum rocking angle is θ_{mo} . Next, this rocking, or angle modulated E_2 vector is passed through the optically active sample, where at any instant, the angle of E_2 has the rotation ϕ_s added to it, as shown at point C. Next, the asymmetrically rocking E_3 vector is passed through polarizer P2, which only passes the y components of E_3 , as E_{4y} at point D. Thus $E_{4y}(t)$ can be written: I (dc)

$$E_{4x}(t) = E_{4yo} \sin[\phi_s + \theta_{mo} \sin(\omega_m t)] \quad (11.10)$$

The sinusoidal component in the $\sin[*]$ argument is from the audio frequency polarization angle modulation. The instantaneous intensity of the light at D, $i_4(t)$, is proportional to $E_{4y}^2(t)$:

$$i_4(t) = E_{4yo}^2 [c \varepsilon_0 / 2] \sin^2[\phi_s + \theta_{mo} \sin(\omega_m t)] \text{ W/m}^2 \quad (11.11)$$

Now by the trigonometric identity, $\sin^2(x) = 1/2 [1 - \cos(2x)]$, we have:

$$i_4(t) = E_{4yo}^2 [c \varepsilon_0 / 2] 1/2 \{1 - \cos[2\phi_s + 2\theta_{mo} \sin(\omega_m t)]\} \quad (11.12)$$

Since the angle argument of the cosine is small, i.e. $|2(\phi_s + \theta_{mo})| < 3^\circ$, we may use the approximation, $\cos(x) \cong (1 - x^2/2)$. Assuming that the instantaneous photosensor output voltage, $v_p(t)$, is proportional to $i_4(t)$, so we can finally write:

$$v_p(t) = K_P i_4(t) \cong K_P \{E_{4y_0}^2(c \varepsilon_0/2)\}^{1/2} \left\{ 1 - \left[1 - \frac{4\phi_s^2 + 4\theta_{mo}^2 \sin^2(\omega_m t) + 8\phi_s\theta_{mo} \sin(\omega_m t)}{2} \right] \right\} \downarrow \quad (11.13)$$

$$v_p(t) = K_V [\phi_s^2 + \theta_{mo}^2]^{1/2} (1 - \cos(2\omega_m t)) + 2\phi_s\theta_{mo} \sin(\omega_m t) \quad (11.14)$$

where $K_V = K_P E_{4y_0}^2(c \varepsilon_0/2)$.

Thus, the photosensor output voltage contains three components:

1. A dc component which is not of interest
2. A double frequency component which has no information on ϕ_s
3. A fundamental frequency sinusoidal term whose peak amplitude is proportional to ϕ_s

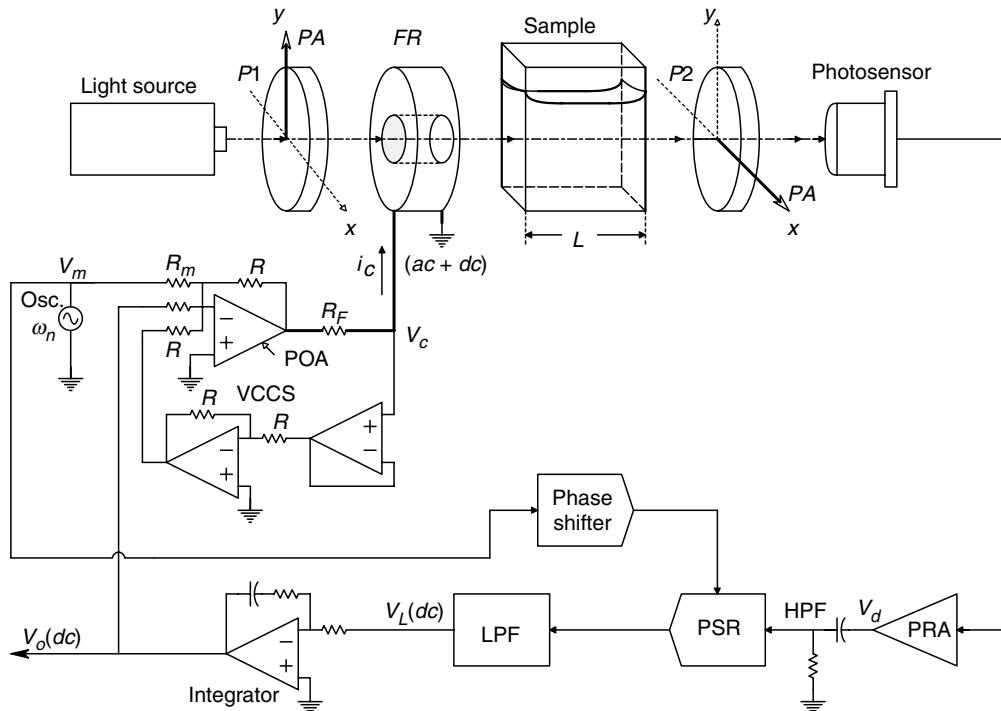
By using a high-pass filter, we can block the dc components, and by using a phase sensitive rectifier synchronized to the FR modulating current sinusoid at ω_m , we can recover a dc signal proportional to ϕ_s , while rejecting the $2\omega_m$ sinusoidal term.

By feeding back a dc current proportional to ϕ_s into the FR, the system can be made closed loop and self-nulling. The first such feedback polarimeter was invented by Gilham in 1957. It has been improved and modified by various workers since then [Northrop, 2002; Cameron and Coté, 1997; Coté, Fox and Northrop, 1990, 1992, 1993; Rabinovitch, March and Adams, 1982]. Instruments of this sort can resolve ϕ_s to better than $\pm 20 \mu\text{degrees}$.

A conventional, closed loop, polarization angle modulated, Gilham polarimeter is shown in Figure 11.6. The dc light source can be a laser or laser diode. (We used a 512 nm (green) diode laser.) The light from the laser is passed through a calcite Glan laser polarizer, P1, to improve its degree of polarization and then through the Faraday rotator. Two things happen in the FR. The input LPL is polarization angle-modulated at frequency f_m . θ_{mo} is made about 2° . After passing through the sample, the angle modulation is no longer symmetrical around the y axis because the optically active analyte's rotation has been added to the modulation angle. As shown above, this causes a fundamental frequency sinusoidal voltage to appear at the output of the high-pass filter (HPF) whose peak amplitude is proportional to the desired ϕ_s . The phase sensitive rectifier and LPF output a dc voltage, V_L , proportional to ϕ_s .

V_L is integrated and the dc integrator output, V_o , is used to add a dc nulling current component to the FR ac input. This nulling current rotates the angle modulated output of the FR so that, when it passes through the sample, there is no fundamental frequency term in the photosensor output, V_d . Thus, the system is at null and the integrator output, V_o , is proportional to $-\phi_s$. Integration is required to obtain a type 1 control system that has zero steady state error to a constant input [Ogata, 1990]. (the integrator is given a zero so that it has a proportional plus integral (PI) transfer function to make the closed loop system stable).

Browne, Nelson and Northrop (1997) devised a feedback polarimetry system that has microdegree resolution, designed to measure glucose in bioreactors. Their system is illustrated in Figure 11.7. The system of Browne *et al* works the same as the polarimeter

**FIGURE 11.6**

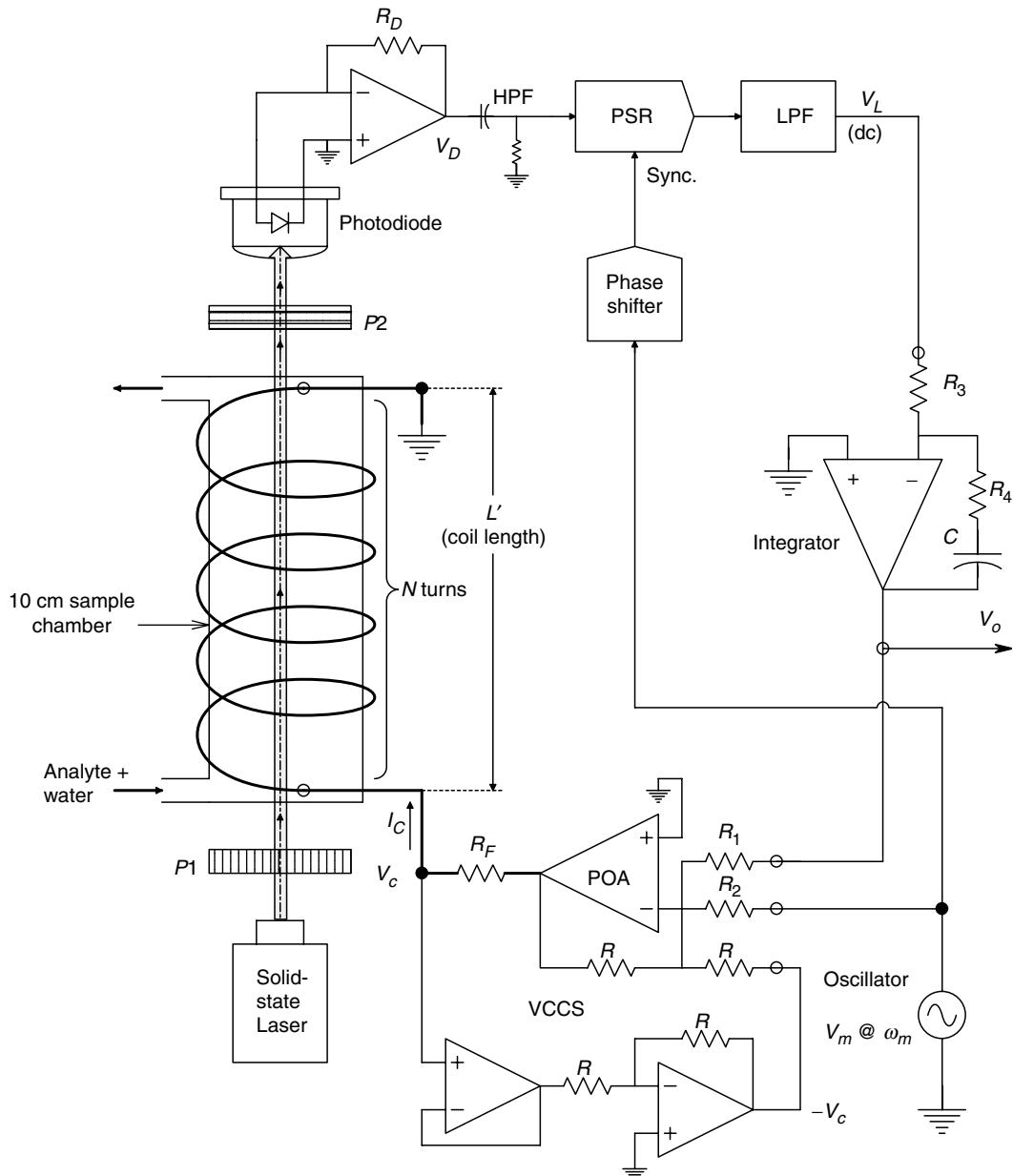
An improved, Gilham type polarimeter devised by the author. This is a Type I feedback system in which the same FR is used to ac modulate the E vector and also to obtain a dc null of the sample's optical rotation.

system in Figure 11.6, with the exception that the FR and the sample cuvette are replaced with a single cylindrical chamber around which the solenoid coil is wound. The 10 cm sample chamber contains D-glucose dissolved in water, as well as other dissolved salts, some of which may be optically active. (Cells and proteins are filtered out, leaving clear liquid in the sample chamber.)

The solenoid current is supplied from a 3 op-amp precision VCCS (Section 4.4). The reason for using a VCCS to drive the coil is because the Faraday magneto-rotary effect is proportional to the axial \mathbf{B} field, which in turn, depends on the current through the coil. Since the coil is lossy, it gradually heats up from the power dissipated by the ac modulation excitation and the dc nulling current. As the coil wire temperature increases, its resistance goes up. Thus, if a voltage source (op-amp) were used to drive the coil, \mathbf{B} would be a function of coil temperature as well as the input voltage, V_c . The \mathbf{B} temperature dependence is eliminated by using the VCCS.

The FR can be eliminated from this polarimeter because the water solvent in the chamber has a Verdet constant and when it is subjected to ac and dc axial magnetic fields, they cause optical rotation of the transmitted LPL. The glucose solute is optically active, so we find that in the same path length, optical rotation occurs because of the dissolved D-glucose and Faraday rotation occurs because of the action of \mathbf{B} on the water.

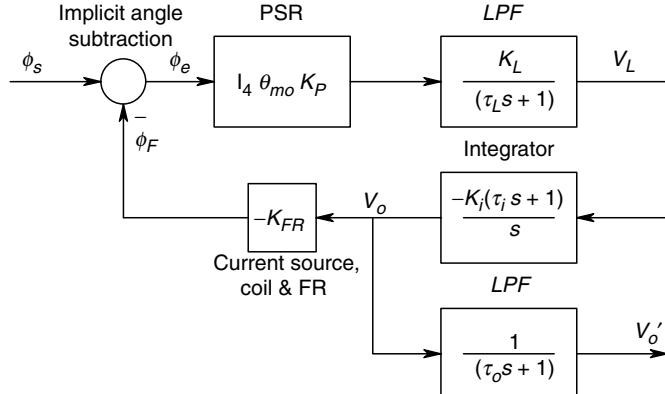
One might ask whether the dissolved glucose itself has a Verdet constant and is also acted on by the axial \mathbf{B} field. It probably does, but the Faraday magneto-rotary effect on the dissolved glucose is much smaller than that on the water solvent. In any case, the Faraday rotation of all solutes present adds (or subtracts) a small amount to the polarization angle modulation produced by the ac \mathbf{B} component acting on the water.

**FIGURE 11.7**

The closed loop, Gilham type polarimeter devised by Browne *et al* (1997). The external glass FR in Figure 11.6 has been replaced by using the water in the analyte solution as a Faraday medium.

The dc (feedback) nulling current too may act on the solutes, affecting system gain, K_{FR} , slightly but uniformly.

It is possible to summarize the operation of this interesting system by a block diagram, shown in Figure 11.8. The input to the measurement system is the physical angle of rotation of the polarization vector, ϕ_s , caused by passing the LPL through the optically active analyte. The negative feedback of the system output voltage, V_o , leads to a Faraday counter-rotation of the LPL, ϕ_F , forming a net (error) rotation, ϕ_e . The dc counter-rotation,

**FIGURE 11.8**

Systems block diagram of the polarimeters of Figures 11.6 and 11.7.

ϕ_F , is simply:

$$\phi_F = V \bar{B} L' = V(3/4) \frac{\mu N}{L'} L' I_c = [V(3/4) \mu N] \left[\frac{-V_o R}{R_1 R_2} \right] = -K_{FR} V_o \quad (11.15)$$

Clearly, $K_{FR} = V(3/4) \mu N R / (R_1 R_2)$. The closed loop transfer function of the system is found from the block diagram:

$$\frac{V_o}{\phi_s}(s) = \frac{-I_4 \theta_{mo} K_P K_L K_i (\tau_i s + 1)}{s^2 \tau_L + s[1 + \tau_i I_4 \theta_{mo} K_P K_L K_i K_{FR}] + I_4 \theta_{mo} K_P K_L K_i K_{FR}} \quad (11.16)$$

It is of interest to consider the dc or steady state response of this closed loop polarimeter to a constant, ϕ_s . This can be shown to be:

$$\frac{V_o}{\phi_s} = \frac{-1}{K_{FR}} \quad (11.17)$$

Equation 11.17 is a rather serendipitous result. It tells us that the system's calibration depends only on K_{FR} , which depends on the transconductance of the VCCS, the dimensions and number of turns of the coil, the Verdet constant and the permeability, μ , of water. The dc gain is *independent* of the light intensity (I_4), the depth of polarization angle modulation (θ_{mo}), the frequency of modulation (ω_m), and the system's gain parameters (K_P , K_L , and K_i) and time constants (τ_L and τ_i).

The prototype system built and tested by Browne *et al* (1997) was evaluated for D-glucose concentrations (x) in the range 60–140 mg/dl. By least mean square error curve fitting, the output voltage (y) was found to follow the linear model,

$$y = 0.0544 x + 0.0846 \quad (11.18)$$

with $R^2 = 0.9986$. System resolution was $10.1 \mu^\circ/\text{mV}$ out, at 27°C with a 10 cm cell and the glucose sensitivity was 55.4 mV/mg\% . Noise in V_o was reduced by low-pass filtering (not shown in Figure 11.7).

Note that in this instrument, op-amps were used exclusively for all active components, including the photodiode current-to-voltage converter (transresistor) (*cf.* Section 2.6.4), the phase sensitive rectifier (*cf.* Section 12.2.3), the low-pass filter (*cf.* Section 7.2), the PI integrator (*cf.* Section 6.6), the oscillator (*cf.* Section 5.5), and the VCCS (*cf.* Section 4.4).

11.3 Design of a System to Detect, Measure and Locate Partial Discharges in High Voltage Coaxial Power Cables

In the process of aging, the insulation of high voltage, underground, coaxial power cables may develop physical defects such as water trees and other cavities. Under certain conditions, these defects can serve as the foci for massive insulation failures (arc overs). Long before catastrophic insulation failure occurs, the water trees and other defect sites, subject to periodic (ac) electric field stress in the dielectric, begin to arc over internally, producing *partial discharges* (PDs). Individual PDs are not fatal events for the cable, however, once initiated, over a period of time, they can cumulatively lead to one. It is our intention here to describe the signal conditioning circuitry involved in measuring PDs and not the sophisticated DSP techniques used to pinpoint their location in a buried cable.

A simple, equivalent, lumped capacitive circuit for a power cable with a PD site is shown in Figure 11.9. C_P is the total capacitance between the inner conductor and the outer braided shield. C_D is the tiny capacitance of the water tree or void across which the PD occurs, and C_S is the total series capacitance between the void and the inner conductor and the braid. Generally, $C_P \gg C_S, C_D$.

The PD occurs when the gas or vapor dielectric in the defect cavity breaks down under electrostatic stress, permitting the charge stored in the defect walls to dissipate, temporarily reducing the electric field in the defect. A simplified electrical equivalent of the arc process is shown in Figure 11.10.

At the next ac cycle, the PD is repeated, etc. The instantaneous applied voltage on the cable at which a given PD occurs is called the *inception voltage*. At 60 Hz, a coaxial,

FIGURE 11.9

Simplified equivalent circuit of a coaxial power cable at 60 Hz. C_D is the equivalent capacitance of the hollow insulation defect that will generate a partial discharge. C_D appears in series with a small equivalent capacitance, C_S , consisting of the capacitance of the good insulation surrounding C_D . C_P is the capacitance of the rest of the cable. In general, $C_P \gg C_D, C_S$.

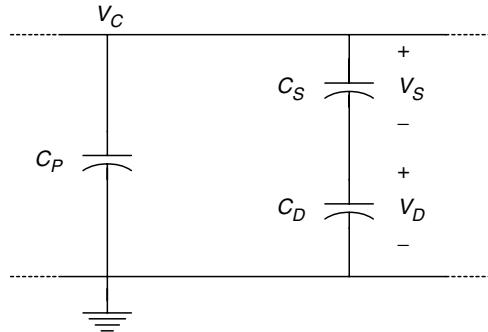
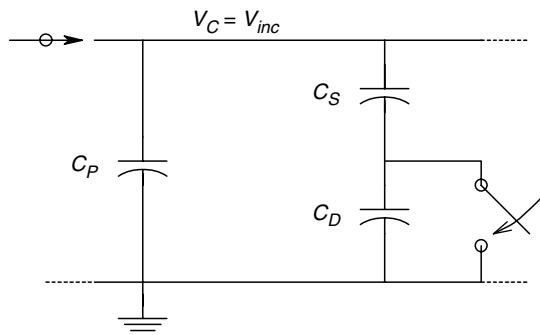


FIGURE 11.10

A simplified schematic illustrating the partial discharge process.



high voltage power cable behaves like a large capacitor, C_p . A typical 15 kV, coaxial power cable has a capacitance of about 60 pF/ft. In the frequency band from 20 kHz to about 50 MHz, the cable behaves like a lossy, RLC transmission line (Section 9.8), with a characteristic impedance, Z_o , of around 45 Ω .

When the water tree or other insulation defect arcs over at the ac inception voltage, the resulting PD injects what is effectively an impulse of charge into the transmission line equivalent circuit of the cable. The PD discharge transient is over in picoseconds, however, a voltage pulse propagates away from the site in both directions on the cable. The direct PD voltage transient recorded at the excitation/measurement (E/M) end of the cable ranges 1–10 mV peak amplitude and several hundred nanoseconds in duration. It is superimposed on the kilovolt range, 60 Hz cable excitation voltage, which presents a problem in its separation and measurement.

The direct PD transient first arrives at the excitation/measurement (E/M) end of the power cable at time T_0 . The transient which travels to the far end of the cable is reflected at its open circuit end and then travels back past the PD site to the E/M end (Section 9.8). The Thevenin impedance of the high voltage source driving the E/M end of the cable is higher than the cable's Z_o , so that all PD transients arriving at the E/M end are reflected to the open circuited, far end of the cable, where they are again reflected, etc. By examining the *reflection series* of a PD pulse at the E/M end of the cable and knowing the physical length of the cable, it is possible to estimate the location of the PD site. A typical reflection series from a single PD, recorded from a 15 kV URD, XLPE dielectric, coaxial power cable in the author's laboratory is shown in Figure 11.11. Note how successive pulses are attenuated and broadened by the cable's low-pass filtering characteristics. In the field, if a PD is judged to be severe, the section of cable containing the PD site is excised and a new cable section is spliced in, or the entire cable may be replaced with a new one.

The initial voltage transient at the PD site can be modeled, following the lumped parameter development of Mason (1965). Three capacitors are used in Mason's model. C_D is the equivalent capacitance of the insulation defect giving rise to the PD. C_S is the equivalent series capacitance between C_D and the cable's outer shield and center conductor. C_p is the total cable capacitance, less C_S and C_D in series. A PD occurs when the 60 Hz voltage on the cable reaches the inception voltage, V_{inc} . At this time, the dielectric of the insulation defect breaks down and the voltage across C_D quickly goes to zero as its charge is neutralized. This situation is illustrated in Figure 11.10. At $t = 0-$, just before the PD occurs, the charge stored in C_p is $Q_p(0-) = C_p V_{inc}$. It is easy to see that the

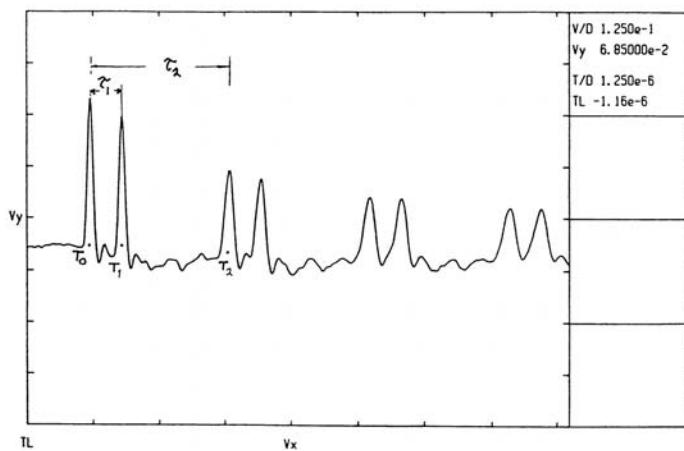


FIGURE 11.11

PD waveforms generated by a defect near the far end of a 730 ft long power cable. Amplitude of the first recorded PD pulse in the series = 375 mV. $\tau_1 = 600 \text{ ns}$, $\tau_2 = 2.640 \mu\text{s}$. Pulses have been conditioned by a voltage gain of 100 as well as high-pass filtering.

charges stored on C_S and C_D are equivalent, and equal to $Q_{SD} = V_{inc}C_S C_D / (C_S + C_D)$. Thus, the voltage across the defect cavity at breakdown is:

$$V_D = Q_{SD}/C_D = V_{inc} \left(\frac{C_S}{C_S + C_D} \right) \quad (11.19)$$

The voltage across C_S is:

$$V_S = Q_{SD}/C_S = V_{inc} \left(\frac{C_D}{C_S + C_D} \right) \quad (11.20)$$

Obviously, $V_{inc} = V_D + V_S$. The PD occurs at $t=0$. At $t=0+$, $V_D \rightarrow 0$, $Q_D \rightarrow 0$ and $V_S \rightarrow V_{inc}$. Now the total charge on C_P and C_S is:

$$Q_{PS} = C_P V_{inc} + Q_{SD} \text{ (on } C_S) \quad (11.21)$$

Assuming the transmission line's series inductance prevents instantaneous longitudinal charge flow, the voltage across the PD site at $t=0+$ is now:

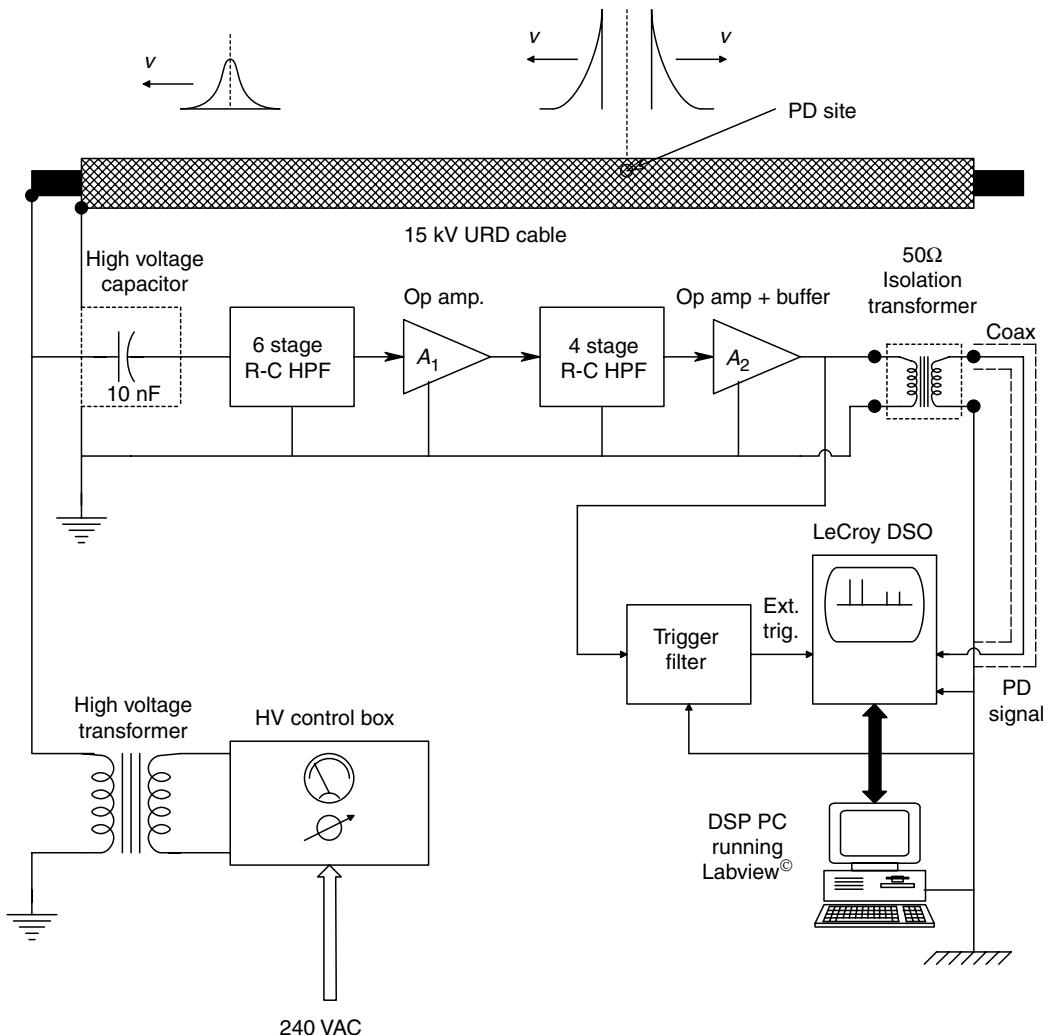
$$V' = \frac{Q_{PS}}{C_P + C_S} = \frac{V_{inc}[C_P + C_S C_D / (C_S + C_D)]}{C_P + C_S} \quad (11.22)$$

The initial voltage change is nearly instantaneous, with a rise time estimated to be in picoseconds. The voltage change at the site is approximately:

$$\Delta V = V' - V_{inc} = V_{inc} \frac{C_P + C_S C_D / (C_S + C_D)}{C_P + C_S} - 1 \cong \frac{-V_{inc} C_S^2}{C_P (C_S + C_D)} \quad (11.23)$$

Let us calculate a typical PD initial amplitude at the PD site. Let $C_P = 10 \text{ nF}$, $V_{inc} = 2,000 \text{ V}$, $C_S = 1 \text{ pF}$, $C_D = 3 \text{ pF}$. From equation 11.23 above, $\Delta V = -50 \text{ mV}$. The 60 Hz excitation on the power cable is over 92 dB larger than the initial PD amplitude. As the PD 'spike' propagates along the cable, it is attenuated and rounded due to the attenuation of high frequencies by the cable's transfer function. This attenuation and rounding are illustrated in the PDs shown in Figure 11.11.

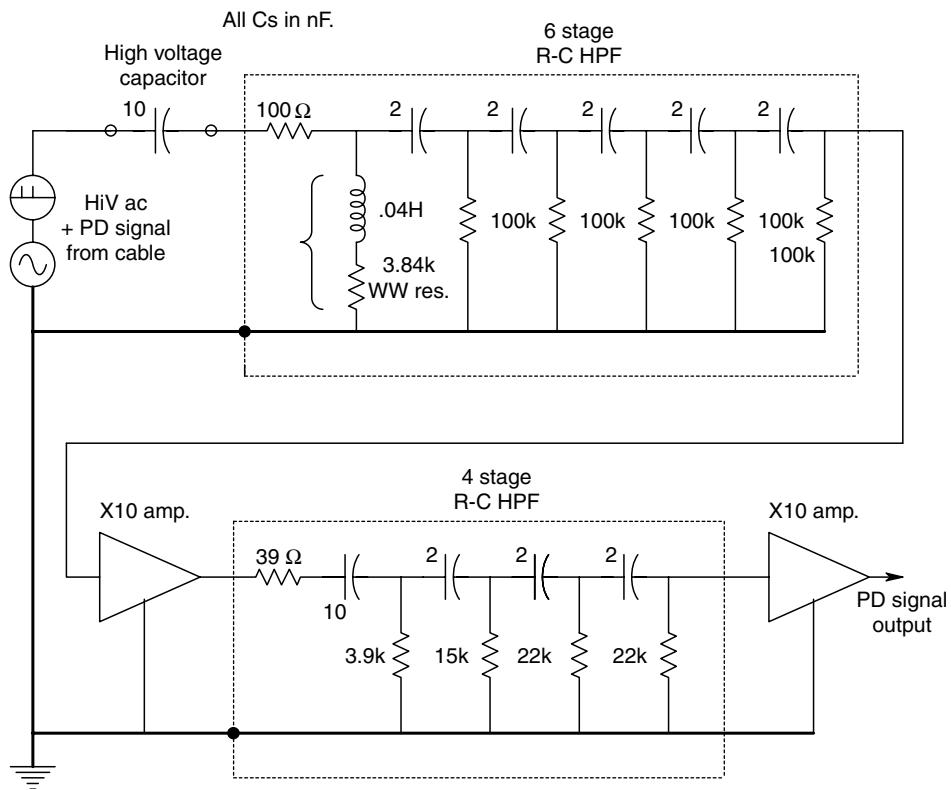
Measurements on the PD reflection series requires the separation of the PD transients from the 60 Hz excitation voltage. Figure 11.12 illustrates a simplified block diagram of the signal acquisition portion of the PD detection, location and characterization system described in the following US patents—Mashikian, Northrop, Bansal and Nikias, US Patent 4,887,041, 12 December 1989; Mashikian, Northrop, Bansal and Palmieri, US Patent 5,272,439, 21 December 1993; Mashikian, Northrop and Sui, US Patent 5,448,176, 5 September 1995. The cable is excited from a clean, variable voltage, 60 Hz HV source. There must be no corona or PDs in the HV power source which can be confused with, or mask, PD signals from the cable. A Tettex AG type 3380/100/70, 70 kV, 10 nF, compressed gas, standard high voltage capacitor is used to couple voltages on the cable under test to the first high-pass filter (HPF1) used to condition the PD signal. A low noise, op-amp, signal conditioning amplifier (A_1) with a gain of 10 is used to boost the PD signal at the output of HPF1. A_1 is followed by a second high-pass filter, HPF2, to further reduce the 60 Hz voltage. The output of HPF2 is conditioned by a second broadband amplifier (A_2) with a gain of 10. AD840 op-amps were used because they have low voltage noise, high f_T and slew rate, and low cost. A_2 drives a 50Ω , high frequency, isolation transformer

**FIGURE 11.12**

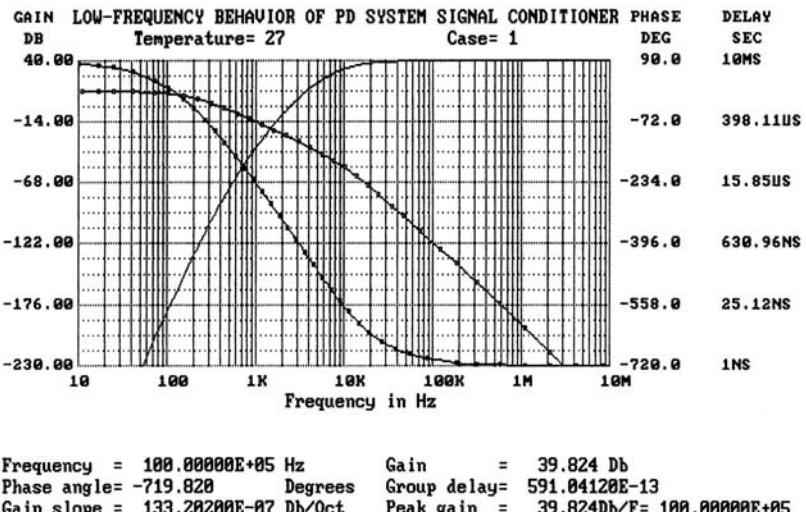
Block diagram of the partial discharge measurement system of Mashikian *et al* (1990). The high voltage capacitor is 10 nF. The op-amp gain stages each have a gain of 10 and -3 dB bandwidths of 5 MHz. The HPFs are shown in detail in Figure 11.13.

(North Hills Electronics, Inc. model 0016PA, 20 Hz–20 MHz, 50Ω , Z_{in} and Z_{out}), which is used to prevent ground loops and protect the DSO in the event of a massive cable arc over. A 50Ω coaxial cable carries the PD signals to the LeCroy digital storage oscilloscope (DSO), which serves to digitize and record the PD transients. Digital records from the DSO's memory or floppy disks are passed on to a PC used for further DSP operations on the signals. The schematics of the HPFs are shown in Figure 11.13. Figure 11.14 shows the frequency response of the two cascaded high-pass filters and amplifiers. The gain of the combined filters and amplifiers is -262 dB at 60 Hz. This means that the 60 Hz HV cable voltage is attenuated by a factor of 7.94×10^{-14} . Thus a 10 kV pk, 60 Hz input to the filter produces a 0.794 nVpk, 60 Hz interference at its output. The combined HPFs have a -3 dB frequency of 25.2 kHz and the passband gain is +40 dB, as expected.

The SNR of weak PDs is improved by signal averaging on the LeCroy DSO. Averaging is only possible given reliable, consistent triggering of the DSO sweep by the peak of

**FIGURE 11.13**

Schematic diagram of the signal conditioning high-pass filters. The op-amp circuits are represented simply by gain of 10 VCVSs.

**FIGURE 11.14**

Frequency response of the circuit of Figure 13.13. Plot made with MicroCap™.

the first (largest) PD transient recorded at the E/M end of the energized power cable. A special, narrow bandpass filter is used on the PD waveforms to more reliably extract the triggering pulses for signal averaging on the DSO. When the PD SNR is too low for reliable averaging on the DSO, we use a cross-correlation technique on the PC to establish a reference time $T = 0$ at the start of each noisy PD pulse group, so that averaging can take place reliably. The DSO sampling period is normally set at 20 ns/sample; 16,000, 8-bit samples are taken on a delayed input signal, so the full extent of the first (trigger) PD pulse can be seen.

After a suitable PD reflection series or averaged reflection series is acquired, it can be transferred to the PC on the IEEE-488 bus for further DSP operations, which include proprietary PD site location algorithms. One location algorithm deconvolves the broad, recorded pulse series with the cable transfer function in order to better estimate the true times of occurrence of the pulses in the reflection series. Another location algorithm uses a maximum likelihood (ML) method with cross-correlation [Knapp *et al*, 1990]. The ML location algorithm has been found to be robust in the presence of noise. PD site location is based on the *a priori* knowledge of the cable length. From the PD reflection series shown in Figure 11.11, we can find the PD propagation velocity on the cable, c_c , from the fact that that it takes τ_2 seconds for a pulse to travel from the E/M end of the cable to the far end and then return, a distance of $2L$ meters. Thus:

$$c_c = 2L/\tau_2 \text{ m/s} \quad (11.24)$$

The c_c on most XLPE high voltage cables is about 200 m/ μ s, or 0.667 times the speed of light. The true time of origin of the PD is initially unknown, so we work with the time differences between the first three pulses in the reflection series. Thus:

$$\tau_1 = T_1 - T_0 \quad (11.25)$$

and

$$\tau_2 = T_2 - T_0 \quad (11.26)$$

The time differences are typically measured using the maxima of the pulses corrected for broadening and attenuation by the cable's transfer function. Equation 11.7 can be used to write:

$$\tau_1 = (2L - x_d)/c_c - x_d/c_c \quad (11.27)$$

Equation 11.27 is solved for x_d , the distance from the E/M end of the cable to the site of PD origin. c_c is eliminated by substituting equation 11.24 into equation 11.27. The final result is:

$$x_d = L(1 - \tau_1/\tau_2) \quad (11.28)$$

Thus, we have only to know the total cable length, L , and estimate the time differences τ_1 and τ_2 in order to find an estimate for x_d .

In its present embodiment, the partial discharge measurement and locating system is able to locate artificial PD sites at known locations on a cable with an accuracy of 99.5% [Knapp *et al*, 1990]. When the PD site is very close to either end of the cable (<30 m from the ends), the PD pulse peaks merge, and estimation of the pulse origin times is not

ordinarily possible. By splicing an extra 30 m of defect free, identical cable to the far end of the cable under test, we obtain distinct separation of the PD pulses and improved location resolution for PD sites closer than about 30 m from the cable ends.

Another problem that affects the SNR of the system and hence, the resolution of x_d , is the pickup of coherent interference in the system's 20 kHz–10 MHz signal bandwidth. Such interference is from RF sources such as local AM broadcast stations. In certain cases, when the cable is above ground, such RFI can be of the same order of magnitude as the PD reflection series, confounding measurements. Our approach to this problem is to do a spectral analysis of the system noise to locate the frequencies of the coherent interference. We then implement one or more adaptive, digital, FIR notch (band reject) filters on the system computer to remove the interference, leaving the PD waveforms largely intact. Of course, we follow good grounding, shielding and guarding practices. As mentioned above, we also use a broadband isolation transformer to reduce common mode noise. However, RFI can be picked up by the cable under test in spite of these practices. System broadband random noise arises in the headstage amplifier and in the high-pass filters. We have used AD840 op-amps for broadband amplifiers. These op-amps have $f_T = 400$ MHz, $\eta = 400$ V/ μ s, and $e_{na} = 4$ nV/ $\sqrt{\text{Hz}}$. Their broadband noise is 10 μ V RMS in a 10 Hz–10 MHz bandwidth.

Our instrumentation allows the detection and reliable site estimation (x_d) of PD pulses as low as 80 μ V peak at the E/M end of the cable using a 5 MHz signal bandpass. Although the estimated PD site location is found in terms of % of the total, known cable length, the actual location of the site is complicated by the fact that the cable plunges into its trench and then snakes a certain amount along its path. In practice, we dig a hole to the cable at the expected site and then inject a transient test pulse into the unenergized cable by magnetic induction. A fast current pulse is passed through a coil with about 25 turns around the outside of the cable. This induces a PD-like transient, and a corresponding reflection series which can be compared to the stored PD series. A coincidence of pulse peaks in the two series means the test coil is very near the PD inception site. Accuracy with the current pulse 'vernier' method is limited by the DSO sampling rate. For repetitive signals, such as the current test pulses, the effective sampling period of the DSO is 2 ns/S. This interval corresponds to a distance of 0.4 m on the cable, giving a comparison location error of 0.2% on a 200 m cable. The results of ongoing development of DSP algorithms for pulse time difference estimation suggest that a routine location accuracy of 99.8% will be possible. This accuracy translates into a ± 0.5 m error in finding a PD site on a 250 m cable. The use of sophisticated DSP filtering techniques on noisy data is what has made the PD measurement instrument a successful, field portable system.

11.4 Design of a Closed Loop, Constant Phase, Pulsed Laser Ranging System and Velocimeter

Laser velocimeters are used by police to measure vehicle velocities on highways. They are also used to measure the relative velocity and range between two space vehicles during docking maneuvers. In the police application, the pulsed laser is generally directed from a stationary transmitter/receiver in line with vehicle travel, such as from a bridge overpass or on a curve, to a non-specular surface on the moving vehicle. A telescope collimated with the laser is used to collect the partially reflected signal and direct it to a fast, sensitive photodetector, such as an avalanche photodiode. There is a Doppler shift on the frequency of the return signal, but this is not what is measured. The time

delay between transmission and the detection of the return pulse is what is quantified. In 3.33 ns, light travels one meter. Unlike Doppler radar velocimeters, the LAser VElocimeter and RAnging (LAVERA) systems are highly directional, subject only to the transmitted laser beam's divergence. It is this specificity that reduces target ambiguity when one or more other vehicles are in close proximity to the desired target. In order to image the laser spot on the target vehicle, the vehicle surface should be a non-specular (unlike a mirror), diffuse reflector with a high albedo (brightness). A mirror-like finish at an angle to the transmitted beam will cause the reflected beam to be directed away from the receiver telescope, and not be sensed, generating a 'stealth' condition. Thus slanted windshields and chrome bumpers on automobiles make poor target surfaces. A black vinyl bumper is also a poor target because of its low albedo. Such poorly responding targets are called 'uncooperative'. A white painted surface is more cooperative, and a retroreflecting surface makes an ideal, 'cooperative' target, as nearly all of the laser's energy is directed back parallel to its beam.

Early LAVERA systems used by police generally operated in an open loop mode, the laser being pulsed at a constant rate (e.g. 381 pps in an early Laser Atlanta Optics, Inc. unit). The time between transmitting any one pulse and the detection of its return image from the moving vehicle is measured and stored in a digital memory. One way of measuring the return time, which is proportional to the target range, is to have the k th transmitted pulse start a fast Miller integrator circuit (similar to the horizontal sweep circuit in an analog oscilloscope), generating a ramp of voltage. A fast, track-and-hold (T&H) circuit follows the ramp until the k th return pulse triggers the T&H circuit to hold the ramp voltage at that instant. This ramp voltage is then converted to digital form by an ADC, whose output is stored in digital memory. The previous, $(k - 1)$ th digital word, proportional to the range at time t_{k-1} is subtracted from the present, k th digital word, which is proportional to the range at time t_k . The difference, appropriately scaled, is proportional to vehicle velocity. In practice, an FIR smoothing routine operates on the sequence of differences to reduce noise. This approach has two serious drawbacks:

1. It does not provide range information
2. Its bandwidth and hence, its ability to respond to sudden changes in vehicle velocity, is limited by the low, constant pulse rate of the laser source.

Another open loop laser LAVERA system, described by Koskinen *et al* (1992), used a higher, fixed, pulse rate (1–50 MHz) and an ECL Time to Amplitude Converter (TAC), which is in effect, a precise phase detector.

The components of a typical LAVERA system consist of a pulsed, laser diode source, usually about $1.5\text{ }\mu\text{m}$ wavelength, collimating optics, a target (moving or stationary, cooperative or uncooperative), a receiving telescope, a high speed photosensor, an electronic signal conditioning amplifier, a time-to-voltage converter to estimate target range, signal processing circuitry (digital and/or analog) to give speed and a readout display. In the description given below of the closed loop, constant phase LAVERA system, we describe the systems architecture of the range and velocity estimating circuits and do not consider the details of the pulse circuits, the laser, optics, the sensor or preamplifier.

Figure 11.15 shows a functional block diagram of the closed loop LAVERA system developed by the author. Note that this system will work either with a pulsed or sinusoidal light intensity output. The loop integrator has a compensatory zero associated with it, giving the closed loop system a circle root locus. The zero is at $s = -1/\tau_c$ in the s plane.

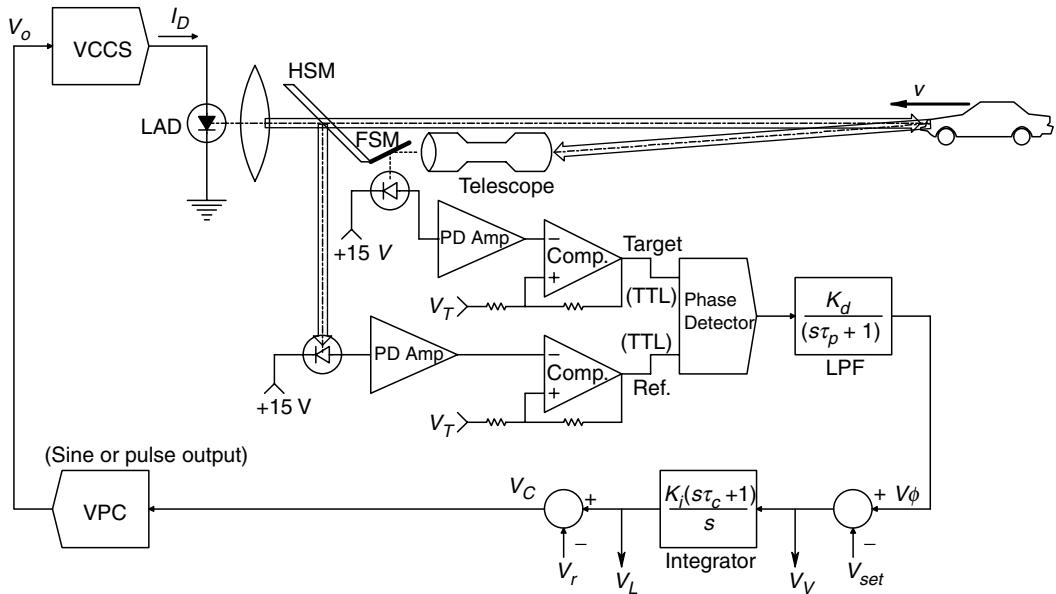


FIGURE 11.15

Block diagram of the closed loop, constant phase, pulsed laser velocity and ranging (LAVERA) system devised by the author. The heart of the system is the voltage-to-period converter controlled oscillator (VPC). V_V is the system output voltage proportional to target velocity, V_L is the output voltage proportional to target range, L .

There are several possible digital phase detector (PD) circuits that can be used. One PD circuit which we used was a digital phase detector based on the Motorola 4044 IC. It compares the phase between the k th transmitted pulse and its 'echo' received $\tau_k = 2L_k/c_s$ later, where L_k is the distance to the target (assumed constant over τ_k) and c is the velocity of light in air. If the frequency of the pulses, f_c , is constant, the phase lag between received and transmitted pulses can be expressed as:

$$\phi = 360 \frac{2L}{cT_c} = (2.40 \times 10^{-6}) L f_c \text{ degrees} \quad (L \text{ in m}, f_c \text{ in Hz}) \quad (11.29)$$

The average voltage at the output of the low-pass filter conditioning the output of the digital phase detector is $V_\phi = K_d \phi$. V_ϕ is next conditioned by an integrator with phase lead compensation, $K_i(s\tau_c + 1)/s$. The output of the integrator, V_L , has a dc level, V_r , subtracted from it to form V_c , the input to the voltage-to-period Converter (VPC). Do not confuse the voltage to period converter with the well-known VCO or voltage controlled oscillator IC. The frequency output of a VCO behaves according to the rule $f_c = K_V V_C + c$. In the VPC, it is the *period* of the oscillator output which is directly proportional to V_c . Thus, for the VPC, we can write:

$$T_c = 1/f_c = K_P V_C + d \quad (11.30)$$

The VPC output, V_o , can be sinusoidal or pulsed, depending on the application. V_o drives a VCCS which provides current to drive the laser diode (LAD). The LAD optical output provides a reference phase input to the PD IC. The entire system operates as a closed loop, constant phase, Type I control system. The period of the VPC output is

automatically adjusted as L changes to maintain a constant ϕ and V_ϕ . Since the closed loop system has a pole at the origin of the s plane, it has zero steady state phase error for constant L (stationary targets) and a small steady state phase error for targets moving at a constant velocity ($dL/dt = v_o$). We can use this general property of a Type I control system to examine the steady state relationships governing the output voltage proportional to velocity, V_V , and the output voltage proportional to range, V_L . If the target is at rest at range L_o , then V_V will be zero, thus $V_\phi = K_d \phi = V_{SET}$. After some algebra, the steady state V_{Lo} is found to be:

$$V_{Lo} = \frac{K_d 2.40 \times 10^{-6} L_o}{K_V V_{SET}} + (V_r - d/K_V) = S_L L_o + \lambda \quad (11.31)$$

where λ is a constant. Let the desired, steady state phase difference in the system be ϕ_o . If we substitute $V_{SET}/K_d = \phi_o$ into equation 11.31 above, we obtain:

$$V_L = \frac{2.40 \times 10^{-6} L_o}{K_P \phi_o} + (V_r - d/K_V) \quad (11.32)$$

Now, let the target be moving at a constant velocity, $dL/dt = v_o$. Now V_V is non-zero and may be shown to be:

$$V_V \cong \frac{dV_L/dt}{K_i} = \frac{K_d 2.40 \times 10^{-6} v_o}{K_i K_V V_{SET}} = S_V v_o \quad (11.33)$$

We now consider the system's dynamics. Clearly, the system is nonlinear, because the input, L , appears as a multiplicative input. Figure 11.16A illustrates the nonlinear systems block diagram of the LAVERA system. In order to *linearize* the system to examine its loop gain, we consider the nonlinear transfer function of the VPC as a linear gain with a dc component, of the form, $f_c = S_V V_C + e$ where S_V and e are determined at the system's operating point. By differentiation:

$$S_V = df_c/dV_C = -K_P/(K_P V_C + d)^2 = -K_P(f_c)^2 = -K_P[\phi_o/(2.40 \times 10^{-6} L_o)]^2 \quad (11.34)$$

The linearized system is shown in Figure 11.16B. Its loop gain is second order, and is given by:

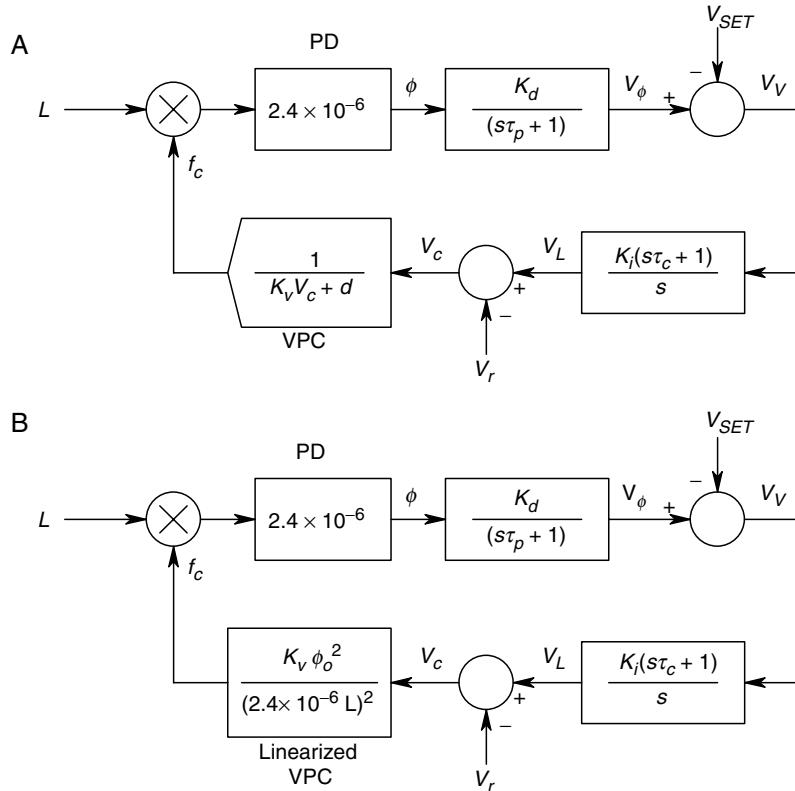
$$A_L(s) = \frac{K_v \phi_o^2 K_d K_i (s\tau_c + 1)}{2.40 \times 10^{-6} L_o s (s\tau_p + 1)} \quad (11.35)$$

It can easily be shown that the closed loop, nonlinear system's natural frequency² is:

$$\omega_n^2 = K_v \phi_o^2 K_d K_i / (\tau_p L_o 2.40 \times 10^{-6}) \quad (\text{r/s})^2 \quad (11.36)$$

The system's damping factor for complex conjugate poles is also nonlinear, and is given by:

$$\zeta = (1/2) \frac{[2.40 \times 10^{-6} L_o + K_v \phi_o^2 K_d K_i \tau_c]}{\sqrt{(K_v \phi_o^2 K_d K_i 2.4 \times 10^{-6} L_o \tau_p)}} \quad (11.37)$$

**FIGURE 11.16**

A. Block diagram of the nonlinear LAVERA system illustrating the VPC and the Lf_c multiplier. B. LAVERA block diagram in which the VPC has been linearized. Its loop gain is given by equation 11.35.

As an example, let the following parameters be used. $L_o = 100 \text{ m}$, $\tau_p = 0.01 \text{ s}$, $K_i = 500$, $\phi_o = 25.0$, $\tau_c = 0.0001 \text{ s}$, $K_d = 0.01$, $K_V = 4.75E-6$. We calculate $S_V = 2.4E-6/(K_i K_P \phi_o) = 4.042E-5 \text{ sV/m/s}$. The range sensitivity is $S_L = 2.4E-6/(K_P \phi_o) = 2.0211E-2 \text{ V/m}$. The undamped natural frequency is:

$$\omega_n = \sqrt{K_v \phi_o^2 K_d K_i / (2.4E - 6 \tau_p L_o)} = 78.64 \text{ r/s}, \quad \text{or } f_n = 12.5 \text{ Hz} \quad (11.38)$$

ω_n is seen to increase as range, L , decreases. The natural frequency calculated above can easily be increased at 100 m by increasing K_i , K_d , etc. and by decreasing τ_p . The closed loop system damping factor at a range of 100 m is $\zeta = 0.64$ (underdamped). If L varies in the range 10–300 m, the SS VPC frequency varies in the range 1.042 MHz–34.7 kHz.

An experimental investigation of the LAVERA system described above was undertaken in the author's laboratory [Nelson, 1999]. A circuit design for a simple VPC is shown in Figure 11.17. A CW, sinusoidal amplitude modulated light beam was used because the amplifiers used to condition the photonic signals required much lower bandwidths as compared to a pulsed output, hence cost was lower. The VPC oscillator was made from an analog reciprocal circuit [Northrop, 2004] that drives a Maxim MAX038 linear VFC VCO. The VPC sinusoidal output drives a 3 op-amp VCCS circuit, which in turn provides current to the two NIR (675 nm), 5 mW, laser diodes. One is the transmitter; the other is

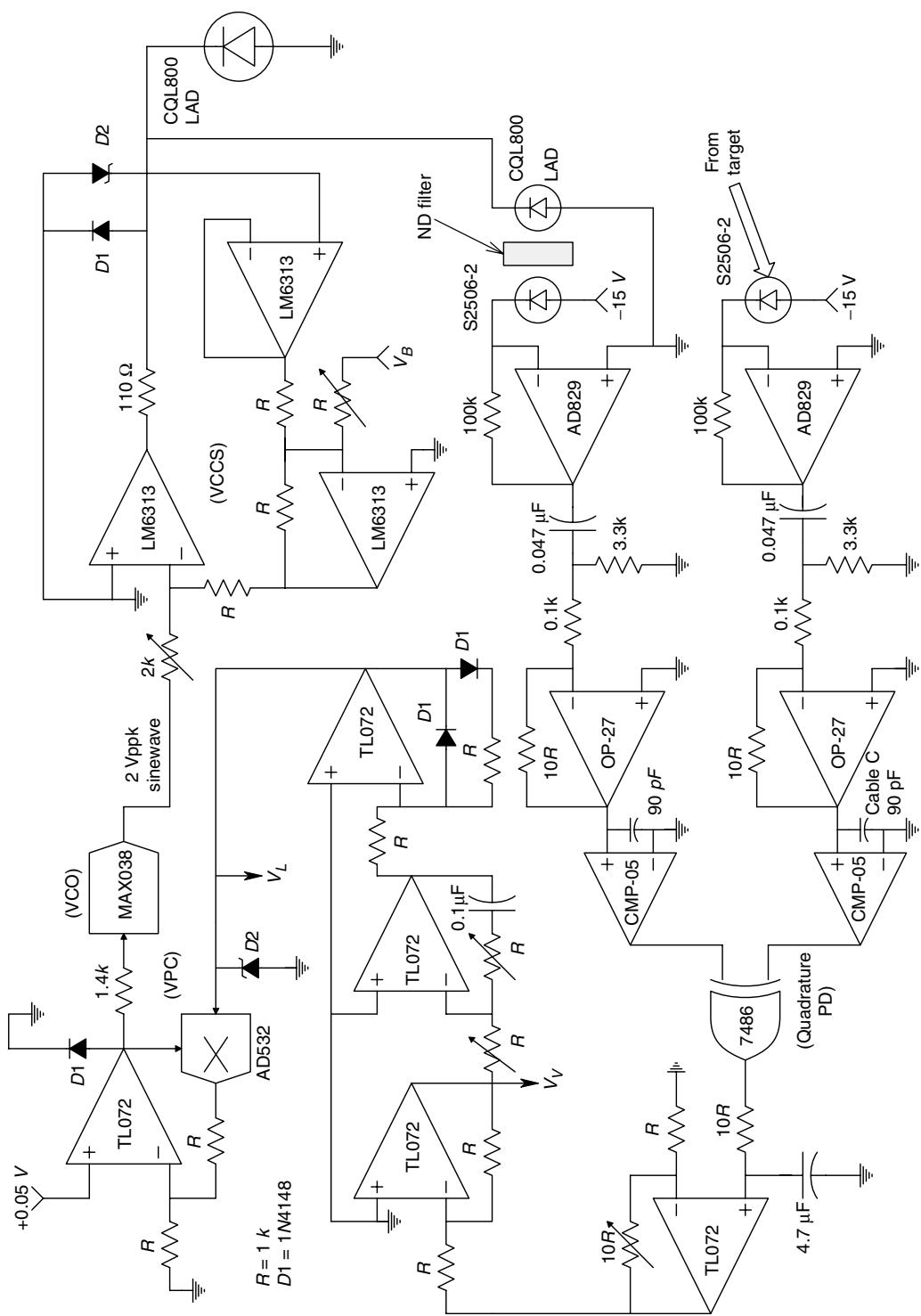


FIGURE 11.17
Simplified schematic of Nelson's 1999 LAVERA system.

used as a phase reference source with a neutral density (ND) filter to prevent saturating the reference photodiode. The sinusoidal signals from the LADs are conditioned, then changed to TTL square waves by a pair of fast comparators. The phase detector in this case is a simple EOR gate followed by a dc shifter and LPF. After integration and rectification to prevent V_L from going <0 , and zener clipping to limit the positive input, V_L drives the VPC, etc. Nelson's system was developed as a prototype navigational aid for the blind. A two channel acoustic output was attached to V_V and V_L . In one embodiment, the frequency of a CW audio tone was made inversely proportional to range, L , and object velocity toward the subject elicited a second signal of clicks whose rate was proportional to v . Object velocity away from the subject produced no clicks; the decreasing frequency of the CW tone signaled negative v .

Note that the architecture of the LAVERA system can be extended to ultrasound to simultaneously measure velocity and distance of objects moving in air or water, or even in the body (such as aneurisms, or heart motion). Unlike Doppler systems, a LAVERA system gives object displacement, as well as velocity. At long ranges, pulsed ultrasound bursts can be used, while for short range applications, CW sound is indicated.

11.5 Design of Capacitive Sensors for the Detection of Hidden Objects

11.5.1 Introduction

Historically, there have been many approaches to sensing small changes in capacitance (ΔC). Some of these techniques were developed to measure small changes in dielectric constant, mechanical displacement (of a diaphragm or cantilever) associated with the measurement of pressure or force, capacitance caused by the presence of buried objects having dielectric constants different from soil (this might include land mines, or ceramic pipes buried in soil or concrete), the presence of humans in places they should not be (including burglars, workers near dangerous machinery such as robot arms and kids near school buses), liquid level changes, water content in foods, etc. A human body is mostly water, thus it has a high dielectric constant, which makes it easy to sense by ΔC methods.

One of the earliest means devised for sensing ΔC is the DeSauty bridge (Section 5.5.1.4 and Figure 11.18). The bridge is generally operated around null. The output of a DeSauty

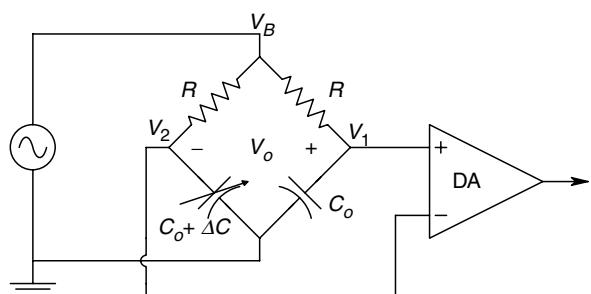


FIGURE 11.18
A De Sauty bridge used to measure ΔC .

bridge, given a small capacitance unbalance, can be written as:

$$\begin{aligned} \mathbf{V}_o &= \mathbf{V}_1 - \mathbf{V}_2 = \mathbf{V}_b \left\{ \frac{1/j\omega C_o}{R + 1/j\omega C_o} - \frac{1/j\omega(C_o + \Delta C)}{[1/j\omega(C_o + \Delta C) + R]} \right\} \\ &\downarrow \\ \mathbf{V}_o &\cong \mathbf{V}_b \left\{ \frac{j\omega RC_o(\Delta C/C_o)}{[j\omega RC_o + 1]^2} \right\} \end{aligned} \quad (11.39)$$

For $\omega = 1/RC_o$ r/s, the bridge has maximum sensitivity. The output can then be shown to be:

$$\mathbf{V}_o = \mathbf{V}_b 1/2(\Delta C/C_o) \quad (11.40)$$

As we stated in Chapter 5, stray wiring capacitances from the bridge corners to ground and the input capacitance of the differential amplifier make the output conditions more complex than the simple output equation, 11.40. Once a DeSautey bridge is taken off the lab bench, its symmetry is hard to preserve, and it requires compensation for stray capacitances to make it sensitive to very small ($\Delta C/C_o$).

Other means of sensing small changes in capacitance have relied on various servo nulling techniques for capacitance bridges. Most have been based on the simple op-amp RC_o circuit shown in Figure 11.19. The ac voltage at the op-amp output is examined to be:

$$\mathbf{V}_o = I_C R + \mathbf{V}_s = \mathbf{V}_s[j\omega CR + 1] \quad (11.41)$$

It is instructive to examine both the *phase* and *magnitude sensitivity* of \mathbf{V}_o . The phase of \mathbf{V}_o is:

$$\phi = \tan^{-1}(\omega RC) \quad (11.42)$$

The phase sensitivity of \mathbf{V}_o with respect to C is:

$$\frac{\phi}{C} = \frac{d\phi}{dC} = \frac{\omega R}{1 + (\omega RC)^2} = \frac{\omega RC}{C[1 + (\omega RC)^2]} = 0.5/C, \quad \text{for } \omega = 1/RC \text{ r/s} \quad (11.43)$$

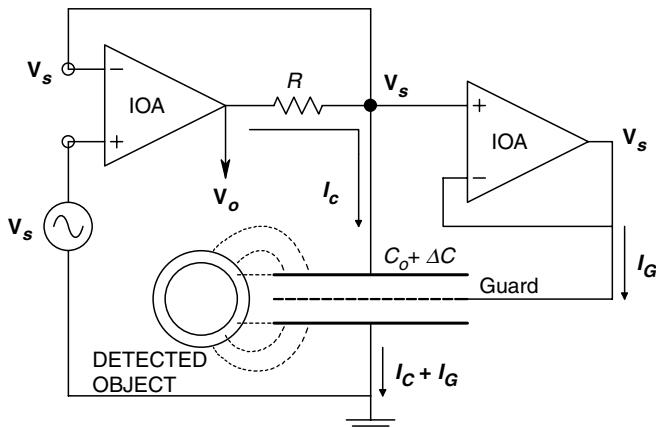


FIGURE 11.19

Active circuit which can measure changes in capacitance as changes of voltage or output phase.

The magnitude of \mathbf{V}_o is:

$$M = |\mathbf{V}_o/\mathbf{V}_s| = \sqrt{[1 + (\omega RC)^2]} \quad (11.44)$$

The magnitude sensitivity with respect to C is:

$$\frac{\frac{M}{S}}{C} = \frac{dM}{dC} = \frac{\omega^2 C^2 R^2}{C\sqrt{[1 + (\omega RC)^2]}} = \frac{1}{C\sqrt{2}} = 0.7071/C, \quad \text{for } \omega = 1/RC \quad (11.45)$$

Phase can be measured under controlled, low noise conditions to millidegrees. Thus:

$$\Delta\phi = \frac{\phi}{C} \Delta C = 0.001^\circ = (0.5/C_o) \Delta C \quad (11.46)$$

Assume $C_o = 100 \text{ pF}$, then a ΔC of *ca.* 0.1 pF could be resolved by direct phase measurement; this is poor resolution. However, it is easier to measure voltage with accuracy than it is phase. For example, if we can resolve $1 \mu\text{V}$ and $V_s = 10 \text{ V}$,

$$V_o = \Delta M |V_s| = \frac{M}{C} |V_s| \Delta C = 10^{-6} \text{ V} = (0.7071/C_o) \times 10 \Delta C \quad (11.47)$$

Thus, the resolvable $\Delta C \cong 1.4 \times 10^{-17} \text{ F}$, or 14 aF . This is a useful range of resolution, which in fact, can be increased by several orders of magnitude by the use of lock-in amplification. In practice, this resolution will be several orders of magnitude less because of system noise and drift.

11.5.2 Self-balancing Circuits Used to Measure $\Delta C/C_o$

Figure 11.20 illustrates a self-nulling DeSauty bridge in which a small change in C_2 unbalances the bridge. We can write $C_2 = C_o + \Delta C$. The unbalance is sensed by a phase sensitive rectifier (PSR). The dc error voltage, V_e , is integrated and added to a dc level, $-V_B$, which maintains static bridge balance. $V_C = V_B - V_i$ is the input to a VCCS with transconductance $G_M \text{ S}$. The VCCS output current, I_{LED} , causes an LED to emit light on a photoconductor, causing its light controlled conductance, G_p , to increase. In general, $G_{PC} = G_p + G_D$. G_D is the photoconductor's fixed dark conductance. Thus, the photoconductor's conductance is changed by V_i , so the bridge rebalances to compensate for the ΔC . The integrator in the feedback loop insures that $V_e \rightarrow 0$ in the steady state, following a step change in ΔC . This means that $V_o = 0$.

$$V_o = \frac{1/(G_p + G_D)}{1/(G_p + G_D) + R_4} - \left(\frac{C_1}{C_1 + C_2} \right) = 0 \quad (11.48)$$

From equation above, we see that null occurs for:

$$C_2 = C_1 R_4 G_{PC} \quad (11.49)$$

or when:

$$C_o + \Delta C = C_1 R_4 (G_p + G_D) \quad (11.50)$$

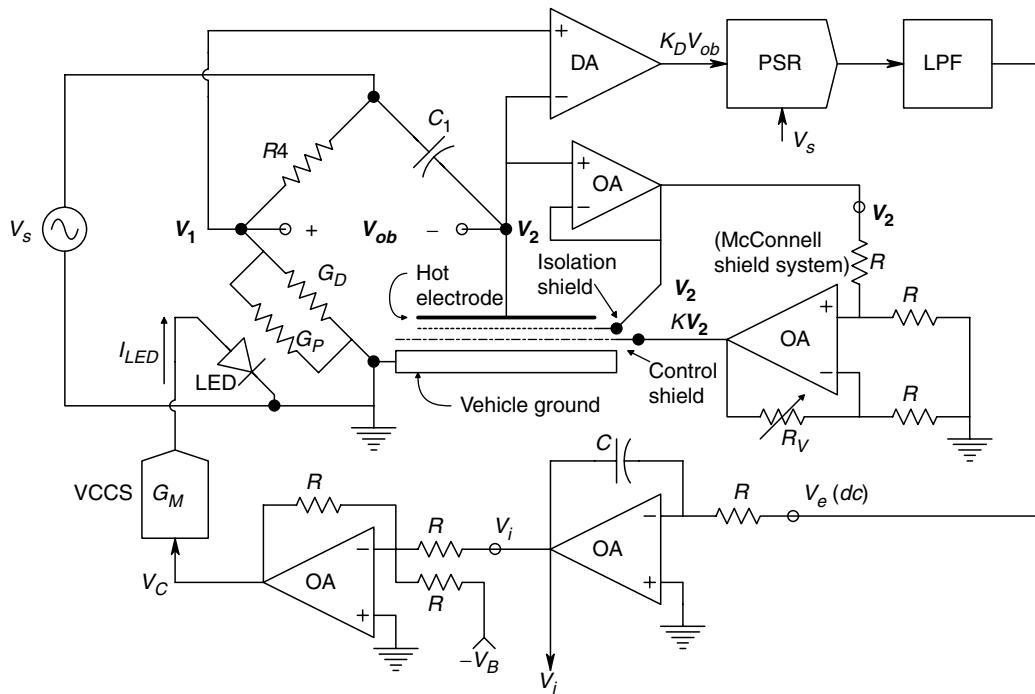


FIGURE 11.20

A self-nulling De Sauty bridge. V_i is proportional to ΔC .

The relation between the photoconductance, G_P , and V_C can be described by:

$$I_{LED} = G_M V_C \quad (11.51)$$

$$P_i = K_{LED} G_M V_C \text{ (optical power, W)} \quad (11.52)$$

$$G_P = K_{PC}P_i = K_{PC}K_{LED}G_MV_C = K_PV_C = K_P(V_B - V_i) \quad (11.53)$$

Substituting equation 11.53 into equation 11.50, we find:

$$C_\theta + \Delta C = C_1 R_4 [K_P(V_B - V_i) + G_D] \quad (11.54)$$

Initial balance is accomplished by setting V_B :

$$V_B = [C_0 - C_1 R_4 G_D] / (C_1 R_4 K_P) \quad (11.55)$$

(Note that V_B can be zero if $C_o = C_1 R_4 G_D$.) Incremental changes in C_2 are nulled by the integrated error output, V_i .

$$V_i = -\Delta C / (C_1 R_4 K_p) \quad (11.56)$$

Note that V_i can have either sign, depending on ΔC . Self-balancing bridges can be quite sensitive, but as we remarked above, unsymmetrical stray capacitances to ground from

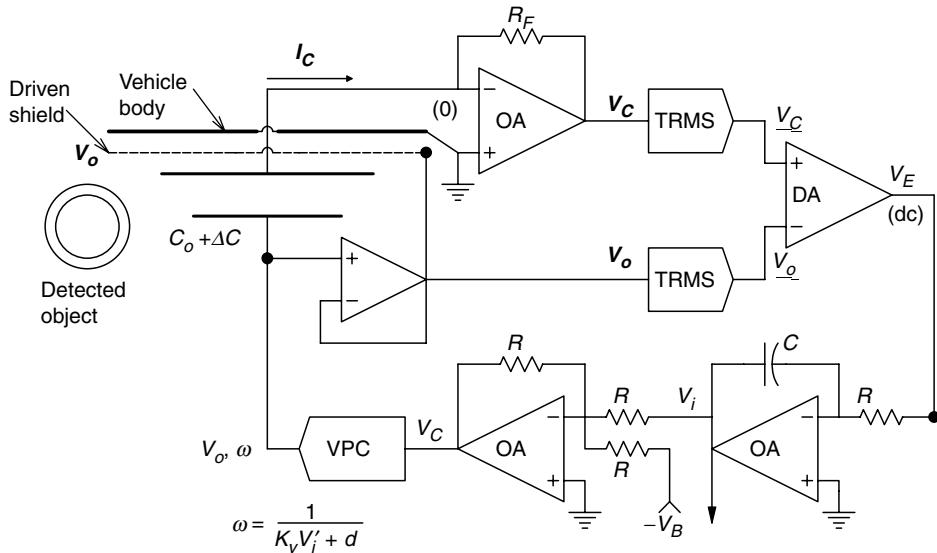


FIGURE 11.21

A self-nulling ΔC measurement system devised by the author. V_i is proportional to ΔC .

the bridge elements can prevent a perfect null of V_o in phase with V_s . Stray capacitances will give V_o a quadrature component.

Another approach to self-nulling detection of ΔC is shown in Figure 11.21. This system, designed by the author, is a parametric control system [Northrop, 2000]. It achieves an automatic null by keeping the ac current through the capacitor constant by varying the frequency of the ac voltage source driving the capacitor. A voltage-to-period converter (VPC) variable frequency oscillator (VFO) is used instead of the customary voltage controlled oscillator in which output frequency is proportional to the dc driving voltage, V_C . In a VPC, the output period is proportional to V_C . That is, $\tau = K_p V_C + d$. The VPC is used to linearize the system's output, V_i . The integrator in the feedback loop forces $V_e \rightarrow 0$ in the steady state. Thus, the dc outputs of the matched true RMS converter chips, V_c and V_o , are equal. From inspection of the figure:

$$I_c = V_o j \omega (C_o + \Delta C) \quad (11.57)$$

$$V_c = V_o j \omega (C_o + \Delta C) R_F \quad (11.58)$$

$$V_c = \frac{V_o}{\sqrt{2}} \omega (C_o + \Delta C) R_F \quad (11.59)$$

$$\underline{V_o} = \frac{V_o}{\sqrt{2}} \quad (11.60)$$

Now at null, in the steady state:

$$\underline{V_c} = \frac{V_o}{\sqrt{2}} \omega R_F C_o (1 + \Delta C / C_o) = \underline{V_o} = \frac{V_o}{\sqrt{2}} \quad (11.61)$$

$$\downarrow$$

$$\omega R_F C_o (1 + \Delta C / C_o) = 1 \quad (11.62)$$

$$\downarrow$$

$$R_F C_o (1 + \Delta C / C_o) = \omega^{-1} = K_p (V_B - V_i) + d \quad (11.63)$$

For steady state null with $\Delta C \rightarrow 0$, V_B is set so that:

$$V_B = C_o R_F / K_P - d / K_P \quad (11.64)$$

Again, the integrator output, V_i , nulls the ΔC component. Hence:

$$V_i = -\frac{R_F C_o}{K_P} \frac{\Delta C}{C_o} \quad (11.65)$$

Using 'typical' numbers, we can find the VPC system's sensitivity. Assume a $V_i = 1 \mu V$ can be detected reliably, also $C_o = 100 \text{ pF}$, $R_F = 10^7 \Omega$ and $K_P = 2 \times 10^{-5} \text{ s/V}$. Thus:

$$10^{-6} = \frac{10^7 \times 10^{-10}}{2 \times 10^{-5}} \frac{\Delta C}{C_o} \quad (11.66)$$

$$\downarrow$$

$$\frac{\Delta C}{C_o} = 2 \times 10^{-8} \quad (11.67)$$

Since the frequency changes with ΔC , it is instructive to examine the *frequency sensitivity*, $S_{\Delta C}^\omega$ with Δ capacitance. Note that:

$$\omega = \frac{1}{K_P V_C + d} \quad (11.68)$$

At null,

$$\omega = \frac{1}{K_P[(C_o R_F / K_P - d / K_P) - R_F \Delta C / K_P] + d} = \frac{1}{R_F(C_o - \Delta C)} \text{ r/s} \quad (11.69)$$

Now,

$$\frac{\frac{\omega}{\Delta C}}{\Delta C} = \frac{d\omega}{d\Delta C} = \frac{-1}{R_F(C_o - \Delta C)^2} \quad (11.70)$$

The fractional change in frequency, $\Delta\omega$, is simply:

$$\Delta\omega = \frac{\omega}{\Delta C} \Delta C \quad (11.71)$$

Let the detectable $\Delta\omega = 10^{-2} \text{ r/s}$ out of $\omega_0 = 10^4 \text{ r/s}$. Again, let $R_F = 10^7 \Omega$ and $C_o = 10^{-10} \text{ F}$. From equation 11.71, we find the detectable $\Delta C = 10^{-15} \text{ F}$, or 1 fF. This amounts to 10 ppm of C_o . Thus the frequency sensitivity is not as great as the V_i sensitivity (0.02 ppm).

11.5.3 Summary

As remarked in the introduction to this section, capacitance sensors have found diverse applications for a number of years. In 1981, Stahovec was issued US Patent 4,300,116 for *Safety Method and Apparatus for Sensing the Presence of Individuals Adjacent [to] a Vehicle*.

Stahovec's system was designed for school bus safety. It was based on the detection of the unbalance voltage of a variant of the DeSauty bridge. His sensing capacitors employed a *driven guard shield* to minimize C_o , while maximizing ΔC . Before stopping the bus, the bridge was servo balanced. It was then left open loop to detect a child near its capacitor electrodes while the bus was halted. Balance was effected using a photoconductor and LED, similar to our system shown in Figure 11.20.

NASA has described several capacitive proximity sensors for use in space and industrial applications. One such system was described by Vranish (1995, 1992a, 1992b) in a Technical Support Package from NASA Tech Briefs. Vranish's system uses a *driven capacitor shield* and senses ΔC by the changes the object causes in the frequency of an oscillator connected to the capacitor sensor about a *ca.* 20 kHz carrier. Apparently, no servo or autobalancing strategy is used in the Vranish system. McConnell (1993) also described a capacitive proximity sensor electrode system with improved object sensitivity that uses *two* driven shields, one with unity gain and the other with a gain that can be set over a range of 0.5–2.0. An embodiment of McConnell's shielding system is shown in Figure 11.20. In McConnell's TSP, he shows an improved sensitivity for a gain of 1.3. The optimum gain will, of course, depend on the actual electrode and shield geometry used.

Use of a capacitive sensor system to sense land mines turns out to have many problems. The work in the author's laboratory has shown that the air gap between the sensing capacitor and the ground must be constant as the sensor is scanned over the ground surface (i.e. the ground must be perfectly flat), the soil must be homogeneous (no large rocks) and the mine must have a significantly different dielectric constant than the soil around it.

11.6 Chapter Summary

In this chapter we described the designs of four novel measurement systems: The first is a novel microdegree polarimeter used to measure the concentration of optically active D-glucose used in bioreactors to grow cells. This system uses a self-nulling design in which the water solvent in the bioreactor's growth medium is used as a Faraday medium to linearly modulate the linearly polarized light used in the instrument. The use of the water solvent for modulation and feedback is novel, because normally, an expensive glass core Faraday rotator would be used.

The second system was designed to measure and locate partial discharges caused by insulation faults on coaxial, high voltage, power cables. These measurements are important to power utilities companies because they indicate the sites where catastrophic insulation failures can and will occur, and where new cable sections should be spliced in order to avoid cable failures. The third example of instrument system design was a closed loop, constant phase, pulsed laser velocity and range measuring (LAVERA) system. It has novel design in its use of a voltage-to-period converter controlled oscillator. The same system architecture can be extended to ultrasonic ranging systems, using either pulsed or CW ultrasound. The LAVERA system provides simultaneous target range and velocity output information.

The fourth example illustrated the use of feedback to automatically null systems used to measure very small changes in object capacitance ($\Delta C/C_o$). In one example, a self balancing bridge run at constant frequency was compared to a novel feedback system in which the current through the capacitive sensor was kept constant by varying the operating frequency by a voltage-to-period converter.

The accuracy of all four systems was seen to be limited by noise. In all cases, low noise amplification must be used. Further improvement of system performance in noise was achieved in the partial discharge measurement system by implementing digital notch filters to remove coherent interference, using cross-correlation to align noisy partial discharge waveforms for ensemble averaging, and other DSP algorithms for PD site location. The effects of noise in the LAVERA system are minimized by using a high power laser, low noise preamplifiers and op-amps, and fast logic.

References and Bibliography

1. Akar, C. and A.M. Shkel, Structural design and experimental characterization of torsional micromachined gyroscopes with non-resonant drive mode, *J. Micromech. Microeng.*, 14, 15–25, 2004.
2. Albert, K.J. *et al*, Cross-reactive chemical sensor arrays, *Chem. Rev.*, 100, 2595–2626, 2000.
3. Albone, E., Mammalian semiochemistry: chemical signaling between mammals, Chapter 21 in *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed., CRC Press, Boca Raton, FL, 1997.
4. Allen, P.E. and D.R. Holberg, *CMOS Analog Circuit Design*, Oxford University Press, NY, 2002.
5. Analog Devices, *Design-In Reference Manual*, Analog Devices, One Technology Way, Norwood, MA 02062-9106, 1994.
6. Anderson, K.F., Constant Current Impedance Measuring System That Is Immune to the Effects of Parasitic Impedances, U.S. Patent No. 5,731,469, December, 1994.
7. Anderson, K.F., *A Conversion of Wheatstone Bridge to Current-Loop Signal Conditioning for Strain Gauges*, NASA Technical Memorandum 104309, Dryden Flight Research Center, Edwards, CA 93523, 1995.
8. Anderson, K.F., NASA's Anderson loop, *IEEE Instrumentation and Measurements Magazine*, 1(1), March, 1998.
9. Angelo, E.J. Jr., *Electronics: BJTs, FETs and Microcircuits*, McGraw-Hill Book Co., NY, 1969.
10. ARRL Staff, *The Radio Amateur's Handbook*, American Radio Relay League, W. Hartford, CT, 1953.
11. Aseltine, J.A., *Transform Method in Linear System Analysis*, McGraw-Hill Book Co., NY, 1958.
12. Balanis, C.A., *Advanced Engineering Electromagnetics*, John Wiley & Sons, NY, 1989.
13. Barnes, J.R., *Electronic System Design, Interference and Noise Control Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
14. Barnes Engineering, *Handbook of Infrared Radiation Measurements*, Barnes Engineering Co., Stamford, CT, 1983.
15. Barney, G.C., *Intelligent Instrumentation, Microprocessor Applications in Measurement and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
16. Bartz, M., What is dithering? *Hewlett-Packard Journal*, December, 1993.
17. Barwicz, A. and W.J. Bock, An electronic high-pressure measuring system using a polarimetric fiber optic sensor, *IEEE Trans. Instrum. Meas.*, 39(6): 976–981, 1990.
18. Beckwith, T.G. and N.L. Buck, *Mechanical Measurements*, Addison-Wesley, Reading, MA, 1961.
19. Berger, A.J., T.-W. Koo, I. Itzkan, G. Horowitz and M.S. Feld, Multicomponent blood analysis by near-infrared Raman spectroscopy, *Appl. Optics*, 38(13), 2916–2926, 1999.
20. Berrie, P.G., Fuzzy logic in the evaluation of sensor data, Chapter 20 in *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed., CRC Press, Boca Raton, FL, 1997.
21. Betzler, K., *Fabry-Perot Interferometer*, At. URL: www.physik.uni-osnabrueck.de/kbetzler/sos/fabryperot.pdf, 2002.
22. Blanchard, A., *Phase-Locked Loops*, John Wiley & Sons, NY, 1976.
23. Bock, W.J., T.R. Wolinsi and A. Barwicz, Development of a polarimetric optical fiber sensor for electronic measurement of high pressure, *IEEE Trans. Instrum. Meas.*, 39(5), 548–555, 1990.
24. Boylstead, R. and L. Nashelsky, *Electronic Devices and Circuit Theory*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 1987.
25. Breer, H., Sense of smell: signal recognition and transduction in olfactory neurons, Chapter 22 in *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed., CRC Press, Boca Raton, FL, 1997.

26. Briand, D., Low-power electronic devices for gas sensing, At URL:www-samlab.unine.ch/Activities/Projects/Gas_sensor/Micro-hotplate.html, 2001.
27. Browne, A.F., T.R. Nelson and R.B. Northrop, Microdegree polarimetric measurement of glucose concentrations for biotechnology applications, *Proc. 23rd Ann. Northeast Bioeng. Conf.*, J.R. Lacourse, ed., UNH/Durham 5/21/22/97, 9–10, 1997.
28. Cameron, B.D. and G.L. Coté, Noninvasive glucose sensing utilizing a digital closed-loop polarimetric approach, *IEEE Trans. Biomed. Eng.*, 44(12), 1221–1227, 1997.
29. Cannon, R.H., Jr., *Dynamics of Physical Systems*, McGraw-Hill Book Co., NY, 1967.
30. Capan, R., A.K. Ray, A.K. Hassan and T. Tanriserver, Poly(methyl methacrylate) films for organic vapour sensing, *J. Phys. D: Appl. Phys.*, 36, 1115–1119, 2003.
31. Caruso, M.J., T. Bratland, C.H. Smith and R. Schneider, AMR magnetic field sensors, www.sensorsmag.com/articles/0399/0399_18/main.shtml, 14, March 1999a.
32. Caruso, M.J. and L.S. Withanawasam, Vehicle detection and compass applications using AMR magnetic sensors, www.ssec.Honeywell.com/magnetic/datasheets/amr.pdf, 1999b.
33. Cease, T.W. and P. Johnston, A magneto-optical current transducer, *IEEE Trans. Power Delivery*, 5(2), 548–555, 1990.
34. Chen, W.-K., *Passive and Active Filters: Theory and Implementation*, John Wiley & Sons, NY, 1986.
35. Cheo, P.K., *Fiber Optics and Optoelectronics*, Prentice-Hall, Englewood Cliffs, NJ, 1990.
36. Chirlian, P.M., *Analysis and Design of Integrated Electronic Circuits*, Harper & Row, NY, 1981.
37. Clarke, K.K. and D.T. Hess, *Communication Circuits: Analysis and Design*, Addison-Wesley Pub. Co. Reading, MA, 1971.
38. Clarke, J. and R.H. Koch, The impact of high temperature superconductivity on SQUID magnetometers, *Science*, 217–223, 14 October, 1988.
39. Cooley, J.W. and J.W. Tukey, An algorithm for the machine computation of complex Fourier series, *Math. Comput.*, 19, 297–301, 1965.
40. Cooper, W.D., *Electronic Instrumentation and Measurement Techniques*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1978.
41. Coté, G.L., M.D. Fox and R.B. Northrop, Optical polarimetric sensor for blood glucose measurement, *Proc. 16th Ann. Northeast Bioeng. Conf.*, R. Gaumond, ed., IEEE Press, NY, 102–103, 1990.
42. Coté, G.L., M.D. Fox and R.B. Northrop, Noninvasive optical polarimetric glucose sensing using a true phase measurement technique, *IEEE Trans. Biomed. Eng.*, 39(7), 752–756, 1992.
43. Coté, G.L., M.D. Fox and R.B. Northrop, Optical Glucose Sensor Apparatus and Method, U.S. Patent No. 5,209,231, 1993.
44. Crum, L.A., Sonoluminescence, *Physics Today*, 22–29, September, 1994.
45. Crum, L.A. and T.J. Matula, Shocking revelations, *Science*, 276, 1348–1349, 30 May, 1997.
46. Culshaw, B., *Optical Fiber Sensing and Signal Processing*, Peter Perigrinus, London, 1984.
47. Cutting, C.L., A.C. Jason and J.L. Wood, *J. Sci. Instrum.*, 32, 425, 1995.
48. Dahake, S.L. et al, Progress in the realization of the units of capacitance, resistance and inductance at the National Physical Laboratory, India, *IEEE Trans. Instrum. Meas.*, 32(1), 5–8, 1983.
49. Dandridge, A., Fiber optic sensors based on the Mach-Zehnder and Michelson interferometers, Chapter 10 in *Fiber Optic Sensors: An Introduction for Engineers and Scientists*, E. Udd, ed., John Wiley & Sons, NY, 1991.
50. Davis, C.M., *Fiber Optic Sensor Technology Handbook*, Optical Technologies, Inc., Herndon, VA, 1986.
51. Delahaye, F., A. Fau, D. Dominguez and M. Bellon, Absolute determination of the farad and ohm, and measurement of the quantized Hall resistance $R_H(2)$ at LCIE, *IEEE Trans. Instrum. Meas.*, 36(2), 205–207, 1987.
52. Deliyannis, T., Y. Sun and J.K. Fiddler, *Continuous-Time Active Filter Design*, CRC Press, Boca Raton, FL, 1998.
53. DePriest, D., *Differential GPS*, At URL: www.gpsinformation.org/dale/dgps.htm, 2003.
54. Didenko, Y.T., W.B. McNamara III and K.S. Suslick, Molecular emission from single-bubble sonoluminescence, *Nature*, 407, 877–879, 19 October, 2000.
55. Dorf, R.C., *Modern Control Systems*, Addison-Wesley, Reading, MA, 1967.

56. Drake, A.D. and D.C. Leiner, A fiber Fizeau interferometer for measuring minute biological displacements, *IEEE Trans. Biomed. Eng.*, 31(7), 507–511, 1984.
57. Drysdale, C.V., A.C. Jolley and G.F. Tagg, *Electrical Measuring Instruments: Part 1*, 2nd ed., John Wiley & Sons, NY, 1952.
58. Du, Z., A Frequency-Independent, High Resolution Phase Meter, MS dissertation in Biomedical Engineering at the University of Connecticut, Storrs, 1993.
59. Egan, W.F., *Frequency Synthesis by Phaselock*, John Wiley & Sons, NY, 1981.
60. Eisthen, H.L., Why are olfactory systems of different animals so similar? *Brain Behav. Evol.*, 59, 273–293, 2002.
61. Eisner Safety Consultants, *Overview of IEC60601-1 Medical Electrical Equipment*, Found at URL: www.EisnerSafety.com, 2001.
62. Endo, T., M. Koyangi and A. Nakamura, High-accuracy Josephson potentiometer, *IEEE Trans. Instrum. Meas.*, 32(1), 267–271, 1983.
63. Exar Integrated Systems, Inc., *Phase-Locked Loop Data Book*, Sunnyvale, CA, 1979.
64. F.A.A., Satellite Navigation Product Teams: Programs—Wide Area Augmentation System, At URL: <http://gps.faa.gov/Programs/WAAS/waas-text.htm>, 2003.
65. Fellows, K.R., The Design of a Non-Dispersive Spectrophotometer to Measure Oxyhemoglobin, MS dissertation in Biomedical Engineering, The University of Connecticut, Storrs, 1997.
66. Foreman, J.W. Jr, E.W. George and R.D. Lewis, Measurement of localized flow velocities in gasses with a laser Doppler flow meter, in *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed., SPIE Press, 8–9, 1993.
67. Foreman, J.W. Jr, R.D. Lewis, J.R. Thornton and H.J. Watson, Laser Doppler velocimeter for measurement of localized flow velocities in liquids, in *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed., SPIE Press, 10-, 1993.
68. Foster, M.W., D.J. Ferrel and R.A. Lieberman, Surface plasmon resonance biosensor miniaturization, *Proc. SPIE*, 2293, 122–131, 1994.
69. Fox, M.D., Multiple crossed-beam ultrasound Doppler velocimetry, *IEEE Trans. Sonics Ultrasonics*, 25(5), 281–286, 1978.
70. Fox, M.D. and J.F. Donnelly, Simplified method for determining piezoelectric constants for thickness mode transducers, *J. Acoust. Soc. Am.*, 64(5), 1261–1265, 1978.
71. Fox, M.D. and W.M. Gardiner, Three-dimensional Doppler velocimetry of flow jets, *IEEE Trans. Biomed. Eng.*, 35(10), 834–841, 1988.
72. Fox, M.D. and L.G. Puffer, Model for short term movements in *Stapelia variegata* L, *Plant Physiol.*, 61, 209–212, 1978.
73. Franco, S., *Design With Operational Amplifiers and Integrated Circuits*, McGraw-Hill Book Co., NY, 1988.
74. Frank, C., Raman analysis in pharmaceuticals, Kaiser Optical Systems, *Raman Review*, At URL: www.kosi.com/raman/resources/ramanreview/1998sept.pdf, September, 1998.
75. Gardner, F.M., *Phaselock Techniques*, 2nd ed., John Wiley & Sons, NY, 1979.
76. Geddes, L.A. and L.E. Baker, *Principles of Applied Biomedical Instrumentation*, John Wiley & Sons, NY, 1968.
77. Geddes, N.J. and C.R. Lawrence, Monitoring Immunoreactions with SPR, Chapter 16 in *Handbook of Biosensors and Electronic Noses*, E. Kress-Rogers, ed., CRC Press, Boca Raton, FL, 349–368, 1997.
78. George, W.K. and J.L. Lumley, The laser Doppler velocimeter and its application to the measurement of turbulence, *J. Fluid Mech.*, 60(2), 321–362, 1973.
79. Ghaussi, M.S., *Electronic Circuits*, D. Van Nostrand Co., NY, 1971.
80. Gilham, E.J., A high-precision photoelectric polarimeter, *J. Sci. Instrum.*, 34, 435–439, 1957.
81. Goldstein, R.J. and W.F. Hagen, Turbulent flow measurements utilizing the Doppler shift of scattered laser radiation, in *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed., SPIE Press, 475-, 1993.
82. Goldstein, R.J. and D.K. Kreid, Measurement of laminar flow development in a square duct using a laser Doppler flowmeter, in *Selected Papers on Laser Doppler Velocimetry*, R.J. Adrian, ed., SPIE Press, 469–474, 1993.

83. Goldwasser, S.M., *Diode Lasers*, 103, web article (Sam's Laser FAQ.) at URL: <http://www.laserfaq.com/laserdio.htm>, 2001.
84. Gowar, J., *Optical Fiber Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
85. Graeme, J.G., *Applications of Operational Amplifiers*, McGraw-Hill, NY, 1974.
86. Gray, P.R. and R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, John Wiley & Sons, NY, 1984.
87. Grebene, A.B., The monolithic phase-locked loop—a versatile building block, *IEEE Spectrum*, 38–49, March, 1971.
88. Greenspan, D. and V. Casulli, *Numerical Analysis for Applied Mathematics, Science and Engineering*, Addison-Wesley Pub. Co., Reading, MA, 1993.
89. Grubb, W.T. and L.H. King, Palladium–palladium oxide pH electrodes, *Anal. Chem.*, 52, 270–273, 1980.
90. Guyton, A.C., *Textbook of Medical Physiology*, 8th ed., W.B. Saunders Co., Philadelphia, 1991.
91. Hamilton, C.A. et al, A 24 GHz Josephson array voltage standard, *IEEE Trans. Instrum. Meas.*, 40(2), 301–304, 1991.
92. Hanlon, E.B. et al, Prospects for *in vivo* Raman Spectroscopy, *Phys. Med. Biol.*, 45(2), R1–R59, 2000.
93. Hannaford, B. and S. Lehman, Short-time Fourier analysis of the electromyogram: fast movements and constant contraction, *IEEE Trans. Biomed. Eng.*, 33(12), 1173–1181, 1986.
94. Hansen, A.T., Fiber-optic pressure transducers for medical applications, *Sensors and Actuators*, 4, 545–554, 1983.
95. Hecht, E., *Optics*, 2nd ed., Addison-Wesley, Reading, MA, 1987.
96. Helfrick, A.D. and W.D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 1990.
97. Herzel, F., H. Erzgräber and P. Weger, Integrated CMOS wideband oscillator for RF applications, *Electron. Lett.*, 37(6), 2001.
98. Hey, J.C. and W.P. Kram, eds., *Transient Voltage Suppression Manual*, 2nd ed., General Electric Co., Auburn, NY, 1978.
99. Hiller, R.A. and B.P. Barber, Producing light from a bubble of air, *Sci. Am.*, 96–98, February, 1995.
100. Hiller, R.A., K. Weninger, S.J. Puttermann and B.P. Barber, Effect of noble gas doping in single-bubble sonoluminescence, *Science*, 266, 248–250, 14 October, 1994.
101. Hobbs, S., A Tour of the Cyranose® 320, At URL: <http://cyranosciences.com/technology/tour.html>, 2000.
102. Hochburg, R.C., Fiber-optic sensors, *IEEE Trans. Instrum. Meas.*, 35(4), 447–450, 1986.
103. Hodgkin, A.L. and A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol., London*, 117, 500–544, 1952.
104. Homola, J., S.S. Yee and G. Gauglitz, Surface plasmon resonance sensors: review, *Sensors Actuators B*, 54, 3–15, 1999.
105. Hong, H.-D., Optical Interferometric Measurement of Skin Vibration for the Diagnosis of Cardiovascular Disease, PhD Dissertation in Biomedical Engineering, University of Connecticut, Storrs, 1994.
106. Hong, H.-D., and M.D. Fox, Detection of skin displacement and capillary flow using an optical stethoscope, *Proc. 19th Ann. Northeast Bioeng. Conf.*, IEEE Press, NY, 189–190, 1993.
107. Huelsmann, L.P., *Active and Passive Analog Filter Design: An Introduction*, McGraw-Hill Book Co., NY, 1993.
108. Igarishi, T., Y. Koizumi and M. Kanno, Determination of an absolute capacitance by a horizontal cross-capacitor, *IEEE Trans. Instrum. Meas.*, 17(4), 226–231, 1968.
109. Jacob, T., *Olfaction: A Tutorial on the Sense of Smell*, At URL: www.cf.ac.uk/biosi/staff/Jacob/teaching/sensory/olfact1.html, 2003.
110. Jaeger, K.B., P.D. Levine and C.A. Zack, Industrial experience with a quantified Hall effect system, *IEEE Trans. Instrum. Meas.*, 40(2), 256–261, 1991.
111. James, H.J., N.B. Nichols and R.S. Philips, *Theory of Servomechanisms*, McGraw-Hill, NY, 1947.
112. Jory, M.J. et al, A surface-plasmon-based optical sensor using acousto-optics, *Meas. Sci. Technol.*, 6, 1193–1200, 1995.

113. Juneau, T. and A.P. Pisano, Micromachined dual axis angular rate sensor, *Proc. IEEE Solid State Sensor Actuator Workshop*, Hilton Head Island, SC, June, 1996.
114. Juneau, T., A.P. Pisano and J.H. Smith, Dual-axis operation of a micromachined rate gyroscope, *Tech. Dig. 9th Int. Conf. Solid-State Sensors Actuators (Transducers '97)*, Chicago, 883–886, June, 1997.
115. Kaiser, S., T. Maier, A. Grossmann and C. Zimmerman, Fizeau interferometer for phase-shifting interferometry in ultrahigh vacuum, *Rev. Sci. Instr.*, 72(9), 3726–3727, 2001.
116. Kaiser Optical Systems, Inc., *Raman Tutorial*, 52, at URL: <http://nte-serveur.univ-lyon1.fr/nte/spectroscopie/raman/H1TUTO~1.htm>, 1998.
117. Kanai, H., K. Noma and J. Hong, Advanced spin-valve GMR-head. *Fujitsu Sci. Tech. J.*, 37(2), 174–182, December, 2001.
118. Kandel, E.R., J.H. Schwartz and T.M. Jessel, *Principles of Neural Science*, 3rd ed., Appleton and Lange, Norwalk, CT, 1991.
119. Katz, B., *Nerve, Muscle and Synapse*, McGraw-Hill, NY, 1996.
120. Keiser, G., *Optical Fiber Communications*, McGraw-Hill, NY, 1983.
121. Keller, P.E. et al, Electronic noses and their applications, *Proc. IEEE Nothcon/Technical App. Conf. (TAC'95)*, Portland, Oregon, 12 October, 1995.
122. Kibble, B.P., Realizing the ampere by levitating a superconducting mass—a suggested procedure, *IEEE Trans. Instrum. Meas.*, 32(1), 144, 1983.
123. Kinley, H., *The PLL Synthesizer Cookbook*, Tab Books, Inc., Blue Ridge Summit, PA, 1980.
124. Kitchin, C. and L. Counts, *RMS to DC Conversion Application Guide*, Analog Devices Inc., Norwood, MA, 1983.
125. Knapp, C.H., R. Bansal, M.S. Mashikian and R.B. Northrop, Signal processing techniques for partial discharge site location in shielded cables, *IEEE Trans. Power Delivery*, 5(2), 859–865, 1990.
126. Koch, D., *Johannes Kepler: His Life, His Laws and Times*, at URL: www.Kepler.arc.nasa.gov/Johannes.html, 2002.
127. Kollár, I, The noise model of quantization, *Proc. IMEKO TC4 Symposium, Noise in Electrical Measurements*, Como, Italy, June 19–21, 1986; OMIKK-Technoinform, Budapest, 125–129, 1987.
128. Koo, T-W. et al, Reagentless blood analysis by near-infrared Raman spectroscopy, *Diabetes Technol. Ther.*, 1(2), 153–157, 1999.
129. Koskinen, M., J. Typpö and J. Kostamovaara, A fast time-to-amplitude converter for pulsed time-of-flight laser rangefinding, *SPIE*, 1633, 128–136, 1992.
130. Krauss, J.D., *Electromagnetics*, McGraw-Hill Book Co. Inc., NY, 1953.
131. Kuo, B.C., *Automatic Control Systems*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 1982.
132. Lambert, J.L., J.M. Morookian, S.J. Sirk and M.S. Borchert, Measurement of aqueous glucose in a model anterior chamber using Raman spectroscopy, *J. Raman Spectroscopy*, 33(7), 524–529, 2002.
133. Lambert, J.L., M. Storrie-Lombardi and M. Borchert, Measurement of physiological glucose levels using Raman spectroscopy in a rabbit aqueous humor model, *IEEE/LEOS Newslett.*, 12, 19–22, April, 1998.
134. Lancaster, D., *Lancaster's Active Filter Cookbook*, 2nd ed., Butterworth-Heinemann, 1996.
135. Langley, R.B., Smaller and smaller: the evolution of the GPS receiver, *GPS World*, 54–58, April, 2000.
136. Lathi, B.P., *Signals, Systems and Communication*, John Wiley & Sons, NY, 1965.
137. Lavallée, M., O.F. Schanne and N.C. Hébert, *Glass Microelectrodes*, John Wiley & Sons, Inc., NY, 1969.
138. Leffingwell, J.C., *Olfaction*. [An excellent illustrated review paper describing the anatomy, neurophysiology, cell physiology, and biochemistry of the vertebrate olfactory process], at URL: www.leffingwell.com/olfaction.htm, 1999.
139. Leffingwell, J.C., Olfaction—III, *Leffingwell Reports*, 1(5), December, 2001a.
140. Leffingwell, J.C., Olfaction—I.. at URL: www.leffingwell.com/olfact2.htm, 2001b.
141. Levitsky, I., S.G. Krivoshylkov and J.W. Grate, Rational design of a Nile red/polymer composite film for fluorescence sensing of organophosphonate vapors using hydrogen bond acidic polymers, *Anal. Chem.*, 73, 3441–3448, 2001.

142. Lewis, N.S., Use of conducting polymer composite chemically sensitive resistors as an 'electronic nose', at URL: www.its.caltech.edu/~mmrc/nsi/nose.htm, 2000.
143. Lion, K.S., *Instrumentation in Scientific Research: Electrical Input Transducers*, McGraw-Hill Book Co., Inc, NY, 1959.
144. Lohse, D., Sonoluminescence: Inside a micro-reactor, *Nature*, 418, 381–383, 25 July, 2002.
145. McConnell, R.L., Capacitive Proximity Sensors With Additional Driven Shields, *NASA Tech Briefs Technical Support Package GSC-13475*, Goddard Space Flight Center, Greenbelt, MD , 1993.
146. McDonald, B.M. and R.B. Northrop, Two-phase lock-in amplifier with phase-locked loop vector tracking, *Proc. Euro. Conf. Circuit Theory Design*, Davos, Switzerland, 6, 30 August–3 September, 1993.
147. Macovski, A., *Medical Imaging Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
148. Marcuse, D., *Principles of Optical Fiber Measurements*, Academic Press, NY, 1981.
149. Maron, S.H. and C.F. Prutton, *Principles of Physical Chemistry*, MacMillan Co., NY, 1958.
150. Mashikian, M.S., R.B. Northrop, D. Sui, Cable Fault Detection Using a High-Voltage Alternating Polarity DC Signal Superposed with a System Frequency AC Signal, U.S. Patent No. 5,448,176, 5 September, 1995.
151. Mashikian, M.S., R.B. Northrop, R. Bansal and C.L. Nikias, Method and Instrumentation for the detection, Location, Characterization of Partial Discharges and Faults in Electric Power Cables, U.S. Patent No. 4,887,041, 12 December, 1989.
152. Mashikian, M.S., R.B. Northrop, R. Bansal and F. Palmieri, Method and Apparatus for the Detection and Location of Faults and Partial Discharges in Shielded Cables, U.S. Patent No. 5,272,439, 1993.
153. Mason, J.H., Discharge detection and measurements, *Proc. IEE*, 112(7), 1407–1422, 1965.
154. Mentelos, R.M., *Electrical Safety in PC Based Medical Products*, App. Note, AN-10, RAM Technologies LLC, Guilford, CT. www.ramtechno.com, 2003.
155. *Merck Manual of Diagnosis and Therapy*, 17th ed. (Centennial Edition), M.H. Beers, and R. Berkow, eds., at URL: www.merck.com/pubs/mmanual/section20/chapter277/277a.htm, accessed October, 2003.
156. Millman, J., *Microelectronics*, McGraw-Hill Book Co., NY, 1979.
157. Möller, K.D., *Optics*, University Science Books, Mill Valley, CA, 1988.
158. Mopsik, F.I., Precision time-domain dielectric spectrometer, *Rev. Sci. Instrum.*, 55(1), 79–87, 1984.
159. Moseley, P.T., J. Norris and D. Williams, *Techniques and Mechanisms in Gas Sensing*, Adam Hilger, NY, 1991.
160. Moss, W.C., D.B. Clarke and D.A. Young, Calculated pulse widths and spectra of a single sonoluminescing bubble, *Science*, 376, 1398–1401, 30 May, 1997a.
161. Moss, W.C., D.B. Clarke and D.A. Young, Star in a jar: a new model for single-bubble sonoluminescence, *Proc. 133rd ASA Meeting*, State College, PA, 1997b.
162. Nagle, H.T., S.S. Schiffman and R. Gutierrez-Osuna, The how and why of electronic noses, *IEEE Spectrum*, 35(9), 22–34, 1998.
163. Nanavati, R.P., *Semiconductor Devices: BJTs, JFETs, MOSFETs, and Integrated Circuits*, Intext Educational Publishers, NY, 1975.
164. National Instruments, *How IVI-C Instrument Driver Technology Enables System Longevity and Platform Portability*, at URL: <http://zone.ni.com/devzone/conceptd.nsf/webmain/...>, 2003a.
165. National Instruments, *IVI Architecture*, At URL: www.ni.com/ivi/ivi_arch.htm, 2003b.
166. National Instruments, *IVI: Comparison of IVI-C and IVI-COM Drivers*, at URL: www.ni.co/ivi_c_com.htm, 2003c.
167. National Instruments, *Measurement and Automation Catalog*, PXI Tutorial, 777–778, 2004.
168. Nelson, R.A., *The Global Positioning System: A National Resource*, at URL: www.aticourses.com/global_positioning_system.htm, 1999.
169. Nelson, T.R., Development of a Type 1 Nonlinear Feedback System for Laser Velocimetry and Ranging, MS dissertation in Biomedical Engineering at the University of Connecticut, Storrs, 1999.
170. Nemat, A., A digital frequency independent phase meter, *IEEE Trans. Instrum. Meas.*, 39(4), 665–666, 1990.

171. Neuman M.R., *Neonatal monitoring*, *Encyclopedia of Medical Devices and Instrumentation*, J.G. Webster, ed., John Wiley & Sons, NY, 1988.
172. Newport Photonics, *Laser Diode Testing*, A tutorial at URL: www.newport.com, 2003.
173. Nicati, P.A. and P. Robert, Stabilised current sensor using Sagnac interferometer, *J. Phys. Eng. Sci. Instrum.*, 21, 791–796, 1988.
174. Nise, N.S., *Control Systems Engineering*, Benjamin Cummings, Redwood City, CA, 1995.
175. Northrop, R.B., *Analog Electronic Circuits: Analysis and Applications*, Addison-Wesley, Reading, MA, 1990.
176. Northrop, R.B., *Introduction to Instrumentation and Measurements*, CRC Press, Boca Raton, FL, 1997.
177. Northrop, R.B., *Endogenous and Exogenous Regulation and Control of Physiological Systems*, CRC Press, Boca Raton, FL, 2000.
178. Northrop, R.B., *Introduction to Dynamic Modeling of Neuro-sensory Systems*, CRC Press, Boca Raton, FL, 2001.
179. Northrop, R.B., *Non-Invasive Instrumentation and Measurements in Medical Diagnosis*, CRC Press, Boca Raton, FL, 2002.
180. Northrop, R.B., *Signals and Systems Analysis in Biomedical Engineering*, CRC Press, Boca Raton, FL, 2003.
181. Northrop, R.B., *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation*, CRC Press, Boca Raton, FL, 2004.
182. Northrop, R.B. and B.M. Decker, Design of a no-touch infant apnea monitor, *Proc. 5th Ann. North-east Bioeng. Conf.*, Pergamon Press, NY, 245–248, 1977.
183. Northrop, R.B. and H.J. Grossman, An integrated-circuit pulse-height discriminator with multiplexed display, *J. Appl. Physiol.*, 37(6), 946–950, 1974.
184. Northrop, R.B. and H.M. Horowitz, An instantaneous pulse frequency demodulator for neurophysiological applications, *Proc. Symp. Biomed. Eng.*, Milwaukee, 1, 5–8, 1966.
185. Northrop, R.B., J.-M. Wu and H.M. Horowitz, An instantaneous frequency cardiotachometer, *Digest 7th Int. Conf. Eng. Med. Biol.*, Stockholm, 417, 1967.
186. Ogata, K., *Modern Control Engineering*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1990.
187. Oliver, B.M. and J.M. Cage, *Electronic Measurements and Instrumentation*, McGraw-Hill, NY, 1971.
188. Olsen, H.F., *Elements of Acoustical Engineering*, D. Van Nostrand Co. Inc., NY, 1943.
189. Ott, H.W., *Noise Reduction Techniques in Electronic Systems*, John Wiley & Sons, NY, 1976.
190. Ouellette, J., Electronic noses sniff out new markets, *The Industrial Physicist*, 26–29, February, 1999.
191. Pallàs-Areny, R. and J.G. Webster, *Sensors and Signal Conditioning*, 2nd ed., John Wiley & Sons, Inc., NY, 2001.
192. Pankratz, C.K., Microchannel plate (MCP) detectors, at URL: www.pha.jhu.edu/~pankratz/detector_pres/node13.html, 1998.
193. Papoulis, A., *Systems and Transforms with Applications to Optics*, McGraw-Hill, NY, 1968.
194. Papoulis, A., *Signal Analysis*, McGraw-Hill Book Co., NY, 1977.
195. Perkin-Elmer, *Avalanche Photodiodes: A User's Guide*, at URL: <http://optoelectronics.perkinelmer.com/library/papers/tp5.htm>, 2003.
196. Peslin, R., et al, Frequency response of the chest: modeling and parameter estimation, *J. Appl. Physiol.*, 39(4), 523–534, 1975.
197. Peterson, J.I. and S.R. Goldstein, A miniature fiberoptic pH sensor potentially suitable for glucose measurements, *Diabetes Care*, 5(3), 272–274, 1982.
198. Peura R.A. and Y. Mendelson, Blood glucose sensors, an overview, *Proc. IEEE/NSF Symp. Biosensors*, 63–68, 1984.
199. Phillips, C.L. and H.T. Nagle, *Digital Control System Analysis and Design*, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.
200. Pimmel, R.L. et al, Instrumentation for measuring respiratory impedance by forced oscillations, *IEEE Trans. Biomed. Eng.*, 24(2), 89–93, 1977.
201. Piquemal, F. et al, Direct comparison of quantized Hall resistances, *IEEE Trans. Instrum. Meas.*, 40(2), 234–236, 1991.

202. Plonsey, R. and D.G. Fleming, *Bioelectric Phenomena*, McGraw-Hill Book Co, NY, 1969.
203. Pöpel, R. *et al*, Nb/Al₂O₃/Nb Josephson voltage standards at 1 V and 10 V, *IEEE Trans. Instrum. Meas.*, 40(2), 298–300, 1991.
204. Proakis, J.G. and D.G. Manolakis, *Introduction to Digital Signal Processing*, Macmillan Publishing Co., NY, 1989.
205. PXI Systems Alliance, *PXI Hardware Specifications*, at URL: www.pxisa.org/, 2003.
206. Quinn, T.J., The kilogram: the present state of our knowledge, *IEEE Trans. Instrum. Meas.*, 40(2), 81–85, 1991.
207. Rabinovitch, B., W.F. March and R.L. Adams, Noninvasive glucose monitoring of the aqueous humor of the eye. Part I. Measurements of very small optical rotations, *Diabetes Care*, 5(3), 254–258, 1982.
208. Rakow, N.A. and K.S. Suslick, A colorimetric sensor array for odor visualization, *Nature*, 406, 710–713, 2000.
209. Reymann, D., A practical device for 1 nV accuracy measurements with Josephson arrays, *IEEE Trans. Instrum. Meas.*, 40(2), 309–311, 1991.
210. Ribeiro, P.C., S.J. Williamson and L. Kaufman, SQUID arrays for simultaneous magnetic measurements: calibration and source localization performance, *IEEE Trans. Biomed. Eng.*, 35(7), 551–559, 1988.
211. Ripka, P., M. Tondra, J. Stokes and R. Beech, AC-driven AMR and GMR magnetosensors, *Eurosensors XII Proc.*, 967–969.
212. Rogers, A.J., Optical technique for measurement of current at high voltage, *Proc. IEEE*, 120(2), 261–267, 1973.
213. Rogers, A.J., Method for simultaneous measurement of voltage and current on high voltage lines using optical techniques, *Proc. IEEE*, 123(10), 957–960, 1976.
214. Rogers, A.J., The electrogyration effect in crystalline quartz, *Proc. Royal Soc. London, A.*, 353, 177–192, 1977.
215. Rogers, A.J., Optical measurement of current and voltage on power systems, *Electr. Power Appl.*, 2(4), 22–25, 1979.
216. Sagnac, G., *Effet tourbillonnaire optique la circulation de l'éther lumineux dans un interféromètre tournante*. *J. Phys. Radium*, 4, 177–195, 1914.
217. Schalkhammer, T., Metal nanoclusters as transducers for bioaffinity reactions, *Monatshefte für Chemie*, 129, 1067–1092, 1998.
218. Schaumann, R. and M.E. Van Valkenburg, *Design of Analog Filters*, Oxford University Press, 2001.
219. Scheid, F.S., *2000 Solved Problems in Numerical Analysis*, McGraw-Hill, NY, 1990.
220. Schilling, D.L. and C. Belove, *Electronic Circuits: Discrete and Integrated*, 3rd ed., McGraw-Hill College Div., NY, 1989.
221. Sears, F.W., *Optics*. Addison-Wesley Press, Inc., Cambridge, MA, 1949.
222. Sears, F.W., *Electricity and Magnetism*, Addison-Wesley Pub. Co., Cambridge, MA, 1953.
223. Shafer-Peltier, K.E., C.L. Haynes, M.R. Glucksberg and R.P. Van Duyne, Toward a glucose biosensor based on surface-enhanced Raman scattering, *J. Am. Chem. Soc.*, 2002.
224. Shida, L. *et al*, SI value of quantized Hall resistance based on ETL's calculable capacitor, *IEEE Trans. Instrum. Meas.*, 38(2), 252–255, 1989.
225. Shields, J.Q., R.F. Dziuba and H.P. Layer, New realization of the ohm and farad using the NBS calculable capacitor, *IEEE Trans. Instrum. Meas.*, 38(2), 249–251, 1989.
226. Simon, H.J., Sensor Using Long Range Surface Plasmon Resonance with Diffraction Double Grating, U.S. Patent No. 5,846, 843, 8 December, 1998.
227. Simpson, D., *Universal Serial Bus (USB) Information*, at URL: www.sss-mag.com/usb.html, 2003.
228. Sirohi, R.S. and H.P. Kothiyal, *Optical Components, Systems and Measurement Techniques*, Marcel Dekker, NY, 1991.
229. Soltani, P.K., D. Wysnewska and K. Swartz, Amorphous selenium direct radiography for industrial imaging, Paper 22, *Proc. Conf. Computerized Tomography for Industrial Applications and Image Processing in Radiology*, Berlin, 123–132, 15–17 March, 1999.

230. Sotzing, G.A. *et al*, Preparation and properties of vapor detector arrays formed from poly(3,4-ethelenedioxy)thiophene-poly(styrene-sulfonate) insulating polymer composites, *Anal. Chem.*, 72, 3181–3190, pdf, 2000a.
231. Sotzing, G.A. *et al*, Highly sensitive detection and discrimination of biogenic amines utilizing arrays of polyaniline/carbon black composite vapor detectors, *Chem. Mater.*, 12, 593–595, 2000b.
232. Stahovec, J.L., Safety Method and Apparatus for Sensing the Presence of Individuals Adjacent to a Vehicle, U.S. Patent No. 4,300,116, 10 November, 1981.
233. Staples, E.J., A new electronic nose, *Sensors*, 16, 33–40, 1999.
234. Stapleton, H. and A. O'Grady, Isolation techniques for high-resolution data acquisition systems, *EDN*, 1 February, 2001.
235. Stark, L., *Neurological Control Systems*, Plenum Press, NY, 1988.
236. Stark, H., F.B. Tuteur and J.B. Anderson, *Modern Electrical Communications*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1988.
237. Stone, H.S., *Microcomputer Interfacing*, Addison-Wesley Pub. Co., Reading, MA, 1982.
238. Stout, M.B., *Basic Electrical Measurements*, Prentice-Hall, Englewood Cliffs, NJ, 1950.
239. Stout, M.B., *Basic Electrical Measurements*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1960.
240. Suslick, K.S., W.B. McNamara III and Y. Didenko, Hot spot conditions during multi-bubble cavitation, in *Sonochemistry and Sonoluminescence*, Crum, L.A. *et al*, eds., Kluwer Publishers, Dordrecht, The Netherlands, 191–204, 1999.
241. Tan, T.T., V.O. Schmitt, Q. Lucas and S. Isz, Electronic noses and electronic tongues, *LabPLus International*, 16–19 September/October, 2001.
242. Tarr, R.V. and P.G. Steffes, Non-Invasive Blood Glucose measurement System and Method Using Stimulated Raman Spectroscopy, U.S. Patent No. 5,243,983, 14 September, 1993.
243. Tarr, R.V. and P.G. Steffes, The noninvasive measure of D-glucose in the ocular aqueous humor using stimulated Raman spectroscopy, *IEEE/LEOS Newslett.*, 12(2), 22–27, 1998.
244. Taub, H. and D.L. Schilling, *Digital Integrated Electronics*, McGraw-Hill, NY, 1977.
245. Taylor, B.N., New international representations of the volt and ohm effective January 1, 1990, *IEEE Trans. Instrum. Meas.*, 39(1), 2–5, 1990.
246. Taylor, B.N., The possible role of the fundamental constants in replacing the kilogram, *IEEE Trans. Instrum. Meas.*, 40(2), 86–91, 1991.
247. Tektronix Product Catalog, Tektronix, Inc., Wilsonville, OR, 1990.
248. Thompson, A.M. and D.G. Lampard, A new theorem in electrostatics and its application to calculable standards of capacitance, *Nature*, 177, 188, 1956.
249. Thompson, H.D. and W.H. Stevenson, *Laser Velocimetry and Particle Sizing*, Hemisphere Pub. Co., 1979.
250. Tompkins, W.J. and J.G. Webster, eds., *Design of Microcomputer-Based Medical Instrumentation*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
251. Tompkins, W.J. and J.G. Webster, eds., *Interfacing Sensors to the IBM® PC*, Prentice-Hall, NJ, 1988.
252. United States Coast Guard, Navigation Center: *DGPS General Information*, at URL: www.navcen.uscg.gov/dgps/default.htm, 2003.
253. van der Ziel, A., *Introductory Electronics*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
254. Vaughan, M. and J.M. Vaughan, *The Fabry Perot Interferometer: History, Theory, Practice, and Applications*, Adam Hilger, NY, 1989.
255. von Klitzing, K., The quantum Hall effect, in *The Physics of the Two-Dimensional Electron Gas*, J.T. Devreese and F.M. Peeters, eds., Plenum Press, NY, 1986.
256. von Klitzing, K., G. Dorda and M. Pepper, New method for high accuracy determination of the fine structure constant based in quantized Hall resistance, *Phys. Rev. Lett.*, 45, 494–496, 1980.
257. Vranish, J.M., Capacitive Proximity Sensor has Longer Range, *NASA Tech Briefs Technical Support Package*, GSC-13377, Goddard Space Flight Center, Greenbelt, MD, 1992a.
258. Vranish, J.M., Capacitive proximity sensor has longer range, *NASA Tech Briefs*, 22–23, August, 1992b.
259. Vranish, J.M., Capacitive sensor with driven shields and bridge circuit, *NASA Tech Briefs*, 34–35, August, 1994.

260. Waggoner, P. *et al*, Canine olfactory sensitivity to cocaine hydrochloride and methyl benzoate, *1st Ann. Symp. Enabling Technol. Law Enforcement*, International Society for Optical Engineering, Boston, 1997.
261. Watriasiewicz, B.M. and M.J. Rudd, *Laser Doppler Measurements*, Butterworths, London, 1976.
262. Webster, J.G., ed., *Medical Instrumentation: Application and Design*, Houghton-Mifflin, Boston, 1978.
263. Webster, J.G., ed., *Medical Instrumentation*, 2nd ed., Houghton-Mifflin, Boston, 1992.
264. West, J.B., ed., *Best and Taylor's Physiological Basis of Medical Practice*, 11th ed., Williams & Wilkins, Baltimore, 1985.
265. Widrow, B. and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, NY, 1985.
266. Williams, C.S., *Designing Digital Filters*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
267. Wilson, W.L., Jr., A Sensitive Magneto resistive MEMS Acoustic Sensor, Final Report, Contract #961004, Rice University, Houston, TX, 28, 1996.
268. Wilson, D.L., *David L. Wilson's GPS Accuracy Web Page*, at URL: <http://users.erols.com/dlwilson/gps.html>, 2003.
269. Wingard, L.B. *et al*, Potentiometric measurement of glucose concentration with an immobilized glucose oxidase/catalase electrode, *Diabetes Care*, 5(3), 199–202, 1982.
270. Wise, D.L., ed., *Applied Biosensors*, Butterworths, Boston, 1989.
271. Wiza, J.L., Microchannel plate detectors, *Nuclear Instruments and Methods*, 162, 587–601, 1979.
272. Wolf, S., Interference signals and their elimination or reduction, *Guide to Electronic Measurements and Laboratory Practice*, Chapter 16, 2nd ed., Prentice-Hall, NJ, 1983.
273. Wolfbeis, O.S., *Fiber Optic Sensors and Biosensors*, Vol 1, CRC Press, Boca Raton, FL, 1991.
274. Wroblewski, W., Field effect transistors (FETs) as transducers in electrochemical sensors, at URL: www.ch.pw.edu.pl/~dybko/csrg/isfet.html, 2002.
275. Yang, E.S., *Microelectronic Devices*, McGraw-Hill Book Co., NY, 1988.
276. Yarov, A., *Optical Electronics*, Saunders, Philadelphia, 1991.
277. Yasui, K., Effect of surfactants on single-bubble sonoluminescence, *Phys. Rev. Eng.*, 58(4), 4560–4567, 1998.
278. Yazdi, Y., F. Ayozi and K. Najafi, Micromachined inertial sensors, *Proc. IEEE*, 86(8), 1640–1658, 1988.
279. Yeh, Y. and H.Z. Cummins, Localized fluid flow measurements with a HeNe laser spectrometer, *Appl. Phys. Lett.*, 4(10), 176, 1964.
280. Ziemer, R.E. and W.H. Tranter, *Principles of Communications: Systems, Modulation, and Noise*, 3rd ed., Houghton-Mifflin Co., Boston, 1990.
281. Zuch, E.L., *Data Acquisition & Conversion Handbook*, 4th ed., Datel-Intersil, Inc., Mansfield, MA, 1981.